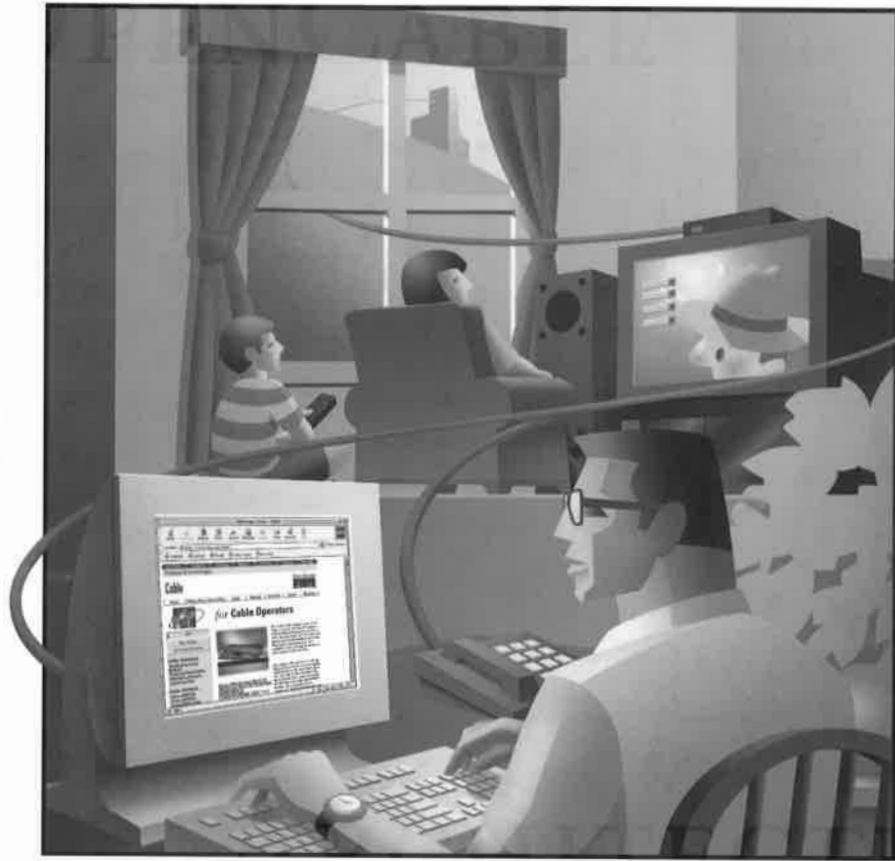


Michael
Adams



**The path to cable compatibility and retail
availability in digital television**

OPENCABLETM ARCHITECTURE

CISCO SYSTEMS

CISCO PRESS
www.ciscopress.com

ENG
TK
6678
A33
2000

LEONARD H. GEMMILL ENGINEERING LIBRARY
UNIVERSITY OF COLORADO AT BOULDER

DATE DUE

ENG	SEP 16	2000	

Demco, Inc. 38-293

University of Colorado at Boulder



U18303 1844534



OpenCable™ Architecture

Michael Adams



Cisco Press
201 West 103rd Street
Indianapolis, IN 46290 USA

OpenCable™ Architecture

Michael Adams

Copyright © 2000 Cisco Press

Cisco Press logo is a trademark of Cisco Systems, Inc.

Published by:

Cisco Press

201 West 103rd Street

Indianapolis, IN 46290 USA

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without written permission from the publisher, except for the inclusion of brief quotations in a review.

Printed in the United States of America 1 2 3 4 5 6 7 8 9 0

Library of Congress Cataloging-in-Publication Number: 99-61715

ISBN: 1-57870-135-X

Warning and Disclaimer

This book is designed to provide information about OpenCable™ architecture. Every effort has been made to make this book as complete and as accurate as possible, but no warranty or fitness is implied.

The information is provided on an "as is" basis. The author, Cisco Press, and Cisco Systems, Inc. shall have neither liability nor responsibility to any person or entity with respect to any loss or damages arising from the information contained in this book or from the use of the discs or programs that may accompany it.

The opinions expressed in this book belong to the author and are not necessarily those of Cisco Systems, Inc.

Feedback Information

At Cisco Press, our goal is to create in-depth technical books of the highest quality and value. Each book is crafted with care and precision, undergoing rigorous development that involves the unique expertise of members from the professional technical community.

Readers' feedback is a natural continuation of this process. If you have any comments regarding how we could improve the quality of this book, or otherwise alter it to better suit your needs, you can contact us through e-mail at ciscopress@mcp.com. Please make sure to include the book title and ISBN in your message.

We greatly appreciate your assistance.

Trademark Acknowledgments

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Cisco Press or Cisco Systems, Inc. cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark. OpenCable is a registered trademark of Cable Labs.

Publisher	John Wait
Executive Editor	Alicia Buckley
Cisco Systems Program Manager	Jim LeValley
Managing Editor	Patrick Kanouse
Project Manager	Kathy Trace
Project Editor	Theresa Wehrle
Copy Editor	Nancy Albright
Team Coordinator	Amy Lewis
Book Designer	Gina Rexrode
Cover Designer	Aren Howell
Compositor	Steve Gifford
Indexer	Tim Wright



Corporate Headquarters
 Cisco Systems, Inc.
 170 West Tasman Drive
 San Jose, CA 95134-1706
 USA
<http://www.cisco.com>
 Tel: 408 526-4000
 800 553-NETS (6387)
 Fax: 408 526-4100

European Headquarters
 Cisco Systems Europe s.a.r.l.
 Parc Evolic, Batiment L1/L2
 16 Avenue du Quebec
 Villebon, BP 706
 91961 Courtaboeuf Cedex
 France
<http://www-europe.cisco.com>
 Tel: 33 1 69 18 61 00
 Fax: 33 1 69 28 83 26

American Headquarters
 Cisco Systems, Inc.
 170 West Tasman Drive
 San Jose, CA 95134-1706
 USA
<http://www.cisco.com>
 Tel: 408 526-7660
 Fax: 408 527-0883

Asian Headquarters
 Nihon Cisco Systems K.K.
 Fuji Building, 9th Floor
 3-2-3 Marunouchi
 Chiyoda-ku, Tokyo 100
 Japan
<http://www.cisco.com>
 Tel: 81 3 5219 6250
 Fax: 81 3 5219 6001

Cisco Systems has more than 200 offices in the following countries. Addresses, phone numbers, and fax numbers are listed on the Cisco Connection Online Web site at <http://www.cisco.com/offices>.

Argentina • Australia • Austria • Belgium • Brazil • Canada • Chile • China • Colombia • Costa Rica • Croatia • Czech Republic • Denmark • Dubai, UAE • Finland • France • Germany • Greece • Hong Kong • Hungary • India • Indonesia • Ireland • Israel • Italy • Japan • Korea • Luxembourg • Malaysia • Mexico • The Netherlands • New Zealand • Norway • Peru • Philippines • Poland • Portugal • Puerto Rico • Romania • Russia • Saudi Arabia • Singapore • Slovakia • Slovenia • South Africa • Spain • Sweden • Switzerland • Taiwan • Thailand • Turkey • Ukraine • United Kingdom • United States • Venezuela

Copyright © 2000 Cisco Systems, Inc. All rights reserved. Printed in the USA. Access Registrar, AccessPath, Any to Any, AtmDirector, CCDA, CCDE, CGDP, CCIE, CCNA, CCNP, CCSI, CD-PAC, the Cisco logo, Cisco Certified Internetwork Expert logo, *CiscoLink*, the Cisco Management Connection logo, the Cisco NetWorks logo, the Cisco Powered Network logo, Cisco Systems Capital, the Cisco Systems Capital logo, Cisco Systems Networking Academy, the Cisco Systems Networking Academy logo, the Cisco Technologies logo, ConnectWay, ControlStream, Fast Step, FireRunner, GigaStack, IGX, Internet Quotient, Kernel Proxy, MGX, Natural Network Viewer, NetSonar, Network Registrar, *Packets*, PIX, Point and Click Internetworking, Policy Builder, Precept, RouteStream, Secure Script, ServiceWay, SlideCast, SMARTnet, StreamView, *The Cell*, TrafficDirector, TransPath, ViewRunner, VirtualStream, VisionWay, VlanDirector, Workgroup Director, and Workgroup Stack are trademarks; Changing the Way We Work, Live, Play, and Learn, Empowering the Internet Generation, The Internet Economy, and The New Internet Economy are service marks; and ASIST, BPX, Catalyst, Cisco, Cisco IOS, the Cisco IOS logo, Cisco Systems, the Cisco Systems logo, the Cisco Systems Cisco Press logo, Enterprise/Solver, EtherChannel, EtherSwitch, FastHub, FastLink, FastPAD, FastSwitch, GeoTel, IOS, IP/TV, IPX, LightStream, LightSwitch, MICA, NetRanger, Registrar, StrataView Plus, Stratum, TeleRouter, and VCO are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and certain other countries. All other trademarks mentioned in this document are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any of its resellers. (9908R)

About the Author

Michael Adams is the principal network architect for Time Warner Cable, responsible for all aspects of networking in the Pegasus Digital Program. He has served as the co-chair of the JEC Digital Standards Sub-Committee, and as chair of the Working Group 3 for the SCTE Digital Video Standards Committee. He is one of the primary architects of the OpenCable initiative, as well as the co-author of the point-of-deployment module interface (OCI-C2) specification and the primary author of the network interface (OCI-N) specification. He is a founding member of the OpenCable Technical Team.

Technical Reviewers

The technical reviewers of OpenCable Architecture are

Yvette Gordon, Vice President of Interactive Services, SeaChange

Andy Scott, Director of Engineering, National Cable Television Association

George Abe, Palomer Ventures

Joe Buehl, Principal Software Architect, Time Warner Cable

Louis Williamson, Senior Project Engineer, Time Warner Cable

Tom Lookabaugh, President, DiviCom

Charlie Kenamer, Senior Director, Engineering Services, AT&T Broadband and Internet Services

Laurie Priddy, President, National Digital Television Center; Senior Vice President, AT&T Broadband and Internet Services

Mike Hayashi, Vice President of Advanced Engineering, Time Warner Cable

Mark Eyer, Principal Staff Engineer, Sony Electronics

John Carlucci, Principal Engineer, DiviCom

Ron Boyer, Senior Project Engineer, Time Warner Cable

Luc Vantalou, Director of Digital TV, SCM Microsystems

Jim Chiddix, Chief Technical Officer, Time Warner Cable

Paul Bosco, Vice President and General Manager, Cable Products and Solutions, Wireless Products and Solutions, Cisco Systems

Dedications

To my father, Roy W. Adams, who said I would write a book someday.

Acknowledgments

Heartfelt thanks to everyone who helped to make this book a reality:

George Abe, for starting the ball rolling

Andy Scott and **Yvette Gordon**, for their excellent comments and additions to every single chapter

Louis Williamson and **Mike Hayashi**, for being wonderful cable mentors and for chapter reviews

Jim Chiddix, for support, encouragement and overall review

John Carlucci and **Tom Lookabaugh**, for unparalleled MPEG expertise

Joe Buehl, for software guidance and reviewing Part 2

Ron Boyer, for advice on the mysteries of HFC

Charlie Kennamer, for *Headend In The Sky* expertise

Laurie Priddy, for OpenCable chapter reviews

Luc Vantalou, for the point of deployment module

Mark Eyer, for writing so many DVS specifications

Paul Bosco, for overall review

Lynette Quinn, for keeping me on schedule

Kathy Trace, for being a great writing coach

Alicia Buckley, for keeping it all on track

Theresa Wehrle, for super copy editnig(sic)

Shelagh, David, and **Jack Adams**, for all those weekends and evenings that I disappeared into the home office.

Leslie Ellis, for collaboration on Chapter 1.

Contents at a Glance

Foreword

Introduction

Part I	Digital Cable Television
Chapter 1	Why Digital Television?
Chapter 2	Analog Cable Technologies
Chapter 3	The Analog Set-Top Converter
Chapter 4	Digital Technologies
Chapter 5	Adding Digital Television Services to Cable Systems
Chapter 6	The Digital Set-Top Converter
Chapter 7	Digital Broadcast Case Studies
Part II	Interactive and On-Demand Services
Chapter 8	Interactive Services
Chapter 9	Interactive Cable System Case Studies
Chapter 10	On-Demand Services
Chapter 11	On-Demand Cable System Case Studies
Part III	OpenCable
Chapter 12	Why OpenCable?
Chapter 13	OpenCable Architectural Model
Chapter 14	OpenCable Device Functional Requirements
Chapter 15	OpenCable Headend Interfaces
Chapter 16	OCI-N: The Network Interface
Chapter 17	OCI-C1: The Consumer Interface
Chapter 18	OCI-C2: The Security Interface
	Glossary

Contents

	Foreword	xx
	Introduction	xxii
Part I	Digital Cable Television	3
Chapter 1	Why Digital Television?	5
	Digital Technology Evolution	5
	Silicon Integration and Moore's Law	5
	Analog-to-Digital Conversion	6
	Convergence with the Personal Computer	7
	Internet Convergence	8
	New Services	8
	New Business Models	9
	Advantages of Digital Television	10
	Channel Expansion	10
	On-Demand Services	10
	Quality	11
	Security	11
	Flexibility	11
	Data Transmission	12
	Summary	12
Chapter 2	Analog Cable Technologies	15
	Analog Channel Expansion	15
	The Hybrid Fiber Coax Upgrade	16
	HFC Topology	18
	Linear Optical Transmission	20
	Return Path Activation	20
	Summary	22
	References	22
Chapter 3	The Analog Set-Top Converter	25
	The Cable Network Interface	27
	Cable Tuner	29
	NTSC Demodulator	29
	Out-of-Band Receiver	29
	Out-of-Band Transmitter	30
	Media Access Control	30

Conditional Access System	30
Analog De-scrambler	31
On-Screen Display	32
Audio Processing	33
Volume Control	33
Digital Music	33
BTSC Stereo and SAP Decoding	34
Microprocessor Subsystem	34
Central Processing Unit	35
Memory Subsystem	35
Display and Keypad	37
RF Modulator	37
RF Bypass Switch	37
Inputs	38
Cable Input	38
Infrared Receiver	38
Diagnostic Port	38
Data Port	39
Outputs	39
RF Output	40
Baseband Video	40
Baseband Audio	40
Infrared Transmitter	41
Software Architecture	41
Operating System Software	42
Device Drivers	42
Applications	43
Case Studies	43
CFT-2200	43
8600X	44
Limitations	45
Lessons Learned	45
Summary	46
References	46
Book	46
Internet Resources	47

Chapter 4	Digital Technologies	49
	Video Compression	50
	MPEG-2 Compression	51
	Other Video Compression Algorithms	55
	Details of MPEG-2 Video Compression	56
	Audio Compression	61
	MPEG-1 Layer 2 (Musicam)	61
	Dolby AC-3	61
	Other Audio Compression Algorithms	62
	Data	62
	System Information	62
	MPEG-2 Systems Layer	63
	Timing and Synchronization	64
	Packetization	64
	Multiplexing	65
	Conditional Access	66
	Limitations of MPEG-2 Systems Layer	67
	Transmission Mechanisms	67
	Baseband Transmission	68
	Broadband Transmission	75
	Summary	82
	References	83
	Books	83
	Periodicals	83
	Standards	83
	Internet Resources	84
Chapter 5	Adding Digital Television Services to Cable Systems	87
	Drivers for Digital Television	88
	Channel Expansion	89
	Direct Broadcast Satellite Competition	89
	High Definition Television	89
	Consolidation with Other Digital Services	90
	Radio Frequency Return Traffic	91

Business Communications	91
Network Management	91
Transmission of Digital Television	91
Out-of-Band Data Communications	93
Drivers	93
Out-of-Band Architectures	94
Forward OOB Channel	95
Reverse OOB Channel	96
Out-of-Band Evolution	100
Out-of-Band Channel Termination	100
Hub-Level Addressing	101
Return Traffic Aggregation	101
Shared Media Access Control	101
Headend-to-Distribution Hub Interconnection	102
LAN Extension Products	103
SONETs	103
ATM Networks	103
IP Networks	104
Summary	104
References	105
Books	105
Periodicals	105
Internet Resources	105
Chapter 6 The Digital Set-Top Converter	107
Cable Environment	108
Overview	109
The Cable Network Interface	110
Cable Input	112
Tuner	112
QAM Demodulator	112
NTSC Demodulator	113
Out-of-Band Channel Termination	113
Out-of-Band Transmitter	113
Media Access Control	114
Telephone Modem	114
Transport Processing	114
Conditional Access System	115

Digital Decryption	116
Analog De-scrambling	117
Video and Graphics Processing	118
MPEG-2 Video Decoding	118
Analog Video Processing	119
On-Screen Display	120
Digital-to-Analog Conversion	121
NTSC Encoding	121
Audio Processing	122
Dolby AC-3 Audio Decoding	123
BTSC Stereo Decoding	123
Audio Synthesis	123
Microprocessor Subsystem	124
Central Processing Unit	125
Memory Subsystem	125
Display and Keypad	127
RF Modulator	127
RF Bypass Switch	128
Inputs	128
Outputs	129
RF Output	129
Baseband Video	129
Baseband Audio	130
S-Video	130
Component Video	130
IEEE 1394	130
Digital Audio	131
Infrared Transmitter	131
Software	131
Device Drivers	132
Operating System	132
Applications	132
Limitations	133
Summary	133
References	133
Books	133
Standards	133
Internet Resource	134

Chapter 7	Digital Broadcast Case Studies	137
	Architectural Comparison	137
	Satellite Distribution to Headends	138
	Headend-to-Subscriber Distribution	140
	Split Security Model	141
	Out-of-Band Channel	142
	Central Versus Local Subscriber Management	142
	DigiCable	143
	Head-End In The Sky Model	145
	Local Subscriber Management	148
	DigiCable Summary	150
	Pegasus	150
	Pegasus Phase 1.0 Goals	151
	Pegasus Request for Proposal	152
	Pegasus Architecture	153
	Digital Broadband Delivery System	156
	Summary	160
	References	160
	Book	160
	Periodical	160

Part II Interactive and On-Demand Services 163

Chapter 8	Interactive Services	165
	Internet Convergence	166
	Goals of Interactive Services	167
	Interactive Versus On-Demand Services	167
	Interactive Services	168
	Navigation	169
	Information Services	169
	Communications Services	170
	Electronic Commerce Services	170
	Video Games	170
	Enhanced Television Services	171
	Applications Model	172
	Client Applications	172
	Server Applications	173
	Distributed Applications	174

Application Requirements	175
Software Download	176
Activation	176
Communications	177
Streaming Media	178
Application Resources	179
Set-Top Resources	179
Software Download Mechanisms	181
Activation and Synchronization Mechanisms	182
Communication Mechanisms	183
Summary	184
References	184
Books	184
Periodicals	185
Standards	185
Internet Resources	185
Chapter 9 Interactive Cable System Case Studies	187
Time Warner Full Service Network	187
FSN Network Architecture Goals	187
Network Overview	189
Services	194
Applications Model	196
Application Requirements	197
Applications Resources	198
Lessons Learned	199
Pegasus	200
Pegasus Phase 1.1 Goals	200
Network Overview	201
Services	201
Applications Model	202
Application Portability	203
Applications Resources	204
Lessons Learned	211
Summary	211
References	212
Standards	212
Internet Resources	212

Chapter 10	On-Demand Services	215
	Interactive Versus On-Demand Services	216
	Internet Comparisons	217
	Goals of On-Demand Services	218
	On-Demand Services	219
	Movies-on-Demand	219
	Music-on-Demand	219
	Post-Broadcast On-Demand	219
	Special Interest Programming	220
	Distance Learning	220
	Library Access	220
	Video Mail	220
	On-Demand Reference Architecture	221
	Provisioning Network	222
	Distribution Network	222
	Media Servers	225
	Conditional Access	226
	Server Placement	226
	Switching Matrix	228
	Set-Top	235
	Summary	235
	References	235
Chapter 11	On-Demand Cable System Case Studies	237
	Time Warner Full Service Network	237
	Full Service Network Overview	237
	Basic Star Architecture	238
	Logical ATM Connectivity	240
	MPEG Mapping into ATM	241
	MPEG Delivery from Server to Set-Top	243
	Lessons Learned	245
	FSN Summary	245
	Pegasus Phase 2.0	245
	FSN Learning Experience	246
	On-Demand Services	247
	New Channels	248
	Server Location	249
	Fiber Transport	250
	Transport Protocol	251

	Switching Matrix	251
	Pegasus Phase 2 Summary	253
	Summary	253
	References	253
	Standards	253
	Internet Resources	253
Part III	OpenCable	255
Chapter 12	Why OpenCable?	257
	Goals of OpenCable	258
	New Suppliers	258
	New Services	258
	Retail Availability	258
	Market Forces	259
	Competition	259
	Technology	261
	Government Regulation	261
	1992 Cable Act	261
	1994 FCC Report and Order	262
	1996 Telecommunications Act	262
	1998 Report and Order on Competitive Availability of Navigation Devices	263
	Digital Carriage	264
	Emergency Alert Systems	265
	Retail Issues	265
	Cable Service Management	266
	New Services	266
	Existing Suppliers	267
	Signal Theft	268
	OpenCable Solutions	269
	The OpenCable Process	269
	The OpenCable Architecture	270
	The OpenCable Specifications	271
	Retail Availability	271
	Summary	272
	References	273
	Books	273
	Periodicals	273
	Internet Resources	273

Chapter 13	OpenCable Architectural Model	275
	OpenCable History	276
	OpenCable Process	278
	Specification Development	278
	Interoperability	279
	OpenCable Reference Diagram	280
	Specified Interfaces	281
	Specified Components	286
	Standards	287
	Relevant Standards Bodies	287
	Summary	290
	References	291
	Books	291
	Internet Resources	291
Chapter 14	OpenCable Device Functional Requirements	295
	Goals	296
	Provide for Integrated Service Environments	296
	Open and Interoperable	297
	Portability	297
	Renewable Security	298
	User Interface	298
	Scaleable	299
	Efficient Application and Network Design	299
	Operational Compatibility	299
	Backward Compatibility	299
	Goals for the OpenCable Set-Top Terminal	300
	OpenCable Device Models	300
	Leased	301
	Retail	302
	Core Services	304
	Core Functions and Features	304
	The Cable Network Interface	306
	Conditional Access System	310
	Video and Graphics Processing	310
	Audio Processing	311
	Microprocessor Subsystem	311
	Remote Control	311

Extension Requirements	311
Performance	311
Summary	312
References	312
Book	312
Standards	312
Internet Resources	313
Chapter 15 OpenCable Headend Interfaces	315
OCI-H1	316
Goals	317
Issues	317
Reference Architecture	318
Development Status	323
OCI-H2	323
Goals	323
Issues	324
Reference Architecture	324
Development Status	325
OCI-H3	326
Goals	326
Issues	326
Development Status	327
Summary	327
References	327
Book	327
Papers	328
Standards	328
Internet Resources	328
Chapter 16 OCI-N: The Network Interface	331
Scope of OCI-N	332
Issues	332
The Frequency-Domain View	334
Channel Types	335

	OCI-N Protocol Layering	336
	In-Band Channels	338
	Out-of-Band Channels	341
	Summary	345
	References	346
	Book	346
	Standards	346
	Internet Resources	346
Chapter 17	OCI-C1: The Consumer Interface	349
	Goals	350
	Issues	351
	OCI-C1 Family	352
	Analog NTSC	355
	Component Video	355
	Home Digital Network Interface	355
	Summary	363
	References	363
	Book	363
	Standards	364
	Internet Resources	364
Chapter 18	OCI-C2: The Security Interface	367
	Reference Diagram	368
	Drivers	369
	Regulatory	369
	Retail	370
	Cable Ready Digital Television	370
	High Definition Television	371
	Retail Cable Issues	371
	Retail Opportunities	372
	Summary of Approaches	373
	System Architecture	374
	Conditional Access System	375
	Out-of-Band Communications	375
	Two-Way Operation	376

One-Way Operation	378
DOCSIS Operation	379
The POD Module	380
POD Module Variants	380
POD Module Architecture	381
PCMCIA Compliance	382
MPEG-2 Transport Stream Interface	382
Out-of-Band Interface	382
CPU Interface	384
Content Protection	385
Why Is Content Protection Required?	385
The Content Protection System	385
Content Protection System Integrity	387
POD Interface Copy Protection	387
Host Revocation	388
Applications	389
Digital Set-Top	389
Digital Set-Top with DOCSIS Cable Modem	391
Cable Ready Digital Television	393
Issues for Cable Ready Devices	398
Summary	399
References	399
Standards	399
Internet Resources	400
Glossary	402

Foreword

by Jim Chiddix, Chief Technical Officer, Time Warner Cable

For sixty years, as the television-watching public we have considered television a passive form of sit back-and-relax entertainment. We've asked the question thousands of times: "What's on?" That question is about to become a meaningless phrase. Something very big is happening. It's a culmination of sorts: past lessons; hard work on broadband plant; and the revolutionary spread of a worldwide network that, for the most part, speaks in a unified language. Suddenly, millions of homes have digital set-top boxes. Initially, these boxes used proprietary technology. But OpenCable opens a door for meshing set-top boxes into television sets, DVD players, and tightly integrated consumer devices sold at retail.

The inevitable shift from analog to digital television is a matter of critical mass. In a sense, we're at a precipice, peering into the misty invisibility of software and resultant applications that will transform television into a medium that suits our times.

As we near the end of the century, it's notable to observe how much time is a continuum of constant connectivity. Our living rooms are our offices, and our offices are our living rooms; we work not 8–5, but in bursts from waking to resting, snaring half-hour respites for pockets of leisure. Our television experience will be an extension of that continuum. With video on demand, we'll watch the television shows we want to watch, when we want to watch them. In determining a way to link the Internet resource to television, we won't necessarily log onto URLs but maybe we'll take advantage of a personalized offering that meets—even surpasses—our expectations.

This isn't the first attempt to define and create a television platform that is in sync with the future. In 1993 in Orlando, Florida, we pulled out all the stops to create a first of its kind, truly advanced broadband network. It involved a great deal of unsustainable and expensive technology, but that wasn't the point. We were fortunate enough to afford that opportunity of experimentation, partly to see what worked and what it cost, but mostly to ascertain what customers would do if they could use their televisions to interact with programs, merchants, and communications.

From those tests, we have quantifiable conjectures about what television can be to a millennium society. We have another half-decade of work behind us on making cable plants capable of digital transmissions and expanded capacity. Computing technology costs have plummeted since our FSN, and performance continues to skyrocket; set-tops today are more powerful than the highest-end FSN boxes, at a fraction of the cost.

Another part of the precipice is software. Courtesy of the Internet, a pervasive set of software and networking standards exist. Those standards are the fabric of OpenCable; their ubiquity significantly widens the pool of potential television applications developers. In Orlando, one of the real challenges we faced was the need to pay enormous sums to attract applications developers to the platform, because software had to be done in proprietary languages against a limited customer base.

That's no longer true. With the pervasiveness of MPEG video combined with the largess of OpenCable, the simple truth is that we'll be getting a lot more television. That cries of interactive navigation systems—consistently the item that tops the satisfaction list among digital cable subscribers. A good navigation system also paves the way for VOI and personalized viewing options (e-commerce and individualized advertisements are within reach, too).

All of these concepts are scrutinized in Michael's book. In our swiftly changing environment, this book is the rare example of a single text that will stand the time test for survivability. That's because this book details the OpenCable foundation, its technical underpinnings, and its commercial potential. Certainly, events will occur that extend the discussions contained in these pages. But developments are nebulous without a foundation; this is the foundation.

My endorsement of Michael and this book rests on the simple fact that he is an expert. He's been in the thick of OpenCable since its earliest days, saw us through the "first" interactive television days, and owns an equally rich understanding of set-top architecture.

Thank you, Michael, for documenting this place in the development of television.

Introduction

The cable industry is currently in the midst of a revolution. Existing systems that were engineered for broadcast television are being called upon to support a host of new applications and services. The race is on to upgrade cable systems to high-speed, two-way communications networks while continuing to support plain old television services.

A lot of consumers, industry analysts, and policy observers are asking “Where is digital TV headed?”, “Is it for real?”, “Will I be able to get it on cable?”, and “What new services will be offered?”

These questions are difficult to answer because they are dependent on so many factors—including market forces, technical realities, investment levels, and learning curves. Nevertheless, here are some safe bets:

- **Competition**—Competition is here to stay. Analog cable systems have to add digital services to compete with direct broadcast satellite, digital terrestrial broadcasting, and even the Internet.
- **Retail Availability**—Government regulation mandates the competitive retail availability of navigation devices. The first digital cable-ready devices will start to appear on the market toward the end of 2000, soon after the July 1st FCC deadline requiring cable operators to make separable security modules available.
- **Convergence**—Convergence, which is the fusion of digital television, data communications and personal computing technologies, will fundamentally change the way we watch television, surf the Internet, and communicate with each other.

OpenCable is the cable industry's response to these market forces. OpenCable is an initiative lead by Cable Television Laboratories (CableLabs) on behalf of the cable operators. OpenCable seeks to set a common set of requirements for set-top equipment so that new suppliers from the consumer electronics and computer industries can start to build equipment for connection to cable systems. This book takes a first attempt at discussing OpenCable issues and progress.

Purpose of this Book

The OpenCable initiative started with an effort by cable operators to find alternative, lower-cost sources for digital set-top converters. The computer and consumer electronics industries have demonstrated an impressive ability to reduce the cost of almost any type of electronic equipment (PCs, laptops, CD players, televisions, VCRs, satellite receivers, DVD players, and so on). Moreover, this reduction in cost has been combined with the rapid introduction of new features and increased levels of performance. However, it quickly became apparent that the computer and consumer electronics industries need a blueprint in order to build a digital cable set-top.

OpenCable is a set of functional requirements and interface specifications that provides this blueprint. This book examines in detail the new architectures being developed by the cable industry as part of the OpenCable initiative. It also reviews the development of digital cable television systems and interactive television services because OpenCable draws on these roots.

The goals of this book are:

- To provide a comprehensive and practical overview of digital cable television systems. This book describes the headend, optical transport, distribution hub, hybrid-fiber coax, and set-top terminal equipment and how these components are interconnected. These topics include some unique aspects of cable systems that are not generally known outside of the cable industry due to their limited publication and recent development.
- To summarize the important issues in digital cable television. This book addresses the competitive, regulatory and technical challenges associated with the introduction of digital cable television services.

- To review the recent developments in interactive television, including the Time Warner Full Service Network (FSN), and to show how the concepts have evolved, since FSN's introduction in 1994, through the development of Time Warner Cable's Pegasus digital program.
- To explain the OpenCable initiative. This book covers the market forces driving the OpenCable initiative and provides a detailed technical analysis of the OpenCable Architecture.

This book is not intended to be a comprehensive engineering reference, but it does offer a roadmap to, and some interpretation of, the myriad specifications for the serious researcher into cable systems architecture.

Nor is this a book on analog cable systems. Where appropriate, I have referred the reader to specific sections of the excellent text *Modern Cable Television Technology; Video, Voice, and Data Communications*, by Ciciora, Farmer, and Large, for answers to most questions on analog cable.

In stating both positive and negative issues surrounding digital cable television (in general) and OpenCable (in particular), it is not my intention to promote or criticize any industry. During my tenure at Time Warner Cable, I have come to know and admire representatives of many companies and organizations involved in all aspects of digital television. I have always been impressed with the effort, brainpower, and public spiritedness of these professionals in business and government, who will no doubt change the way we all receive and respond to information in our homes.

Audience

This book is for the following audiences:

- Anyone who would like to learn more about digital cable systems. If you need a broad-based familiarity with the recent developments in digital television, this book provides an accessible introductory text whether you are a business development manager, industry analyst, legal counsel, or regulator.
- Engineers and technical managers who are already involved in the cable industry and would like to learn more about the digital technology that is enhancing (and, in some cases, replacing) familiar analog techniques.
- Application developers who want to deploy their skills in developing new services for digital cable systems.
- Network developers who are being called upon to build broadband networks that seamlessly interleave video, audio, and data communications techniques.

My overall goal in writing this book is to provide a networking perspective on digital cable systems. I feel it is important to give the business development, marketing, and applications development teams a good background in the network architecture that is so fundamental to digital cable systems. This fact, more than any other, has prompted me to pick up my word processor and write. I hope this book will introduce these concepts to a wider audience and act as a valuable reference to my colleagues in the cable industry.

Structure

The body of this book is divided into three main parts:

- **Part I—Digital Cable Television.** Chapter 1 presents, in high-level terms, the nature of the digital revolution and its impact on television. Chapter 2 provides a brief overview of analog cable technology and the hybrid fiber coax (HFC) upgrade path. Chapter 3 introduces the advanced analog set-top converter. Chapter 4 explains some of the key digital technologies that will be referred to in the rest of the book. Chapter 5 explains how digital television services are added to existing analog cable systems. Chapter 6 describes the hardware and software characteristics of a digital set-top. Chapter 7 contains case studies of the two predominant choices for digital broadcast systems available to North American cable operators: General Instruments DigiCable system and Scientific Atlanta's Digital Broadband Delivery System (DBDS).

- **Part II—Interactive and On-Demand Services.** Chapter 8 discusses the development of interactive services for cable systems. Chapter 10 introduces on-demand services, which provide a dedicated multi-media stream to the customer. Chapter 9 and 11 provide case studies of interactive and on-demand services respectively from the Time Warner Full Service Network and the Time Warner Pegasus program.
- **Part III—OpenCable Architecture.** Chapter 12 introduces the market drivers for OpenCable. Chapter 13 describes the OpenCable Architectural model. Chapter 14 explains the functional requirements for an OpenCable device, for example a digital set-top or a digital cable ready receiver. Chapter 15 introduces the OpenCable headend interfaces (OCI-H1, OCI-H2, and OCI-H3). Chapter 16 describes the OpenCable Network Interface (OCI-N). Chapter 17 describes the Consumer Interface (OCI-C1), which connects the OpenCable device to the consumer devices in the home. Chapter 18 describes the interface to a replaceable security module (OCI-C2).

Features and Text Conventions

The following text design and content features used in this book are intended to help understanding of the subject matter:

- Key terms are italicized the first time they are used and defined. Similarly, all abbreviations are expanded the first time they are used.
- Bullet points are used to outline the key topic areas at the start of each major section. Section sub-headings use consistent terminology to ease quick reference to specific information.
- Chapter summaries placed at the end of each chapter provide a brief recapitulation of the subject matter covered in that chapter.
- References to further information are included at the end of every chapter.

Timeliness

While writing this book, I made numerous changes and additions to reflect the most recent standards developments, FCC directives, and technical innovations. In this rapidly evolving field, some changes will no doubt have occurred by the time you read this book. Nevertheless, the material contained in this book will provide you with a solid foundation for understanding the complex issues surrounding this intriguing subject.

Digital Cable Television

- Chapter 1 Why Digital Television?
- Chapter 2 Analog Cable Technologies
- Chapter 3 The Analog Set-Top Converter
- Chapter 4 Digital Technologies
- Chapter 5 Adding Digital Television Services to Cable Systems
- Chapter 6 The Digital Set-Top Converter
- Chapter 7 Digital Broadcast Case Studies

Why Digital Television?

Television pictures live in an analog world—with infinite possibilities of hue and color—which is exactly what the human eye requires for a realistic and satisfying viewing experience. What's more, television as we know it is now 60 years old and enjoys nearly 100% penetration in 100 million U.S. households.

Given that, does the “if it ain't broke, don't fix it” axiom apply? Not exactly. Although the shift to digital television will require serious engineering attention and capital expenditures, the resultant product will add greatly to quality and new service opportunities. How and why did digital television evolve and what is driving digital television into our homes?

Digital Technology Evolution

Digital technology has fascinated engineers since the first digital computers were developed. It wasn't just the fascination among engineers for the next new gadget. Well, there was some of that. But, at the same time, digital techniques represented a more efficient way of doing what had already been done with vacuum tubes (or “valves”). Ultimately, transistor technology replaced vacuum tubes because the high density and low-power dissipation characteristics of semiconductor junctions were ideal for fabricating the complex digital circuits needed to build computers.

In lockstep with the rapid evolution of digital technology, engineers continued to push the envelope in applying digital technology to every conceivable problem. Consequently, the digital push has become something of a Holy Grail in engineering. In fact, there is currently so little interest in analog electronics that it is sometimes necessary to tempt analog-savvy engineers out of retirement because so few younger engineers are entering the field with any interest or experience in analog techniques.

Silicon Integration and Moore's Law

The fuel-propelling digital technology clearly came from the computer industry, which has spent the last fifty years in a race to build faster computers. In practical application, this means reducing the physical dimensions of transistor junctions as a way to increase the clock frequency of a circuit—without increasing power dissipation. The semiconductor segment leapt to action, finding ways to shrink transistor sizes so that more of them could squeeze onto the silicon die of an integrated circuit.

For example, early transistor-transistor logic (TTL) chips contained between about 30 and 100 transistors—a huge technical breakthrough at the time. TTL was quickly replaced with complementary metal-oxide semiconductor (CMOS). With the reduced power dissipation of CMOS, larger integrated circuits were built. Soon, an entire computer processor could be placed on a single chip. Enter the microprocessor.

This phenomenon gave rise to one of the most frequently cited harbingers of technological change: Moore's Law. Hardly a technical presentation goes by—regardless of industry segment—without someone mentioning Moore's Law in the context of startlingly swift technological growth.

Moore's Law stems from Intel chairman emeritus Gordon Moore, who observed that the number of transistors on a chip was doubling every 18 months. When he confirmed this trend, it was dubbed Moore's Law, an axiom that continues with no end in sight until perhaps 2030. (For more information on Moore's Law, see <http://www.intel.com/pressroom/kits/bios/moore.htm>.) Millions of transistors are now routinely placed on a silicon die, and many chips are now I/O-limited, which means that the cost of the chip has more to do with the number of leads and the packaging cost than the number of transistors it contains.

Very large scale integration (VLSI) encourages the designer to place as many functions as possible on a single chip. The ultimate goal: a single chip that performs all the functions of a product (whether it is a television receiver, a set-top, or a personal computer). Because it is tricky to mix analog and digital functions on a chip, it makes sense to do all possible functions in the digital domain. For example, relatively complicated digital circuits are replacing even trivial analog functions, such as audio mixing.

Analog-to-Digital Conversion

When it was realized (in the 1970s) that almost all analog processing could be done with more precision and much greater flexibility in the digital domain, the race was on to shift more analog functions to digital.

The first step in this process is called *analog-to-digital (A/D) conversion*. The analog signal is sampled (measured in time close enough together to adequately represent the analog signal) and its instantaneous value is represented as a binary value. After A/D conversion, most analog signal processing can be done in the digital domain. This technique is known as *digital signal processing (DSP)*. After signal processing, a process known as *digital-to-analog (D/A) conversion* reconstructs the (modified) analog signal.

Early DSP applications had an analog input and an analog output, but soon the digital representation became the reference signal that was stored or transmitted. An early user of these techniques was the music industry, which embraced digital techniques so roundly, it is nearly impossible to purchase cassette tapes and vinyl records today. Their successor, the compact disc, was introduced in the 1980s and harnessed Moore's Law as a way to

dramatically improve sound quality and the amount of music stored per CD (relative to analog predecessors).

By the 1990s, digital techniques had evolved (thanks in part to Moore's law) to tackle the hundredfold increase in bandwidth of video (compared to audio).

A/D conversion and DSP are now cost-effective tools for television services and have found their way into most of the technologies described in Chapter 4, "Digital Technologies." This trend shows no sign of slowing down and continues to drive the migration to digital television.

Convergence with the Personal Computer

Moore's Law also made the development of personal computers (PCs) practical. Early PCs were very limited in performance and memory—remember when 4 MB of RAM was a big deal? But reductions in price and quantum leaps in performance combined to create a multibillion dollar industry around the PC. Standalone PCs remain somewhat limited in what they can do; all applications must be loaded from stored media, and it is still somewhat slow and cumbersome to share data with other PC users. Still, PC networking is transforming the PC; it is now possible to pipe in applications and data from the Internet and to use the PC as a communications tool. The development of standard protocols to support World Wide Web services also introduced a new mode for research and entertainment.

PCs are now powerful enough to perform sophisticated multimedia processing (using digital signal processing). Suddenly, *convergence* has reemerged as a buzzword to describe the personal computer as the focus of entertainment, computing, and communications services in the home.

In the home, the notion of convergence will also create a *divergence* of in-home electronics, where the swift impact of Moore's Law creates customized, inexpensive chip-sets that can be installed in many communications and entertainment gadgets. The cable modem gained wide U.S. acceptance in 1999; industry analysis firm Paul Kagan Associates (PKA) anticipates that 1.6 million cable subscribers will use a cable modem, at \$40 per month, to link to the Internet at high speed (27 Mbps shared over a node versus 56 Kbps via dial-up). That figure could leap to 20 million subscribers by 2005, according to PKA (including other broadband connectivity devices, such as DSL and wireless modems).

By the end of 2000, as many as three million advanced digital set-tops (that include a cable modem) will populate U.S. homes. Add to that DVD players, personal organizers, and boxes such as those made by TiVo and Replay that enable truly on-demand television viewing.

Convergence has many faces, but it is really just the parallel application of evolving digital technologies across different fields. The technology of the Internet and the World Wide Web are already finding their way into advanced analog set-top converters (see Chapter 3, "The

Analog Set-Top Converter”). This means that services previously experienced only with a PC can now be experienced on the family television (with a wireless keyboard). This trend will continue at a faster pace and is certainly helping to drive us into a digital world, although it has yet to be seen which platform will be dominant in the home.

Internet Convergence

Over a ten-month period—since I began the research for this book—the Internet gestated from a powerful communications tool looking for a way to grow up to a multibillion dollar industry. The Internet has become a useful means for such tools as e-mail and research/fact-finding, which is now an inexorable part of the lives of nearly 50 million users.

It also stands to lighten the wallets of U.S. consumers by billions of dollars over the coming years, through electronic commerce—a phenomenon that saw early adopter usage during the Christmas 1998 season, but will almost certainly attract the masses going forward. This is largely because the Internet uses a technological language, known as Internet Protocol (IP), that will undoubtedly be the fulcrum that pries open a cornucopia of advanced broadband services.

This isn't tomfoolery, and its impact on digital video applications is dramatic. The belief in heady revenues from providing Web-based entertainment services led to the concept of an Internet-based entertainment device, such as Microsoft's WebTV, which surpassed the one million-user mark in 1999. OpenCable boxes will be similarly outfitted to deliver Web-based information, interactivity, and e-commerce to cable subscribers. If the trend line toward a “free” Internet continues, service providers will be all the more inclined to seek revenue sources, such as e-commerce and advertising, outside of subscription fees.

Obviously the Internet and the World Wide Web are cool showcases for the potential for digital entertainment services, and it would be foolish to ignore them. Very foolish! Building television systems with digital technology is a smart investment decision considering this potential. Of course, building digital into television is just one piece of the puzzle. Service providers also need to link their digital products and services into existing—and new—back-office systems, such as provisioning and customer service. Again, many of these solutions exist for the PC-oriented Internet. Just as it was clever to repurpose existing digital techniques for use in television, another smart investment is to identify, repurpose, and use the many back-end Internet software management packages for content management, network management, and so on.

New Services

The gradual process of blending entertainment with communications has already started and will continue; Internet sites will continue folding in digital video clips, and broadcast and cable providers will continue folding in Internet content to their programs. When you click on an Internet site and browse from business news to people news to movie clips, you

are experiencing this today. New services will take further steps in this direction and will use a combination of digital television technology with enhancements based on high-speed communications technologies that use the increased bandwidth of cable systems.

It is possible to enhance analog television services; a small handful of interactive television players has been pursuing the category since the mid-1990s. So far, the techniques have seen little endorsement, mostly because the services are somewhat awkward to use and limited in scope. The possibilities for enhancing digital television are much greater, and there is a cleaner integration path because everything is in the digital domain. Examples are interactive services (see Chapter 8, "Interactive Services") and on-demand services (see Chapter 10, "On-Demand Services").

New Business Models

New services enable new business models; there are many new business models that are currently being tested, such as

- Video-on-demand (VOD)—The first VOD business model was published last year (www.schange.com). The reduction in media server cost combined with the development of software to support VOD services has triggered the deployment of commercial VOD systems.
- Linked advertising—Linked advertising gives customers the ability to customize how they receive advertisements over television. For example, they can skip them altogether (perhaps for a fee—we'll see how that model evolves!), request more information (such as a coupon), or specify areas for which they would like to see ads (cars, household products, financial, and so on).
- Linked merchandising—This model lets customers buy products that are shown in a program (for example, the hero's leather jacket or the heroine's little black dress) and provides discounts for services with the purchase of entertainment services (for example, a discount pizza with a movie purchase).
- Linked communications—Examples include linking a chat session to a live television broadcast, allowing the customer to play along with a television game show, and so on.

Of course, nobody knows whether these or other business models will unlock new revenue streams. A persistent sociological argument exists throughout the category about whether people want to *lean back* and let television entertain them, or whether most people prefer to do heavy interactivity on their *lean-forward* PC. There are also issues around family viewing—will Mom and Dad sit by patiently while the kids click off to who-knows-where during a television broadcast? and vice versa?

Nonetheless, most of these models are enabled by the evolution to digital television. Moreover, a software-based approach can be flexible enough to support these and many other potential business models. In other words, software applications can be developed and deployed independently of the hardware infrastructure, allowing many incremental business opportunities.

Advantages of Digital Television

Digital technology brings a number of concrete advantages when applied to television services. The most important of these are

- **Channel expansion**—Digital compression allows the cable operator to expand the number of channel offerings by a factor of 10.
- **On-demand services**—Digital compression also provides bandwidth economies that make video-on-demand commercially viable.
- **Quality**—Although there is a carefully controlled *loss* of quality inherent in the digital compression process, further transmission losses in quality due to noise and distortion can be eliminated in a well-designed digital system.
- **Security**—Digital scrambling, or *encryption*, provides a very secure means to enforce conditional access (where only people who pay for the service get the service) with no loss of quality.
- **Flexibility**—Digital compression formats can be changed instantaneously to suit the program material, which means that multiple video and audio streams are supported.
- **Data transmission**—Digital transmission systems are agnostic about the nature of the payload they carry and, as such, contain a multiplex of video, audio, and data. This allows the transmission of any kind of data, which may be used for program enhancement, program downloading, Web-casting, and many other applications.

Channel Expansion

Channel expansion adds more programming choices to a service tier. Channel expansion is not a new concept, but cable operators, terrestrial broadcasters, and satellite operators were historically limited by spectrum availability. For many years, cable operators expanded their channel offerings by upgrading their systems to carry more and more analog channels. Meanwhile, terrestrial broadcasters were confined to channels allocated by the FCC. Satellite operators added more channel capacity by launching more satellites—at considerable expense.

Digital compression allows approximately a 10:1 expansion in the number of channels for a given amount of spectrum. In fact, it was the development of digital compression techniques that made direct satellite broadcasting an economically palatable proposition, giving rise to such companies as DirecTV and EchoStar.

On-Demand Services

Digital compression is an essential element of on-demand services because it reduces the bandwidth occupied by each channel by a factor of 10. On-demand services allocate a dedicated digital program stream (or channel) to each active user. Chapter 11, "On-Demand Cable System Case Studies," discusses on-demand services in detail.

Quality

It is difficult to maintain the quality of an analog television picture from its origination at the television camera until its final appearance at the television receiver. Each time it is amplified, copied, or transmitted, noise and distortion creep into the signal. In the studio, these effects can be minimized by the careful use of high-quality equipment. However, the path to the customer's home is more difficult to maintain. As a result, some noise and distortion mar most analog television pictures, in the form of *snowy* and *ghosted* images.

Analog-to-digital conversion transforms the analog signal from the television camera into a digital stream—before noise and distortion have a chance to impair the signal. After the signal is digitized, it can be copied perfectly as many times as required. Assuming the cable system is properly maintained, only noise and distortion between the D/A converter and the display itself can degrade the picture. As a result, digital television pictures are almost completely free of noise and distortion. (See Chapter 4 for a discussion of some of the potential drawbacks of digital compression.)

Security

To collect payment for premium television services—the most common examples are HBO and Showtime—it is necessary to prevent unauthorized viewers from watching the programs. In analog television, scrambling is used, but this analog transformation of the signal introduces distortion. It is also relatively easy to defeat analog scrambling (see Chapter 3). In 1999, the cable industry could lose more than \$5 billion to unauthorized access of premium cable signals.

A digital stream, by contrast, is a string of ones and zeros. As such, the science of cryptography can be applied to encrypt the bit stream into literally trillions of different permutations. Without the correct *key*, deciphering the encrypted stream is like searching for a needle in a haystack the size of New York. With the correct key, however, the original bit stream can be recovered with perfect accuracy and no loss in quality is incurred.

Flexibility

Digital techniques are very flexible, in part because they decouple the encoding process from the transmission process. The encoding process can be changed to sample the picture with higher resolution or with a faster refresh rate, and the only impact is a change in the bit rate of the digital stream. The transmission process accepts any digital stream as long as the bit rate does not exceed the transmission rate. Thus, it is possible to send 4 video streams and 19 audio streams, if that is desired, as long as there is adequate transmission capacity. If a very high-quality video stream is required, it can be designed to fill almost the entire transmission payload. In fact, that is the basis for high definition television.

Data Transmission

The transmission process used for digital television is completely agnostic of the payload. The bits might be used to send video, audio, or data with equal ease. Thus, it is possible to add data to a television program. (This is also possible with analog television using the vertical blanking interval [VBI] lines, but the data is sent as a series of white or black dots and the payload is limited to about 19 Kbps on each VBI line.) Using digital transmission techniques, much higher data rates are possible, in the order of megabits per second.

Summary

This chapter presented a brief overview of some of the history and technical advantages that favor the transition of television from analog to digital. The concrete advantages of digital technology are

- Channel expansion
- On-demand services
- Quality
- Security
- Flexibility
- Data transmission

However, there are other, less measurable, aspects of digital television that include

- Industry trends toward a digital approach based on very large scale integration of silicon chips and digital signal processing
- Convergence with the computer and communications industries and the influence of the Internet and the World Wide Web
- Potential new service and revenue opportunities that are enabled by digital technology

Analog Cable Technologies

Cable systems have evolved rapidly since 1990 through the use of fiber-optic technology. Until this time, cable systems were designed to support broadcast of the same channel lineup throughout the cable system and suffered from limitations in frequency range and signal quality. With the introduction of fiber optics into cable distribution networks, the characteristics of the cable system are changed dramatically:

- The frequency range is increased to 870 MHz (or higher), expanding the analog channel lineup and creating an additional spectrum for digital channels.
- The signal quality is greatly improved by reducing RF amplifier cascades, minimizing noise and distortion.
- The cable system is effectively segmented into many parallel distribution networks. Although the broadcast signals are split and fed into every distribution network, the segmentation enables narrowcast services, such as multiple commercial insertion zones and video-on-demand.
- The cable system is much more reliable and easier to maintain because it incorporates physical link redundancy and because of the smaller serving area of the node.

There are many excellent references that describe the coaxial and fiber-optic technologies in detail (see *Modern Cable Television Technology; Video, Voice, and Data Communications* by Walter Ciciora and others); this chapter briefly addresses the implications for network architecture of these technologies.

Analog Channel Expansion

Fiber-optic distribution allows the cable operator to expand the analog channel lineup and to create an additional spectrum for digital channels, allowing a gradual migration from analog to digital services. This provides advantages to both the cable operator and the customer:

- Analog revenues continue to fund the core business of the cable television industry during the migration to digital services.
- Customers can still use their existing television receivers. Practically all television receivers available in the retail market support *only* the NTSC analog standard, and these receivers have a long projected life expectancy of 10 to 15 years.

- Multiple television receivers can be supported at no extra cost to the customer for unscrambled analog services. Additional televisions do not require a set-top converter to receive *basic* (or unscrambled) analog services.
- Cable operators can continue to provide analog cable services well into the twenty-first century, well beyond the proposed sunset date of 2007 for analog terrestrial broadcast.

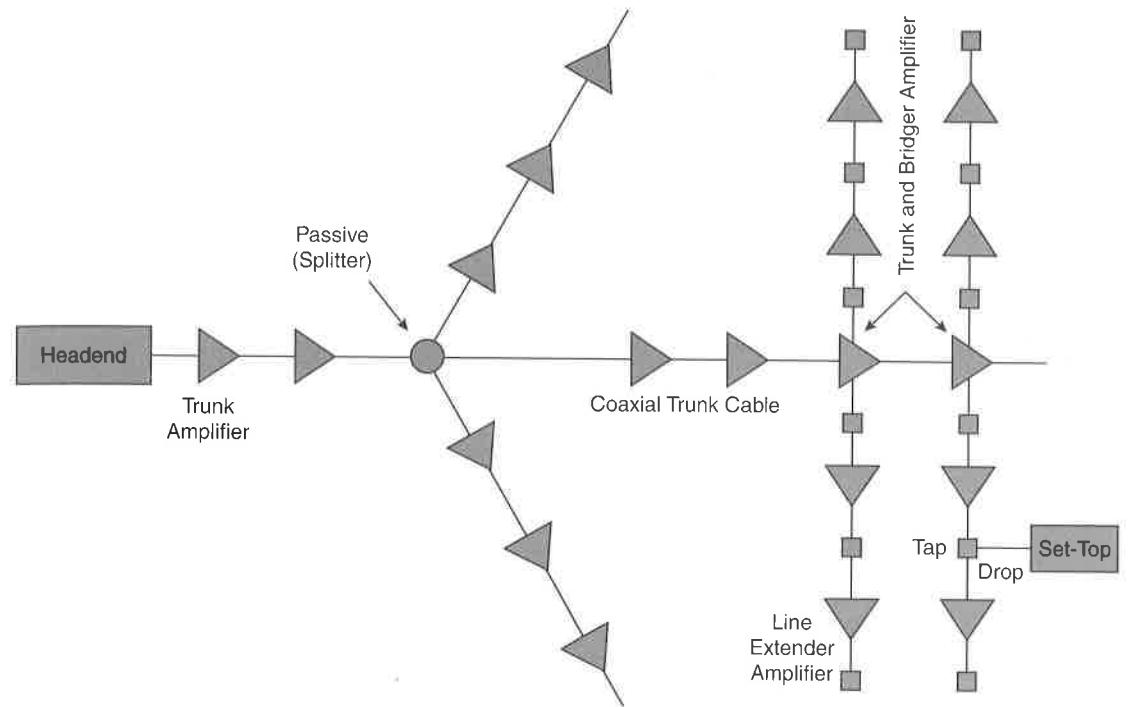
The primary reason for a plant upgrade is to improve signal quality and to reduce system outages. However, an important requirement for any upgrade of the cable distribution network is to support and expand analog services. With this in mind, how can the upgrade also move the cable industry forward toward the deployment of advanced services, such as digital broadcast, interactive services, and on-demand services?

The Hybrid Fiber Coax Upgrade

The upgrade path for nearly all cable operators is based on hybrid fiber coax (HFC) architecture. The HFC architecture is attractive to the cable operator because it improves the analog performance of the cable plant, increases the number of channels, and provides a migration path toward the support of digital services. The HFC upgrade also makes the cable system easier to maintain and reduces system outages, improving overall service to the customer.

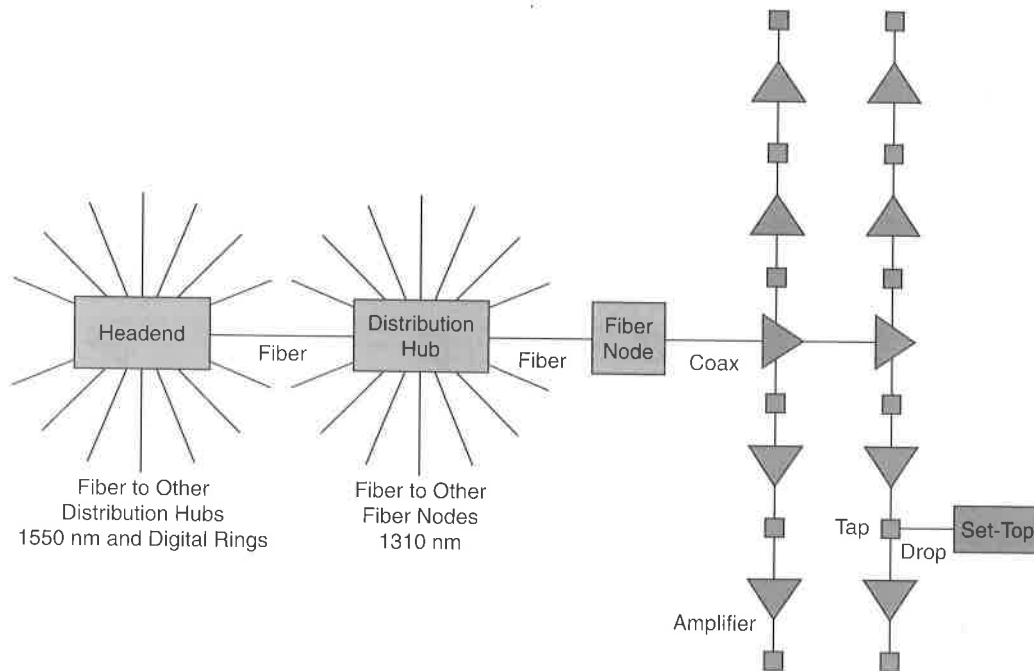
Figure 2-1 shows part of a cable distribution network before the fiber upgrade. The trunk amplifiers are optimized to compensate for the signal loss introduced by coaxial cable over the considerable distance from the headend to the furthest customers, with the lowest possible distortion and noise. Even so, the cascade of 24 or more trunk amplifiers required to reach the serving area introduces significant distortion and noise.

The trunk fans out at convenient geographical points in the network to form a tree-and-branch structure. Figure 2-1 shows one such branch point, where a passive device (such as a splitter or a directional coupler) directs the signal to feed additional trunks. In this way, the trunk fans out to feed many thousands of homes from the single cable that leaves the headend. However, the branching trunk cable makes up only about 12% of the total cable length in a typical cable system; the role of the trunk is to feed the *serving areas*, which contain the other 88% of the cable length.

Figure 2-1 Coaxial Cable Distribution Network

At the serving area, the distribution network fans out rapidly from the trunk by means of bridger amplifiers and line extender amplifiers to the customer homes. This final section of the distribution network is commonly known as the *last mile* because it is usually within one mile of the customer. Finally, a portion of the signal is coupled to a *drop cable* at the *tap*. The drop cable runs from the tap to the set-top in the customer's home.

In Figure 2-2, the cascade of trunk amplifiers is replaced by a fiber-optic link. The RF signal modulates the optical signal at the headend and is detected and converted back to RF at the fiber node. (In many systems, the signal is boosted at the distribution hub using optical amplification, or by conversion to an RF signal followed by conversion back to an optical signal.)

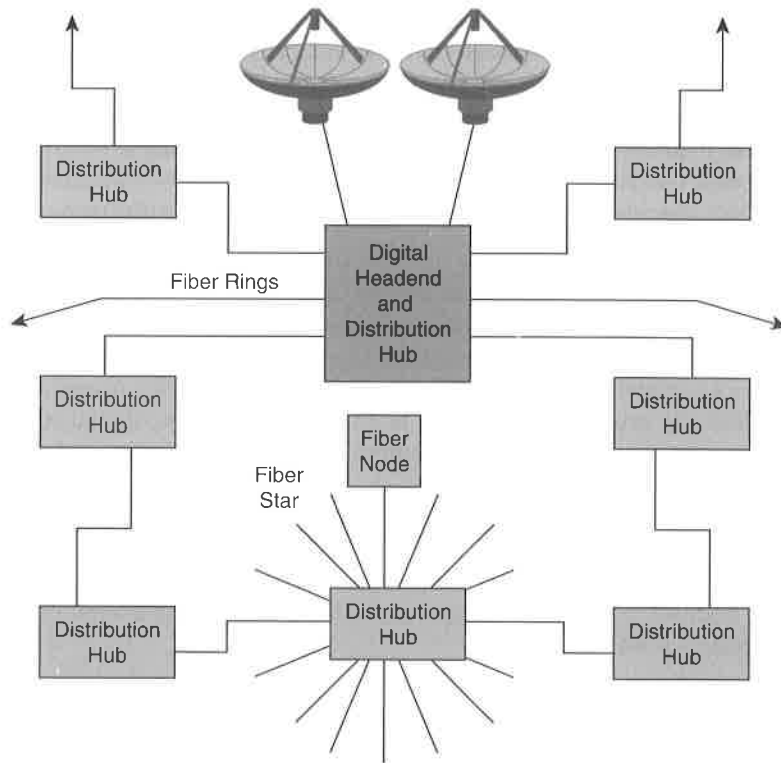
Figure 2-2 HFC Distribution Network

The performance of the optical link is superior to that of the coaxial link because of the reduction of amplifier actives and the isolation of system outages. The service and signal quality to the customer is improved. More importantly, the distribution network is segmented into a large number of parallel distribution networks, allowing the reuse of the RF spectrum (narrowcasting).

HFC Topology

Most hybrid fiber coaxial networks are based on the three-level hierarchy shown in Figure 2-3, which includes a headend (origination of content), a distribution hub (typically 20,000 homes passed), and a fiber node (typically 500 homes passed). The fiber node converts the optical signal into RF for the last mile distribution to the set-top. Fiber nodes are usually connected to a distribution hub in a star network to reduce cost. A star network is acceptable because a fiber cut affects only one fiber node and a very limited number of customers.

Figure 2-3 HFC Topology



The distribution hub is a different matter entirely—it supplies a large number of fiber nodes, and the consequences of a loss of signal are far greater. Distribution hubs are usually fed by two fiber spans that are *diverse-routed*—that is, they take a different physical route. A convenient way to achieve diverse-routing is by means of a ring structure (see Figure 2-3). Typically, four to six distribution hubs are placed on a ring. A distribution hub might feed a very large number of fiber nodes, depending mainly on the customer density; 40 is an average number of fiber nodes per distribution hub.

At the top level of the hierarchy is the headend. Headend facilities are similar in many ways to distribution hubs, but the headend also is a satellite receiver site. In most cases, the headend site also serves as a distribution hub for the fiber nodes closest to the headend.

Linear Optical Transmission

Hybrid fiber coax networks are based on linear optical transmission, which uses *amplitude modulation* of an optical carrier. This is in contrast to digital (on-off) modulation of an optical carrier, as specified in the Synchronous Optical Network (SONET) standard. Linear optical transmission has been developed and refined by the cable industry to meet its specific requirements; the most important of these is reducing distortion and noise while increasing optical output power. Linear long-reach optics allow the cable industry to reduce amplifier cascades, improving the quality of existing analog distribution systems.

Optical fiber is an ideal medium for cable transmission because its bandwidth is virtually unlimited (NHK [the Japan Broadcasting Corporation] has demonstrated transmission at 3,000 *gigabits* per second over a single fiber) and because the attenuation of light in a fiber is low (at 1550 nanometers, attenuation is about 0.25 dB per kilometer and it is 0.34 dB per kilometer at 1310 nm). The use of digital transmission in analog cable systems is prohibitively expensive because of the expense of analog-to-digital conversion. Instead, the intensity (or amplitude) of the output of a laser is modulated to build an end-to-end analog link.

For successful analog modulation of an optical carrier, it must have certain characteristics:

- The linewidth, or optical purity of the carrier, must be narrow. The distributed feedback (DFB) laser incorporates a diffraction grating to limit the oscillation to a single wavelength.
- The linear region of the DFB laser is used where the light output is proportional to the modulation current. Likewise, the linear region of the optical detector must be used.
- Noise sources must be kept to a minimum at the laser and at the detector.

For these reasons, adjustment of analog-modulated optical links is more complicated than their digitally modulated counterparts. Furthermore, degradation of an analog link is more visible than that of a digital link. Nevertheless, the cable industry has deployed HFC upgrades widely with excellent results [Chiddix].

Return Path Activation

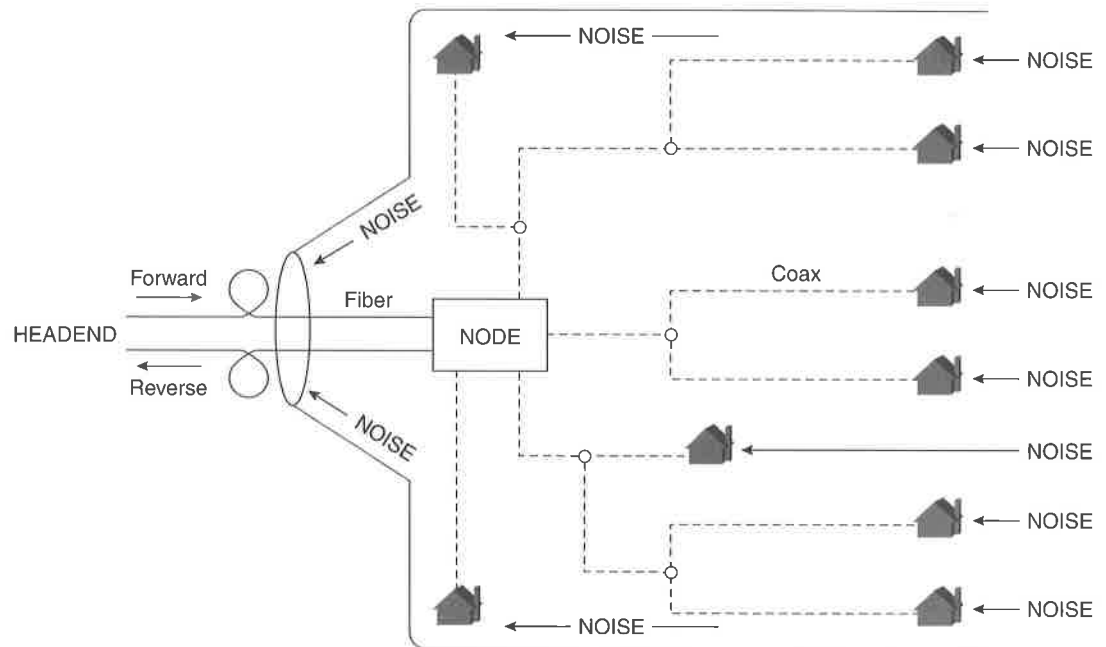
Two-way operation of the cable system plant is essential to support advanced services, such as impulse pay-per-view, interactive, on-demand, and cable modem. Although activation of the return path is possible in a coaxial-only cable system, the HFC upgrade makes it much more practical because it segments the system into small groups of homes passed.

In the coaxial portion of the network, the signal is amplified in both the forward (toward the customer) and reverse directions. This is achieved over a single cable by partitioning based on frequency spectrum. In the forward direction, frequencies from 54 to 870 MHz are amplified, and a high-pass filter attenuates frequencies below 54 MHz. In the reverse direction, frequencies from 5 to 40 MHz are amplified, and a low-pass filter attenuates

frequencies above 40 MHz. The combination of a high-pass filter in the forward direction and a low-pass filter in the reverse direction is called a *diplex filter*.

In the reverse direction, the coaxial system is just like a giant antenna that aggregates the signal from every customer (see Figure 2-4). Unfortunately, this arrangement also aggregates *noise* and *ingress* signals—noise that is generated in the return amplifiers and ingress that is picked up due to problems, such as loose connectors or perhaps poorly shielded equipment connected to the cable in the home. If the return signals from too many homes are aggregated, the noise and ingress will swamp the desired signal from the set-top or cable modem. HFC divides the coaxial system into small groups of about 500 homes passed and limits the amount of noise and ingress aggregation.

Figure 2-4 *Noise Funneling*



Another factor is the relatively small amount of usable bandwidth in the reverse direction. The frequency range is limited from between about 5 MHz to 40 MHz, which limits the number of return carriers that can be allocated. To make matters worse, certain portions of the spectrum might be unusable because of strong radio interference pickup from citizen's band radios and other EMI sources that operate and radiate in the return frequency band. Fortunately, the return bandwidth required by 500 homes passed is relatively small and can be accommodated in the remaining part of the return spectrum.

At the fiber node, the return signal modulates a return laser, and a separate fiber is used to carry the signal back to the distribution hub. At the distribution hub, the return signal is converted back into an RF signal and fed to a demodulator. The demodulator converts the signal back into the digital bit-stream that was transmitted by the set-top or cable modem. Often, a digital fiber-optic transmission system provides the final leg of the return system path from the distribution hub to the headend (see Chapter 5, "Adding Digital Television Services to Cable Systems").

Summary

This chapter provided a brief overview of analog cable technology and the hybrid fiber coax upgrade path that has been adopted by cable operators. The HFC architecture not only provides a number of improvements in performance and reliability for analog cable systems but also paves the way for the introduction of digital services.

The HFC upgrade effectively segments the cable system into a large number of parallel distribution networks. This architecture is supported by a three-level topology of headend, distribution hubs, and fiber nodes. The segmentation of the distribution network enables narrowcasting for on-demand and cable modem services.

The HFC upgrade also enables return path activation by subdividing the cable system into small groups of homes passed. The problems of noise and ingress are reduced, and the limited return bandwidth is reused by each subnetwork.

References

Baldwin, Thomas F. and D. Stevens McVoy. *Cable Communication*. Upper-Saddle River, NJ: Prentice-Hall, 1983.

Ciciora, Walter, James Farmer, and David Large. *Modern Cable Television Technology; Video, Voice, and Data Communications*. San Francisco, CA: Morgan Kaufmann Publishers, Inc., 1999.

Chiddix, James A. "Time Warner Cable's Development of Broadband Optical Fiber Technology for the Cable Television/Telecommunications Industry." Time Warner Cable. October 4, 1994 (see technical papers at <http://www.pathfinder.com/corp/twcable/>).

The Analog Set-Top Converter

The set-top converter is probably the most talked-about element in the cable system, and it is where a great deal of the capital budget is spent. The set-top converter is variously hated, tolerated, and accepted by cable customers. It has been the focus of much cable legislation and regulation.

Initially, set-top converters were deployed to give access to more channels because early televisions couldn't tune most of the frequencies used by cable systems. The set-top converter also gave the subscriber desirable features, such as remote control, volume control, and channel number display. These were appreciated by customers until cable ready televisions made these features redundant and set-top converters began to get in the way of features on the cable ready television. The set-top converter continues to play a role in giving the subscriber new features and services; for example, the advanced analog set-top converter introduces program guides, digital music, and VCR commanders. These features increase customer acceptance of set-top converters even though they might interfere with some of the features of cable ready television (for example, picture-in-picture).

As the number of channels increased and pay channels appeared, access control became more important. Cable systems introduced *traps* (filters), or *interdiction* (jamming) to deny access to premium services. (For more information on program denial techniques in analog cable systems, see *Modern Cable Television Technology; Video, Voice, and Data Communications* by Walter Ciciora and others.) In cable systems that use traps or interdiction, customers do not require a set-top converter if they have a cable ready television. Other cable customers purchase only basic services and do not require a set-top converter.

Nonetheless, most cable customers in North America are familiar with a set-top converter that is leased to them by their cable operator. Analog scrambling has been widely deployed by cable systems because traps are too easy to defeat and interdiction is extremely expensive. In cable systems that use analog scrambling, a set-top converter provides the only alternative to access pay services.

Analog set-top converters have evolved considerably over the past 20 years and can be divided into four major categories:

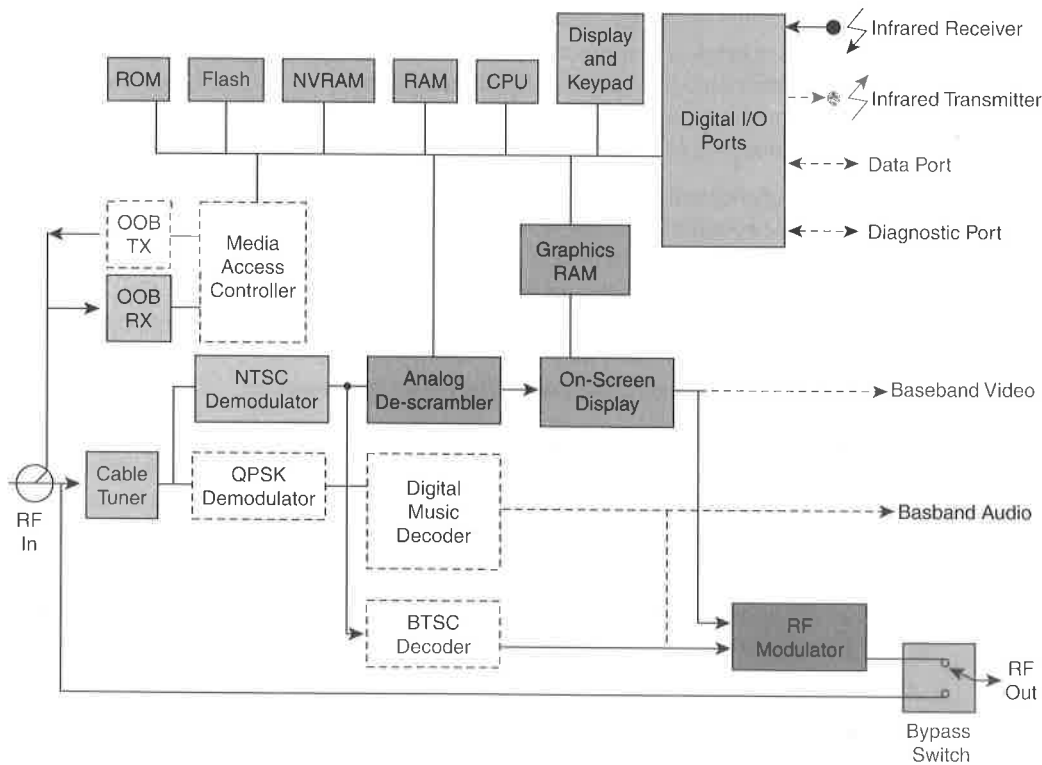
- *Basic set-top converters* that perform only a tuning function—Basic set-tops are almost obsolete because nearly all television receivers are now *cable ready* (that is, they have an F-connector input and can tune according to the cable frequency plan).

- *De-scrambling set-top converters* that perform tuning and conditional access functions—The authorized services enabled for the set-top must be programmed into it before it is installed in the customer's home.
- *Addressable set-top converters* that perform tuning and conditional access functions—The authorized services enabled for a particular set-top are controlled remotely, from the headend, by sending messages that are addressed to the set-top (and that set-top alone).
- *Advanced analog set-top converters*, which are addressable set-tops that support additional features, such as an on-screen display (OSD), an electronic program guide (EPG), and impulse pay-per-view (IPPV).

Since 1995, advanced analog set-top converters have been replacing other set-top types in many cable systems. Therefore, this chapter focuses on the advanced analog set-top converter.

Figure 3-1 shows the block diagram of a typical advanced analog set-top. Although there are a number of different implementation approaches, this block diagram describes most advanced analog set-tops. Optional blocks are shown using broken lines.

Figure 3-1 Block Diagram for a Typical Advanced Analog Set-Top



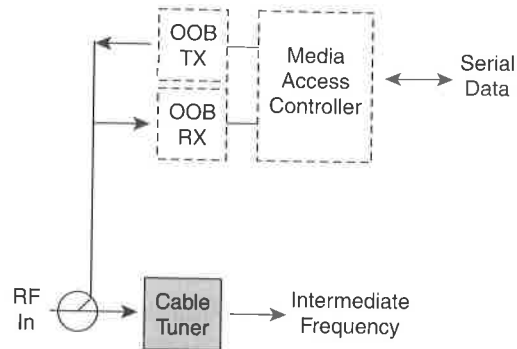
The next sections consider Figure 3-1 in detail after breaking it down into its major subsystems:

- Cable network interface—This subsystem is the interface between the cable system and the set-top. This interface includes the cable tuner and, optionally, an out-of-band (OOB) receiver, an OOB transmitter, a Media Access Controller (MAC), and a VBI decoder.
- Conditional access system—This subsystem provides the means by which access to specific services is granted to the user, based on payment for those services. In analog set-tops, a video de-scrambling circuit performs this function. (Some manufacturers also use some form of audio scrambling.)
- On-screen display (OSD)—Visual information is presented to the user by means of an OSD, which overlays text and graphics onto the video output.
- Audio processing—Most set-tops include a remote volume control and, occasionally, other audio processing options are offered to the customer, such as Broadcast Television System Committee (BTSC) stereo decoding or digital music decoding.
- Microprocessor subsystem—This subsystem is the brains of the set-top and includes a microprocessor with ROM, Flash, NVRAM, RAM, an LED display, and a keypad.
- Inputs—All set-tops have a coaxial F-connector input. In addition, user control signals are received via an infrared receiver from a remote control. Other inputs might include a data port and a diagnostic port.
- Outputs—All set-tops have an F-connector output for a National Television Systems Committee (NTSC) channel modulated at broadcast channel 3 or channel 4. (Other channels are available as well.) In addition, the set-top could have *baseband* (or *composite*) video and baseband audio outputs.
- Infrared transmitter—Set-tops can include an infrared transmitter to control a VCR.
- Remote control—Advanced analog set-tops are usually supplied with a remote control (not shown in Figure 3-1).

The Cable Network Interface

The cable network interface is shown in Figure 3-2. It is completely specified for clear (unscrambled) video, but most of the cable network interface is proprietary and varies from one set-top model to another.

Figure 3-2 Cable Network Interface



The standard aspects are

- The cable frequency plan, specified by the Electronic Industry Association (EIA) standard EIA-542 "Cable Television Channel Identification Plan" (see Chapter 8 of *Modern Cable Television Technology; Video, Voice, and Data Communications* by Walter Ciciora and others).
- The video and audio channel format, specified by the National Television Systems Committee (NTSC) specification, which includes extensions for color.
- Stereo and Secondary Audio Program (SAP) encoding, specified by the BTSC standard.

The proprietary aspects of the cable network interface are

- The video de-scrambling system—Not only is this different from one vendor's system to another, but a single vendor might have a number of different scrambling systems.
- The out-of-band signaling system—Each vendor's system is proprietary, and there is only limited compatibility across the different set-top models within a vendor's product range.

The proprietary aspects of the analog cable network interface present some difficult problems for the cable operator. After an operator has selected a set-top, that set-top effectively dictates the specification of the cable network interface. The operator is locked in to that set-top model and any compatible models (which are usually available only from the same vendor).

On the other hand, the existence of a number of proprietary scrambling systems discourages signal theft. As long as a scrambling system is not deployed too widely, it is not economic for the pirate to try to defeat it. Unfortunately, simple analog scrambling systems based on video sync-suppression and video inversion have become so widely deployed that they present an easy target for signal theft.

The cable network interface consists of the following components:

- A cable tuner
- An out-of-band receiver
- An optional out-of-band transmitter
- An optional media access control circuit

These components are described in more detail in the following sections.

Cable Tuner

The cable tuner selects a single 6 MHz channel between 50 and 860 MHz and rejects all other channels. The frequency is converted in two steps to an intermediate frequency (IF) that is suitable for demodulation by the NTSC demodulator. Tuning requirements for analog set-tops are fundamentally straightforward and have not changed for some time; the tuner performance is tightly specified to include image rejection, tuning range, sensitivity, and so on over the entire radio-frequency spectrum of the cable plant. Analog tuner performance is specified in EIA-23, and the cable frequency plan is specified by EIA-542.

NTSC Demodulator

The NTSC demodulator detects and recovers the video and audio signals from the tuned channel. Although designs vary, the requirements for NTSC demodulation are well known, and it is a standard component in the analog set-top.

Out-of-Band Receiver

The out-of-band receiver tunes and demodulates a narrow, digitally modulated carrier and feeds the recovered serial data to the media access controller. The out-of-band channel specification is proprietary and the channel is of quite limited capacity (in the order of kilobits per second).

Some addressable set-tops lack an out-of-band receiver and one or more VBI lines are used to send addressed messages to the set-top in lieu of the out-of-band channel. (The VBI lines are video scan lines that are not displayed by the television receiver, and that are used to send coded data, including closed captioning information.)

Out-of-Band Transmitter

The out-of-band transmitter modulates a serial data stream onto a narrow carrier for transmission upstream to the cable headend. The out-of-band channel specification is proprietary and typically uses a very robust modulation technique such as frequency shift keying (FSK). When analog set-tops with out-of-band transmitters were first deployed, the return system was not very sophisticated and the noise floor was typically high due to noise funneling (see the section Return Path Activation in Chapter 2, “Analog Cable Technologies”). Therefore, the out-of-band transmitter was designed to operate in that environment.

In addition, the impulse pay-per-view application uses a store-and-forward approach that retries each out-of-band channel transaction until it is successful. Even in the presence of errors, the transaction is eventually completed.

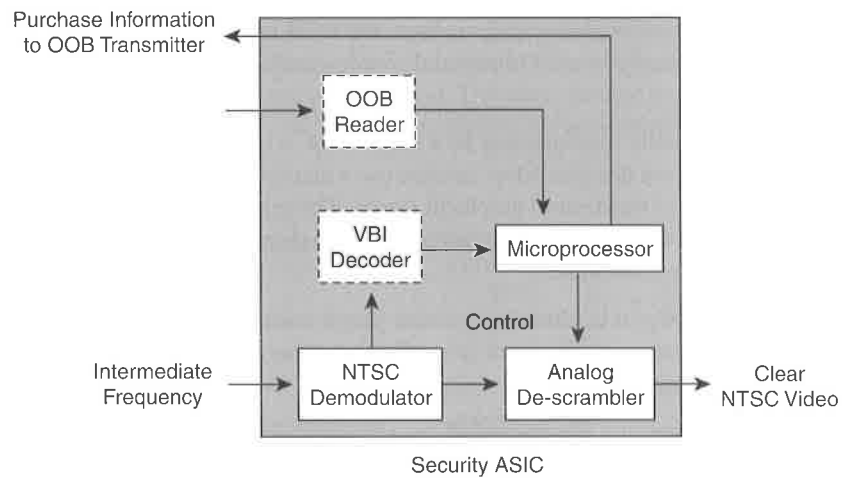
Media Access Control

The media access control (MAC) function ensures that only one set-top transmits at any given time on the cable return path (which is a shared medium). In analog set-tops, this function is typically accomplished by polling. In a polled system, the set-top is allowed to transmit only in response to a poll. This is a very simple MAC to implement, but it can take hours to poll every set-top in turn. Therefore, polled systems are impractical for any kind of interactive signaling.

Conditional Access System

The conditional access system provides the means by which access to specific services is granted to the user based on payment for those services. The conditional access system, depicted in Figure 3-3, uses embedded security components in the set-top. Most modern analog set-top converters use an Application Specific Integrated Circuit (ASIC) that contains the analog descrambler/demodulator, OOB data reader, microprocessor, and data decryptor. Many of the early set-top converters were defeated by monitoring various points in the unit to see how signals changed; embedding everything in a single ASIC prevents this.

Every aspect of the analog conditional access system is highly proprietary. It is common for set-top purchase contracts to include financial remedies in the case of a security breach that might include the cost of upgrading or replacing security elements. Thus, the set-top vendor has every incentive to make the conditional access system as obscure and complicated as possible in an effort to defeat signal theft. This approach has been nicknamed *security through obscurity*.

Figure 3-3 Conditional Access System

Analog De-scrambler

The analog de-scrambler is designed to restore a scrambled channel to its original form. There are a number of analog scrambling techniques, including video sync-suppression, video inversion, line dicing, line rotation, line shuffling, and time jitter. However, most scrambling systems in North America rely on video sync-suppression and video inversion. (See Chapter 18 of *Modern Cable Television Technology; Video, Voice, and Data Communications* by Walter Ciciora and others for an excellent tutorial on analog scrambling techniques.)

Analog scrambling has a number of limitations:

- There is a slight degradation in picture quality due to the analog scrambling and de-scrambling process.
- Cable ready televisions allow the customer to tune to scrambled channels and to occasionally display distorted, but recognizable, pictures for brief periods (usually when a vertical bar at the left edge of the picture provides the receiver with a substitute for the horizontal synchronization pulse).
- Some analog scrambling systems do not mask the audio channel of a scrambled channel. This has limited the transmission of adult programming to late at night (the so-called *safe-harbor times*) on systems without audio masking.
- Analog scrambling is far too easy to defeat. A quick survey of the World Wide Web (WWW) yields a list of many individuals and companies offering "retail" set-tops that will de-scramble all channels on a given cable system.

On-Screen Display

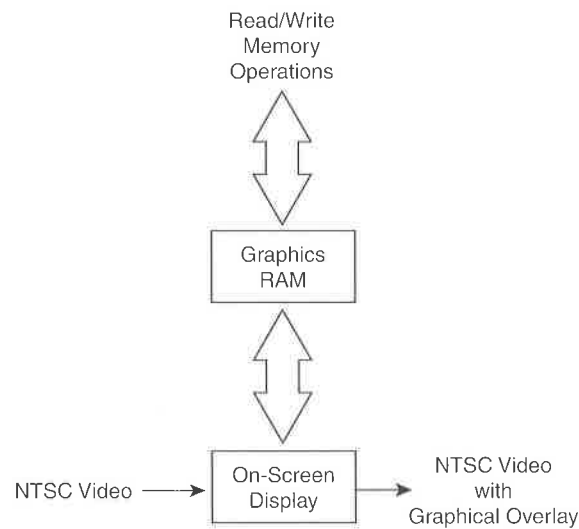
Figure 3-4 depicts the on-screen display (OSD), which generates a graphical overlay on the video output. In advanced analog set-tops, the OSD is quite primitive compared to digital set-tops. It is usually limited to text and simple graphics and supports only a small palette of colors.

The OSD is usually implemented by a single chip, which might be integrated with other functions in recent designs. Most designs use a dual-ported memory and allow the CPU to write a bitmap of the desired graphical image. The graphical image is typically combined with the NTSC signal by selecting either the video signal or the graphical image in memory, based on its x-y coordinate.

The combined signal is filtered to prevent illegal transitions in the NTSC signal. The best example of filtering is called *anti-alias filtering*, where adjacent field lines are filtered to prevent flickering.

Despite its limitations, the on-screen display supports a menu-driven user interface, an electronic program guide and, often, other applications. (See the section Applications later in this chapter.)

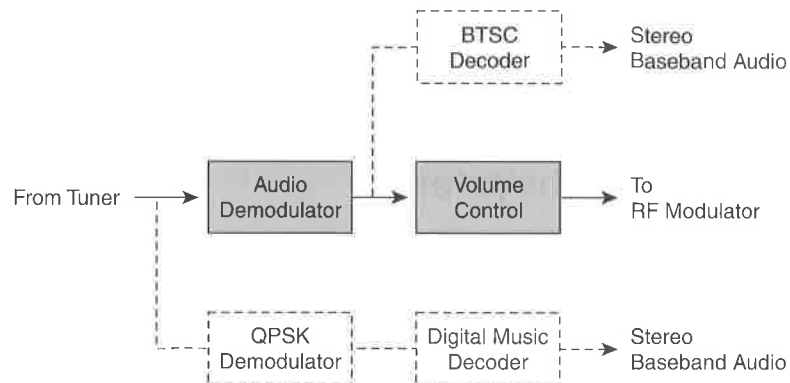
Figure 3-4 *On-Screen Display*



Audio Processing

The audio processing capabilities of analog set-tops are typically limited to a volume control function, though BTSC stereo decoding and digital music services have occasionally been offered. Figure 3-5 shows that the analog carrier is usually fed into the RF modulator without any modification. This arrangement has the advantage that BTSC stereo and the secondary audio program carriers are available for demodulation by circuitry in the television receiver.

Figure 3-5 Audio Processing



Volume Control

A volume control function is commonly included in analog set-tops (refer to Figure 3-5). This feature is very useful when the set-top is used with older televisions without a remote control, or if the customer wants to use the set-top remote control for all functions.

The volume control function works by adjusting the level of the audio signal that is fed into the RF modulator. Signal-to-noise ratio inevitably suffers, but many customers do not seem to notice. Because the BTSC decoder in the television receiver gives good stereo separation only if the audio deviation is set to its optimum value, the set-top has a *best-stereo* setting that sets the volume control to the optimum level.

Digital Music

Two proprietary digital music systems have been developed in North America:

- Digital Music Express (DMX)
- Music Choice

These systems use a narrowband quaternary phase shift keying (QPSK) modulated carrier to send digital stereo music signals through the cable system. Typically, a separate device is required to receive digital music, and this has slowed its acceptance. However, it is possible to combine the digital music receiver circuitry into the set-top box, as shown in Figure 3-5. (See the “CFT-2200” section, later in this chapter.)

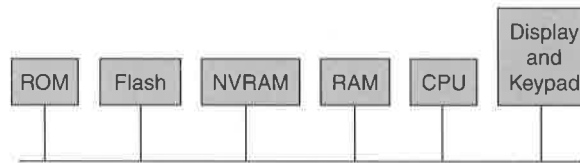
BTSC Stereo and SAP Decoding

The BTSC/SAP feature is extremely rare on analog set-tops and, although it has been offered as an option (refer to Figure 3-5), few set-tops with this feature have been deployed. (See the “CFT-2200” section, later in this chapter.) A BTSC/SAP decoder provides the option of baseband audio outputs that can be used as input to a stereo system or the television receiver.

Microprocessor Subsystem

The microprocessor subsystem, depicted in Figure 3-6, differentiates the advanced analog set-top from earlier analog set-tops.

Figure 3-6 *Microprocessor Subsystem*



The microprocessor subsystem enables a new range of features for the cable operator and the cable customer:

- The advanced set-top contains firmware that supports a very flexible, menu-driven, user interface that, in combination with the OSD, allows greater ease-of-use and more complex features to be offered to the customer.
- The electronic program guide (EPG) allows the customer to scan the program offerings using the OSD, usually by means of a table of channels (the rows of the table) and times (the columns of the table). The EPG relies on program guide data that is usually sent to all set-tops periodically in the out-of-band channel or it might be embedded in the VBI.
- In some advanced analog set-tops, the firmware can be downloaded over the network so that additional features (or bug-fixes) can be distributed to set-tops that have been deployed. For example, a change in the EPG display could be made.

These features, particularly the EPG, have been well received by cable customers.

The microprocessor subsystem has a number of components, which are shown in Figure 3-6 and described in the following sections:

- The central processing unit
- The memory subsystem
- The display and keypad

Central Processing Unit

The central processing unit (CPU) is typically an 8- or 16-bit microprocessor. By trailing the microprocessor price-performance curve of the processors commonly used in personal computers, the analog set-top limits the cost of the CPU to just a few dollars.

CPU performance is low, typically between 1 and 3 million instructions per second (mips), but this is adequate for the limited set of embedded applications that a set-top has to perform if they are coded efficiently. The firmware is typically written in a mix of assembly language and C.

Memory Subsystem

The memory subsystem contains a number of different types of memory, each having a specific purpose. In more recent designs that use a microcontroller, part of the memory might be contained on the same chip as the microprocessor.

The different types of memory are

- Read-only memory
- Flash memory
- Nonvolatile random access memory
- Random access memory

Read-Only Memory

ROM contains the firmware for the set-top. ROM is the least expensive type of memory and is used for storing program instructions and data that are never changed. Because "never" is a long time, programs are usually not committed to ROM until they are thoroughly debugged, which requires exhaustive, thorough testing of the program. This testing takes time, so early versions of the set-top might use programmable read-only memory (PROM) to speed the development cycle. When the code is stable, it will typically be stored as masked-ROM because it is the cheapest type of ROM. (In the masked-ROM process, the program is lithographically etched into the chip as the final metal connection layer.)

Flash Memory

Flash memory is also referred to as electrically erasable programmable read-only memory (EEPROM). Flash memory can be erased and rewritten thousands of times, so it is ideal for storing program code and data that might need to be changed due to bug-fixes or other enhancements. Flash memory is typically two to three times more expensive than ROM.

Some advanced analog set-tops support field upgrade of their functions, and this is achieved by downloading new program code and data over the network. Such field upgrades can be very useful to the cable operator but are also fraught with peril! If the download mechanism is not well designed, the set-top might be left in an undefined state or might fall back to a basic set of functions.

One disadvantage of certain types of flash memory is that it might not support the access speed required by the microprocessor. In this case, the contents of the flash memory can be copied to RAM before execution, or the microprocessor can be programmed to insert wait cycles. The first solution requires additional RAM, whereas the second solution can seriously affect performance.

Nonvolatile Random Access Memory

NVRAM is used to store any program variables or user data that must survive when the set-top loses AC power. There are two common techniques for implementing NVRAM:

- The use of NVRAM chips—These chips are very reliable, but they are expensive and available only in limited sizes (2 KB is used in some advanced analog designs).
- The use of battery-backed RAM—This is standard RAM that is powered by a small on-chip battery. The RAM size can be many times larger than NVRAM chips, but if the battery fails it must be replaced.

Random Access Memory

RAM is read/write memory that retains its contents only while it is powered. RAM is used to store data and program code. You are probably familiar with the dramatic reduction in the price of RAM for personal computers. The same trends make analog set-tops with 1 or 2 MB of RAM economical.

One of the main uses for RAM in an analog set-top is the storage of program guide data. The program guide data is sent to the analog set-top at a rate of only 10 to 20 Kbps. This is adequate because the program schedule data is fairly static, that is, it does not change very much or very often. The program guide data is stored in RAM as it is received and is then available for instantaneous access by the electronic program guide.

Display and Keypad

In an advanced analog set-top, customers use the OSD and the remote control for most of their interaction with the set-top. However, remote controls have a way of slipping down the back of the couch or exhausting their batteries, so a keypad is usually included on the set-top itself for this eventuality.

For convenience, the OSD is supplemented by a 4-digit display that can be used to display the channel number or the time.

RF Modulator

The RF modulator (refer to Figure 3-1) takes the composite NTSC video and audio carriers and places them on channel 3 or channel 4 for output to the television receiver. Channel 3 or channel 4 selection might be by a switch on the set-top or by software control. Output channel options are given so that interference from local broadcast stations operating on the same channel can be avoided.

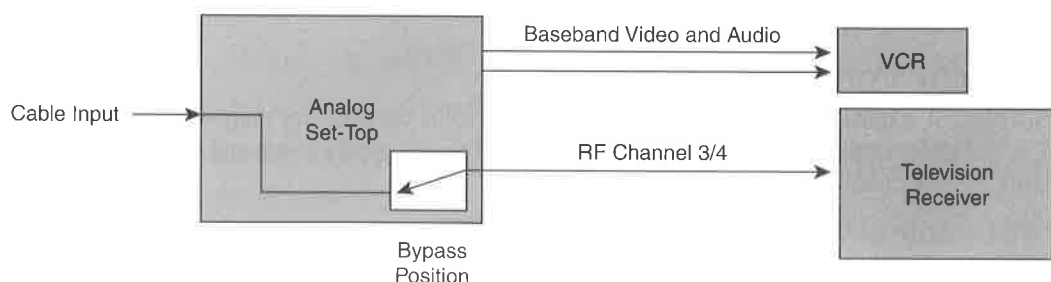
RF Bypass Switch

The RF bypass switch (refer to Figure 3-1) is actually a coaxial relay that returns to the bypass setting when the set-top is powered off. In this position, the RF bypass switch allows all cable channels to pass through the set-top unchanged to the television receiver. (This behavior is mandated by the 1992 Cable Act.)

When the set-top is powered, the RF bypass switch normally connects the output of the RF modulator to the set-top output (and to the television receiver connected to it).

However, some analog set-tops include a remote control button that toggles the bypass switch. This is intended to allow the customer to use the set-top for a watch-and-record function, as shown in Figure 3-7.

Figure 3-7 *Watch and Record Using the Bypass Switch*



The VCR is connected to the baseband video and audio outputs of the set-top and records a program (which might be scrambled) while the RF signal is bypassed to the television receiver so that the customer can view any unscrambled service. (There are probably a few people in North America who regularly use this feature, but it causes great confusion to everyone else.) This feature might be combined with automatic control of the VCR and is discussed in more detail in the section “Infrared Transmitter,” later in this chapter.

Inputs

The advanced analog set-top has the following inputs:

- The cable input
- An infrared receiver
- A diagnostic port
- A data port

The cable input is the familiar 75-ohm F-connector but the other inputs use proprietary connectors and standards. These four inputs are described in the following sections.

Cable Input

The cable input is a 75 ohm female F-connector. It is the only radio frequency input to the shielded set-top unit. Shielding is important to prevent off-air television broadcasts from interfering with the cable channels.

Infrared Receiver

An infrared receiver is mounted on the front of the set-top box to allow line-of-sight operation with an infrared remote control. Advanced analog set-tops are often supplied with a *universal* remote control, which can also control the television and the VCR but often causes considerable confusion to the average customer.

Diagnostic Port

It is quite common for the set-top to have a serial data port for diagnostic purposes. This port is typically used during development of the set-top firmware and might be completely disabled in a production unit.

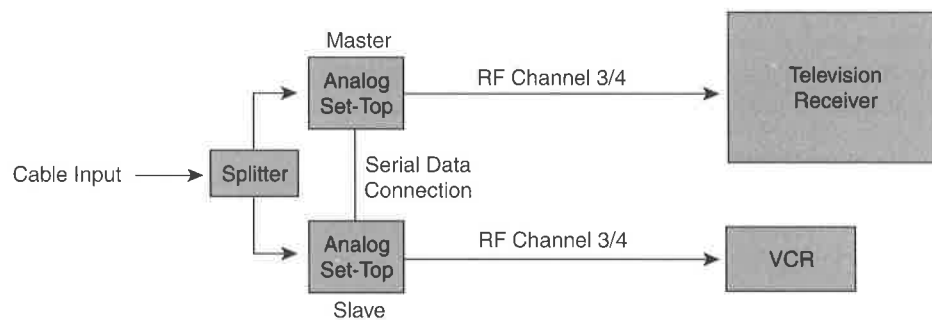
Data Port

A data port on the set-top can be used for various expansion purposes. The most common of these is to satisfy the requirements stated in Section 17 of the 1992 Cable Act:

- To watch a program on one channel while simultaneously using a video cassette recorder to tape a program on another channel
- To use a video cassette recorder to tape two consecutive programs that appear on different channels
- To use advanced television picture generation and display features

These requirements are colloquially known as the watch-and-record and picture-in-picture (PIP) requirements and can be satisfied if the set-top contains multiple tuners and multiple de-scramblers. However, the customer demand for these features is low, and it is generally not economical to manufacture a special set-top for this purpose. Instead, the data ports of two standard (single-tuner, single de-scrambler) set-tops are connected with a serial data connection so that they behave as a dual-tuner, dual de-scrambler set-top, as shown in Figure 3-8. In this arrangement, all user interface functions are handled by one set-top (the master), and the second set-top (the slave) merely provides a second tuner and de-scrambler function for the master set-top.

Figure 3-8 *Dual Set-Top Arrangement*



Outputs

The advanced analog set-top has the following outputs:

- RF channel 3/4 output
- Baseband video
- Baseband audio
- Infrared transmitter

The RF output is always present and uses the standard 75 ohm F-connector. All other outputs are optional. The outputs are discussed in the following sections.

RF Output

An RF output provides the simplest and most common method of connection to a television receiver or VCR. As discussed previously in this chapter, the RF output carries

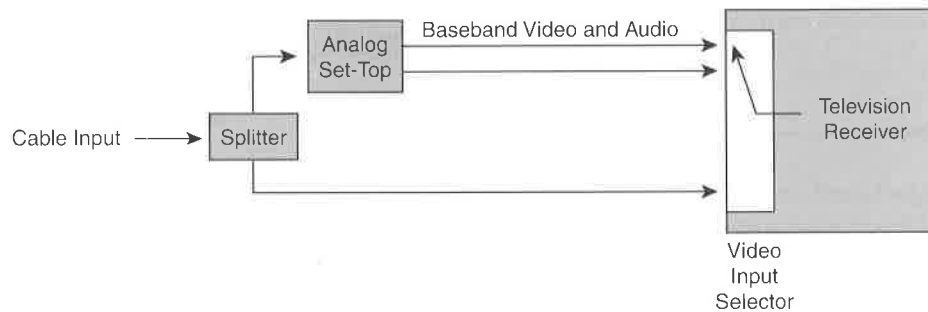
- A clear NTSC signal modulated onto channel 3 or channel 4. This signal is present only when the set-top is active.
- The entire cable spectrum from the RF input. This signal is present when the set-top is powered off or when the bypass switch is activated.

Baseband Video

An optional baseband video output is provided for connection to the *video-in* input of a television or VCR. This signal is sometimes called a *composite signal* because it contains the baseband luminance information and the modulated chrominance information according to the NTSC standard. The RCA-style connector is the *de facto* standard for this signal.

One advantage of the baseband video signal is that it allows the customer to use the input selector on the television receiver to switch between the output of the set-top and the broadband cable signal, as shown in Figure 3-9.

Figure 3-9 *Baseband Video Output*



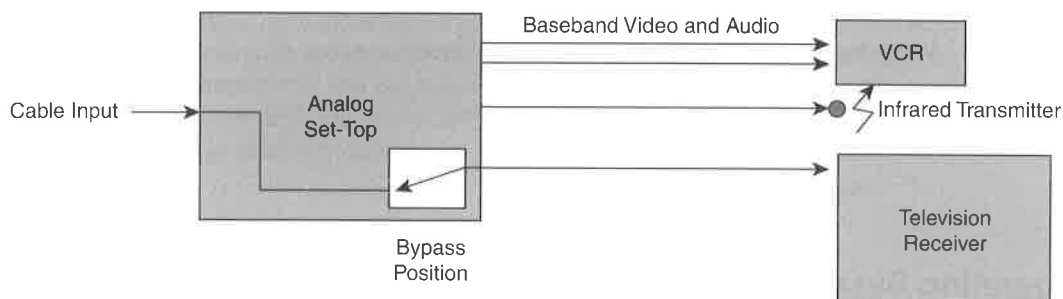
Baseband Audio

An optional baseband audio output is provided for connection to the *audio-in* input of a television or VCR. This signal is usually monaural because a BTSC decoder is rarely included in an analog set-top. Even when a stereo baseband audio output is available, most customers are disappointed to learn that a hi-fi stereo VCR cannot make stereo recordings from its baseband audio inputs because it does not have a BTSC stereo encoder.

Infrared Transmitter

An optional infrared transmitter is used to control a VCR. (This is sometimes called an *IR blaster*.) If correctly configured, as shown in Figure 3-10, this can simplify the VCR recording considerably. It allows the customer to select and record a program without programming the VCR.

Figure 3-10 Watch and Record Operation



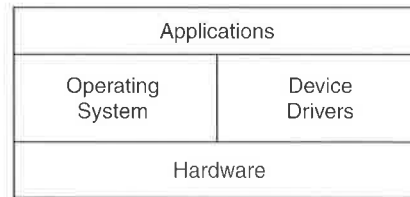
The operation is as follows:

- The customer selects a program to record using the EPG. This causes the set-top to set an internal timer that will wake up the set-top when the program starts.
- When the program starts, the set-top wakes up, tuned to the desired program. It then sends a control sequence to the infrared transmitter (which has been positioned to illuminate the VCR's infrared receiver). The control sequence turns on the VCR and starts recording the program.
- When the program finishes, the set-top sends a control sequence to the infrared transmitter that stops recording the program and turns off the VCR. The set-top then goes back to sleep.

What could be simpler? But did you remember to put a blank tape in the VCR? And did you remember to leave the VCR in its off state? (If you left it on, the infrared signal to turn it on will have turned it off!)

Software Architecture

The software architecture of an advanced analog set-top is designed to support a limited set of embedded applications. In this environment, the software architecture shown in Figure 3-11 is quite adequate.

Figure 3-11 *Set-Top Software Architecture*

The software can be divided into three main categories:

- Operating system
- Device drivers
- Applications

Operating System Software

The set-top operating system is quite simple because only a few functions are required:

- Task scheduling—Task scheduling provides the basic mechanism for ensuring that applications are run in sequence so that each has a turn to use the processor.
- Timers—Timers are used to trigger task execution at some future time and are used by unattended recording and sleep timer applications.
- Interrupt handling—The operating system dispatches interrupts to the appropriate device driver.

A number of commercial operating systems have been developed that satisfy these requirements; examples are pSOS, VRTX, and VxWorks. It is also quite simple to develop an operating system with these functions in-house.

The operating system is simple because all other functions—for example, memory management—are done by the application.

Device Drivers

Device drivers are programs that work directly with the hardware devices and are often written in assembly language for optimum performance. The device driver hides the details of the hardware from the applications. However, in an embedded applications environment, the distinction between a driver and an application is sometimes blurred.

Applications

The set-top applications perform all the desired functions of the set-top and are typically coded in C language or assembly language. The applications developer is responsible for ensuring that the application does not corrupt its own or other applications data, because there is no memory management in the operating system or hardware. Therefore, the applications are *trusted* and must be tested thoroughly, because if an application fails, it causes the set-top to crash or hang up.

The set-top applications are not portable; that is, they cannot be readily moved to a different set-top platform because they rely too much on the specifics of the underlying hardware. Until recently, this was not a significant issue, because each set-top software development effort is typically done separately as a standalone project and the software development effort was relatively small.

Case Studies

Two advanced analog set-tops are being widely deployed in North America. They are the CFT-2200 from General Instrument and the 8600X from Scientific Atlanta.

This section provides a brief overview of the hardware and software features of these advanced analog converters. The limitations of advanced analog converters are discussed together with the lessons that have been learned from deploying them.

CFT-2200

The CFT-2200 is the latest in a series of advanced analog set-tops from General Instrument (GI), and several million units have been deployed.

The CFT-2200 has a single, 50 to 860 MHz tuner and supports GI analog scrambling formats. It uses a secure microprocessor smart card to provide for upgrade of the conditional access system.

The CFT-2200 has an out-of-band receiver that tunes a proprietary control channel in the FM band. A factory-installed module provides an RF transmitter for two-way operation.

The CFT-2200 supports an infrared transmitter and has an optional data port that uses the RS-232 standard.

The CFT-2200 uses a Motorola 68000 microprocessor with a performance of about 10 mips. GI calls this the Feature Expansion Module Processor, and it is powerful enough to run a number of different applications:

- EPG—GI gives the cable operator a choice of an EPG developed in-house or the Prevue Express EPG, which is licensed from Prevue Networks Incorporated.

- Impulse pay-per-view—The secure processor and return module allow the customer to buy an event (typically a movie) and collect the purchase information from the set-top. The event purchase price is added to the next monthly bill.
- Enhanced audio—An optional hardware module supports digital music services offered by Music Choice. A BTSC stereo decoder is optional.
- Internet browsing and e-mail—The WorldGate TV online application uses the VBI lines to send Web pages at up to 128 Kbps to the set-top. The software also supports email and chat services. An infrared keyboard is supplied to allow the customer to type.
- Enhanced programming—An application provided by Wink Communications uses the VBI lines to carry data to the set-top that is associated with the television programming. The Wink application enhances the programming with text and graphics. These enhancements are *authored* to the program before it is transmitted. Wink also allows the customer to interact with the program enhancements.
- Interactive services—The LocalWorks application developed by CableSoft allows local information to be provided to the customer.

8600X

The 8600X Home Communications Terminal (HCT) is the latest in a series of advanced analog set-tops from Scientific Atlanta (SA), and several million units have been deployed.

The 8600X HCT has a single, 50 to 860 MHz tuner and supports SA analog scrambling formats and GI-compatibility mode.

The 8600X HCT lacks an out-of-band receiver. Instead, it uses the VBI lines to insert control data into one or more channels at the headend. The 8600X HCT is programmed to tune to these channels whenever the set-top is switched off from the remote control or the keypad. The disadvantage of this approach is that while the set-top is left switched on and tuned to a channel that does not contain the control data, it does not receive any control updates and could eventually become deauthorized. This problem can also occur when the set-top is connected to a switched wall outlet and has no AC power when it is not being used.

The 8600X HCT has options for an RF transmitter for two-way operation. The 8600X HCT supports an infrared transmitter and has a data port. The 8600X HCT uses an Intel microprocessor with a performance of about 3 mips. This processor is used to support the following bundled applications:

- EPG—The EPG application is developed by SA.
- Impulse pay-per-view—The secure processor and return module allow the customer to buy an event (typically a movie) and collect the purchase information from the set-top. The event purchase price is added to the next monthly bill.

The 8600X also supports a plug-in module, known as the genius card, that contains a **second** processor and additional memory. This card is required to run supported third-party applications:

- Enhanced programming—An application provided by Wink Communications uses the VBI lines to carry data to the set-top that is associated with the television programming. The Wink application enhances the programming with text and graphics. These enhancements are *authored* to the program before it is transmitted. Wink also allows the customer to interact with the program enhancements.
- Internet browsing and e-mail—The WorldGate TV online application uses the VBI lines to send Web pages at up to 128 Kbps to the set-top. The software also supports e-mail and chat services. An infrared keyboard is supplied to allow the customer to type.

Limitations

Advanced analog set-tops have a number of limitations that have slowed the deployment of advanced services:

- The advanced analog set-top has performance, memory, and graphics limitations that do not allow support of sophisticated applications.
- The advanced analog set-top has no application sharing support and is generally limited to running a single interactive application.
- Communication between the set-top and the headend are limited to a single, narrow channel that is shared by all set-tops. This limitation has been partially solved by Wink Communications and Worldgate by using multiple VBI lines, but it is still a fundamental issue that prevents scaling of the system.
- Any new applications must be ported to the proprietary hardware platform. This requires an agreement with the set-top provider and takes time. As the number of applications grow, and as third-party application developers develop more applications, the lack of application portability becomes a significant issue.

Lessons Learned

The advanced analog deployment experience has taught the cable industry a great many lessons that can be applied to digital cable deployment and which provide some of the foundations for the OpenCable architecture:

- The cable network interface must be consistent across set-tops from all vendors, and in practice this can be achieved only by establishing a cable network interface standard.
- The conditional access system must be upgradable on an ongoing basis to provide continued protection from signal theft.

- A scalable, two-way communications network is required for interactive applications support.
- As applications become more sophisticated, they require more processor and graphics performance and more memory.
- A standard applications environment is required to allow applications to be easily ported to new set-top platforms as they become available.
- The operating system must have the capability of supporting concurrent execution of more than one application.

Summary

This chapter has covered the hardware and software characteristics of the analog set-top converter. There are many aspects of analog set-tops that make the services that cable operators can provide to their customers less than ideal. The analog set-top fosters proprietary interfaces and provides a limited applications platform. It does not always interoperate well with other video equipment—such as television receivers, VCRs, and home-theater equipment—in the customer's home. There are many valuable lessons to be learned from the analog set-top in the transition to digital cable systems.

References

Book

Ciciora, Walter, James Farmer, and David Large. *Modern Cable Television Technology; Video, Voice, and Data Communications*. San Francisco, CA: Morgan Kaufmann Publishers, Inc., 1999.

Internet Resources

CableSoft Web site

www.cablessoft.com

Consumer Electronics Manufacturers Association Web site

www.cemacity.org

General Instrument Web site

www.gi.com

Music Choice Web site

www.musicchoice.com

PreVue Web site

www.prevue.com

Scientific Atlanta Web site

www.sciatl.com

Wink Web site

www.wink.com

Worldgate Web site

www.wgate.com

Digital Technologies

Digital technology is becoming pervasive in all types of services. As computing power continues to increase according to Moore's Law (see Chapter 1, "Why Digital Television?"), more and more functions can be tackled in the digital domain. An excellent example is the transmission of television pictures.

Nevertheless, digital technology is not a panacea. The complexity of the techniques can introduce reliability and quality issues. In addition, only a few engineers thoroughly understand all these techniques, making us more reliant on a smaller number of *de facto* standard chip-sets.

This chapter introduces a number of key digital technologies. If you are familiar with these technologies, you might want to skim through or skip over this chapter. There are many excellent texts (see the list of references at the end of the chapter) that explain how these digital technologies work. This chapter does not attempt to repeat them, but instead provides a commentary on why these techniques are so important and what they mean in practical terms. This chapter discusses

- Video compression—The basic principles of the video compression algorithms commonly used for entertainment quality video, the importance of choosing the correct parameters for video encoding, and some alternative video compression algorithms.
- Audio compression—The basic principles of MPEG-2 audio compression and Dolby AC-3 audio compression.
- Data—Arbitrary *private* data can be carried by the underlying layers.
- System information—Tabular data format used by the digital receiver device to drive content navigation, tuning, and presentation.
- Timing and synchronization—A mechanism is required to recover the source timing at the decoder so that the presentation layer can synchronize the various components and display them at exactly the intended rate.
- Packetization—The segmentation and encapsulation of elementary data streams into transport packets.
- Multiplexing—The combining of transport streams containing audio, video, data, and system information.

- Baseband transmission—The various mechanisms for carrying digital transport streams: the Digital Video Broadcast (DVB) asynchronous serial interface (ASI), Synchronous Optical Networks (SONET), Asynchronous Transfer Mode (ATM), and Internet Protocol (IP).
- Broadband transmission—The digital transmission payload must be modulated before it can be delivered by an analog cable system. Quadrature Amplitude Modulation (QAM) has been selected as the best modulation for *downstream* transmission in cable systems. Other modulation techniques include quaternary phase shift keying (QPSK) and vestigial side band (VSB) modulation.

Figure 4-1 shows how each of these techniques are layered. This diagram illustrates the in-band communications stack for a digital cable set-top and is discussed in Chapter 6, “The Digital Set-Top Converter,” in more detail.

Figure 4-1 Layered Model for Digital Television

Video Compression	Audio Compression	Data	System Information
Timing and Synchronization			
Packetization			
Multiplexing			
Baseband Transmission			
Broadband Transmission			

Video Compression

Image compression has been around for some time but video compression is relatively new. The processing requirements to compress even a single frame of video are large—to compress 30 frames (or 60 fields) per second of video requires massive processing power (delivered by rapid advances in semiconductor technology).

Nonetheless, digital video must be compressed before it can be transmitted over a cable system. Although other compression algorithms exist, the dominant standard for video compression is MPEG-2 (from Moving Picture Experts Group). Although MPEG-2 video compression was first introduced in 1993, it is now firmly established and provides excellent results in cable, satellite, terrestrial broadcast, and digital versatile disk (DVD) applications.

This section discusses

- MPEG-2 video compression—The basics of the MPEG-2 video compression algorithm and why it has become the dominant standard for entertainment-video compression
- Other video compression algorithms—Why other video compression algorithms have their applications and why they are unlikely to challenge MPEG-2 video compression in the entertainment world for some time
- Details on MPEG-2 video compression—Some more details on the use of MPEG-2 video compression and its parameters

MPEG-2 Compression

MPEG-2 video compression is the *de facto* standard for entertainment video. MPEG-2 video compression is popular for a number of reasons:

- It is an international standard [ISO/IEC IS 13818-2].
- MPEG-2 places no restrictions on the video encoder implementation. This allows each encoder designer to introduce new techniques to improve compression efficiency and picture quality. Since MPEG-2 video encoders were first introduced in 1993, compression efficiency has improved by 30 to 40%, despite predictions by many that MPEG-2's fundamental theoretical limitations would prevent this.
- MPEG-2 fully defines the video decoder's capability at particular levels and profiles. Many MPEG-2 chip-sets are available and will work with any *main level at main profile (MP@ML)*-compliant MPEG-2 bit-stream from any source. Nevertheless, quality can change significantly from one MPEG-2 video decoder to another, especially in error handling and video clip transitions.
- MPEG-2 video compression is part of a larger standard that includes support for transport and timing functions.

Moreover, MPEG-2 is likely to remain as the dominant standard for entertainment video because it has been so successful in establishing an inventory of standard decoders (both in existing consumer electronics products and in the chip libraries of most large semiconductor companies). Additional momentum comes from the quantity of real-time and stored content already compressed into MPEG-2 format. Even succeeding work by the MPEG committees has been abandoned (MPEG-3) or retargeted to solve different problems (MPEG-4 and MPEG-7).

MPEG-2 is a *lossy* video compression method based on motion vector estimation, discrete cosine transforms, quantization, and Huffman encoding. (*Lossy* means that data is lost, or thrown away, during compression, so quality after decoding is less than the original picture.) Taking these techniques in order:

- *Motion vector estimation* is used to capture much of the change between video frames, in the form of best approximations of each part of a frame as a translation (generally due to motion) of a similar-sized piece of another video frame. Essentially, there is a lot of *temporal redundancy* in video, which can be discarded. (The term *temporal redundancy* is applied to information that is repeated from one frame to another.)
- *Discrete cosine transform (DCT)* is used to convert spatial information into frequency information. This allows the encoder to discard information, corresponding to higher video frequencies, which are less visible to the human eye.
- *Quantization* is applied to the DCT coefficients of either original frames (in some cases) or the DCT of the residual (after motion estimation) to restrict the set of possible values transmitted by placing them into groups of values that are almost the same.
- *Huffman encoding* uses short codes to describe common values and longer codes to describe rarer values—this is a type of *entropy coding*.

The foregoing is a highly compressed summary of MPEG-2 video compression (with many details omitted). However, there are so many excellent descriptions of MPEG compression (see *DTV: The Revolution in Electronic Imaging*, by Jerry C. Whitaker; *Digital Compression for Multimedia: Principles and Standards*, by Jerry D. Gibson and others; *Testing Digital Video*, by Dragos Ruiu and others; and *Modern Cable Television Technology; Video, Voice, and Data Communications*, by Walter Ciciora and others) that more description is not justified here. Instead, the following sections concentrate on the most interesting aspects of MPEG:

- What are MPEG-2's limitations?
- What happens when MPEG-2 breaks?
- How can compression ratios be optimized to reduce transmission cost without compromising (too much) on quality?

MPEG Limitations

If MPEG-2 is so perfect, why is there any need for other compression schemes? (There are a great many alternative compression algorithms, such as wavelet, pyramid, fractal, and so on.) MPEG-2 is a good solution for coding relatively high-quality video when certain transmission requirements can be met. However, MPEG-2 coding is rarely used in Internet applications because the Internet cannot generally guarantee the quality of service (QoS)

parameters required for MPEG-2-coded streams. These QoS parameters are summarized in Table 4-1.

Table 4-1 *MPEG-2 QoS Parameters for Entertainment Quality Video*

Bit Rate	1.5–20 Mbps Constant or Variable Bit Rate (CBR or VBR)
Bit error rate (BER)	Less than 1 in 10^{-10}
Packet/cell loss rate	Less than 1 in 10^{-8}
Packet/cell delay variation	Less than 500 nS

As you can see from the table, for entertainment-quality video, MPEG-2 typically requires a reasonably high bit rate, and this bit rate must be guaranteed. Video-coding will, in general, produce a variable information rate, but MPEG-2 allows for CBR transmission facilities (for example, satellite transponders, microwave links, and fiber transmission facilities). As such, MPEG-2 encoders attempt to take advantage of every bit in the transmission link by coding extra detail during less-challenging scenes. When the going gets tough—during a car chase, for example—MPEG-2 encoders use more bits for motion and transmit less detail. Another way to think of this is that MPEG-2 encoding varies its degree of loss according to the source material. Fortunately, the human visual system tends to work in a similar way, and we pay less attention to detail when a scene contains more motion. (This is true of a car chase whether you are watching it or you are in it!)

MPEG-2 coded material is extremely sensitive to errors and lost information because of the way in which MPEG-2 puts certain vital information into a single packet. If this packet is lost or corrupted, there can be a significant impact on the decoder, causing it to drop frames or to produce very noticeable blocking artifacts. If you think of an MPEG-2 stream as a list of instructions to the decoder, you can understand why the corruption of a single instruction can play havoc with the decoded picture.

Finally, MPEG-2 is extremely sensitive to variations in transmission delay. These are not usually measurable in synchronous transmission systems (for example, satellite links) because each bit propagates through the system according to the clock rate. In packet- or cell-based networks, however, it is possible for each packet-sized group of bits to experience a different delay. MPEG-2 was designed with synchronous transmission links in mind and embeds timing information into certain packets by means of timestamps. If the timestamps experience significant *jitter* (or cell delay variation), it causes distortions in audio and video fidelity due to timing variations in the sample clocks—for example, color shifts due to color subcarrier phase variations.

MPEG-2 Artifacts

What are MPEG artifacts? In practice, all lossy encoders generate *artifacts*, or areas of unfaithful visual reproduction, all the time; if the encoder is well designed, all these artifacts will be invisible to the human eye. However, the best laid plans sometimes fail; the following are some of the more common MPEG-2 artifacts:

- If the compression ratio is too high, there are sometimes simply not enough bits to encode the video signal without significant loss. The better encoders will progressively *soften* the picture (by discarding some picture detail); however, poorer encoders sometimes break down and overflow an internal buffer. When this happens, all kinds of visual symptoms—from bright green blocks to dropped frames—can result. After such a *breakdown*, the encoder will usually recover for a short period until once again the information rate gets too high to code into the available number of bits.
- Another common visible artifact is sometimes visible in dark scenes or in close-ups of the face and is sometimes called *contouring*. As the name suggests, the image looks a little like a contour map drawn with a limited set of shades rather than a continuously varying palette. This artifact sometimes reveals the macro-block boundaries (which is sometimes called *tiling*). When this happens, it is usually because the encoder allocates too few quantization levels to the scene.

NOTE

Macro-blocks are areas of 16-by-16 pixels that are used by MPEG for DCT and motion-estimation purposes. See Chapter 3 of *Modern Cable Television Technology; Video, Voice, and Data Communications* by Walter Ciciora and others, for more details.

- High-frequency *mosquito* noise will sometimes be apparent in the background. Mosquito noise is often apparent in surfaces, such as wood, plaster, and wool, that contain an almost limitless amount of detail due to their natural texture. The encoder can be overtaxed by so much detail and creates a visual effect that looks as if the walls are crawling with ants.

There are many more artifacts associated with MPEG encoding and decoding; however, a well-designed system should rarely, if ever, produce annoying visible artifacts.

MPEG-2 Operating Guidelines

To avoid visible artifacts due to encoding, transmission errors, and decoding, the entire MPEG-2 system must be carefully designed to operate within certain guidelines:

- The compression ratio cannot be pushed too high. Just where the limit is on compression ratio for given material at a certain image resolution and frame rate is a subject of intense and interminable debate. Ultimately, the decision involves engineers and artists and will vary according to encoder performance (there is some expectation of improvements in rate with time, although also some expectation of a law of diminishing returns). Table 4-2 gives some guidance based on experience.

Table 4-2 *MPEG-2 Resolution Versus Minimum Bit Rate Guidelines*

Material	Resolution	Minimum Bit Rate (CBR)
Movies	360 x 240 (CIF)	1.5 Mbps
Movies	360 x 480 (half)	2 Mbps
Movies	540 x 480 (3/4)	3 Mbps
Movies	720 x 480 (full)	4 Mbps
Sports (video)	540 x 480 (3/4)	5 Mbps
Sports (video)	720 x 480 (full)	6 Mbps

- The transmission system must generate very few errors during the average viewing time of an event. For example, in a two-hour movie, the same viewers may tolerate very few significant artifacts (such as frame drop or green blocks). In practice, this means that the transmission system must employ forward error correction (FEC) techniques.

Other Video Compression Algorithms

There are a great many alternative video compression algorithms, such as wavelet, pyramid, fractal, and so on (see Chapter 7 of *Digital Compression for Multimedia: Principles and Standards* by Jerry D. Gibson and others). Many have special characteristics that make them suitable for very low bit rate facilities, for software decoding on a PC, and so on. However, it is unlikely that they will pose a significant threat to MPEG-2 encoding for entertainment video in the near future.

Compression Processing Requirements

Let's take a full-resolution video frame that contains 480 lines, each consisting of 720 pixels. The total frame, therefore, contains 345,600 pixels. Remember that a new frame arrives from the picture source every 33 milliseconds. Thus, the pixel rate is 10,368,000 per second. Imagine that the compression process requires about 100 operations per pixel. Obviously, a processor with a performance of 1,000 million instructions per second (mips) is required.

In practice, custom processing blocks are often built in hardware to handle common operations, such as motion estimation and DCT used by MPEG-2 video compression.

Details of MPEG-2 Video Compression

The following sections detail some of the more practical aspects of MPEG-2 video compression:

- **Picture resolution**—MPEG-2 is designed to handle the multiple picture resolutions that are commonly in use for broadcast television. This section defines what is meant by picture resolution and how it affects the compression process.
- **Compression ratio**—MPEG-2 can achieve excellent compression ratios when compared to analog transmission, but there is some confusion about the definition of compression ratios. This section discusses the difference between the MPEG compression ratio and the overall compression ratio.
- **Real-time MPEG-2 compression**—Most of the programs delivered over cable systems are compressed in real-time at the time of transmission. This section discusses the special requirements for real-time MPEG-2 encoders.
- **Non-real-time MPEG-2 compression**—Stored-media content does not require a real-time encoder, and there are certain advantages to non-real-time compression systems.
- **Statistical multiplexing**—This section explains how statistical multiplexing works in data communications systems and what special extensions have been invented to support the statistical multiplexing of MPEG-2 program streams.
- **Re-multiplexing**—Re-multiplexing, or grooming, of compressed program streams is discussed, including a recent technique that actually allows the program stream to be dynamically reencoded to reduce its bit rate.

Picture Resolution

MPEG-2 compression is a family of standards that defines many different profiles and levels. (For a complete description of all MPEG-2 profiles and levels, see Chapter 5 of *DTV: The Revolution in Electronic Imaging* by Jerry C. Whitaker.) MPEG-2 compression is most

commonly used in its main profile at its main level (abbreviated to MP@ML). This MPEG-2 profile and level is designed for the compression of standard definition television pictures with a resolution of 480 vertical lines.

The resolution of a picture describes how many pixels are used to describe a single frame. The higher the resolution, the more pixels per frame. In many cases, the luminance information is coded with more pixels than the chrominance information.

NOTE

The retina of the human eye perceives more detail with rod cells, which are sensitive only to the intensity of light—the *luminance*—and perceives less detail with cone cells, which are sensitive to the color of light—the *chrominance*.

Chroma subsampling takes advantage of the way the human eye works by sampling the chrominance with less detail than the luminance information. In the MPEG-2 main profile (MP), the chrominance information is subsampled at half the horizontal and vertical resolution compared to the luminance information. For example, if the luminance information is sampled at a resolution of 480 by 720, the chrominance information is sampled at a resolution of 240 by 360, requiring one-fourth the number of pixels. This arrangement is called 4:2:0 sampling. The effect of 4:2:0 sampling is to nearly halve the video bandwidth compared to sampling luminance and chrominance at the same resolution.

Compression Ratio

The *compression ratio* is a commonly misused term. It is used to compare the spectrum used by a compressed signal with the spectrum used by an equivalent NTSC (National Television Systems Committee) analog signal. Expressed this way, typical compression ratios achieved by MPEG-2 range from 6:1 to 14:1. Why is the term confusing?

If you take the same video signal and modulate it as an analog signal (uncompressed) and compress it using MPEG-2, you have two very different things. The analog signal is an analog waveform with certain bandwidth constraints so that it fits into 6 MHz, whereas the MPEG-2 elementary stream is just a string of bits that cannot be transmitted until further processing steps are taken. These steps include multiplexing, transport, and digital modulation, and they all affect how much bandwidth is required by the compressed signal.

To compare apples with apples, you must take the same video signal and convert it to an uncompressed digital signal (this is actually the first step in the compression process and is termed *analog-to-digital conversion* or simply *sampling*). You can now compare the uncompressed digital signal with the MPEG-2 compressed elementary stream for a true comparison of the input bit rate and the output bit rate of the compression process. A full-resolution uncompressed video signal sampled in 4:2:0 (see the previous section, “Picture Resolution”) requires 124.416 Mbps. MPEG-2 can squeeze this down to about 4 Mbps with

little or no loss in perceived quality; this is a true compression ratio of 124:4 or 31:1. This is very different than the commonly quoted range of 6:1 to 14:1.

To continue the math, take the 4 Mbps video elementary stream and add an audio stream at 192 Kbps to create a program stream at 4.192 Mbps. Add information to describe how the streams are synchronized and place the data into short transport packets for efficient multiplexing with other streams. You now have a payload of approximately 4.3 Mbps. Using 64-QAM modulation (see the section Broadband Transmission in this chapter), six 4.3 Mbps streams fit into its 27 Mbps payload. Thus, we could express this as a 6:1 compression ratio.

This is all very confusing! In this example, a video signal with a 31:1 *MPEG-2* video compression ratio is roughly equivalent to an *overall* compression ratio of 6:1. (If the example employs 256-QAM and statistical multiplexing, you might achieve an overall compression ratio of 12:1 although the *MPEG-2* video compression ratio is still 31:1.)

In this book, the terms *MPEG-2 video compression ratio* and *overall compression ratio* will be used to distinguish these very different measures.

Real-Time MPEG-2 Compression

Real-time compression is commonly used at satellite up-links to compress a video signal into a digital program stream as part of the transmission (or retransmission) process. Very often, the encoder runs for long periods of time without manual intervention. There must be sufficient *headroom* in the allocated bit rate to allow the encoder to operate correctly for all kinds of material that it is likely to encode. (*Headroom* refers to available, but normally unused, bits that are allocated to allow for the video compression of difficult scenes.) Each channel requires a dedicated encoder, so price is a definite issue for multichannel systems. The encoder must also be highly reliable, and in many cases automatic switching to a backup encoder is required.

Non-Real-Time MPEG-2 Compression

Non-real-time encoders are technically similar to real-time encoders, but have very different requirements. In fact, they may encode in real-time but their application is to encode to a stored media (such as a tape or disc), and a highly-paid *compressionist* usually monitors the compression of each scene. (Compressionists are studio engineers who not only understand how to operate the encoding equipment but also apply their artistic judgment in selecting the best trade-off between compression ratio and picture quality.) Therefore, encoder price is less of an issue and performance is extremely important because the compressed material will be viewed over and over again. In the case of digital versatile disks (DVDs), no annoying visible artifacts, however subtle, can be tolerated, because the picture quality will be carefully evaluated by a magazine reviewer.

Statistical Multiplexing

Statistical multiplexing is a technique commonly used in data communications to extract the maximum efficiency from a CBR link. A number of uncorrelated, bursty traffic sources are multiplexed together so that the sum of their peak rates exceed the link capacity. Because the sources are uncorrelated, there is a low probability that the sum of their transmit rates will exceed the link capacity. However, although the multiplex can be engineered so that periods of link oversubscription are rare, they will occur. (See Murphy's law!) In data communications networks, periods of oversubscription are accommodated by packet buffering and, in extreme cases, packet discard. (The Internet is a prime example of an oversubscribed, statistically multiplexed network where packet delay and loss may be high during busy periods.)

Video material has a naturally varying information rate—when the scene suddenly changes from an actor sitting at a table to an explosion, the information rate skyrockets. Although MPEG-2 is designed to compensate by encoding more or less detail according to the amount of motion, the encoded bit rate may vary by a ratio of 5 to 1 during a program.

This makes MPEG-2 program streams excellent candidates for statistical multiplexing, except for the fact that MPEG-2 is extremely sensitive to delay and loss. As such, statistical multiplexing cannot be used for MPEG-2 if there is any probability of loss due to oversubscription.

Therefore, statistical multiplexing has been specially modified for use with MPEG by the addition of the following mechanisms:

- A series of real-time encoders are arranged so that their output can be combined by a multiplexer into a single multi-program transport stream (MPTS). Each encoder has a control signal that instructs it to set its target bit-rate to a certain rate.
- The multiplexer monitors the sum of the traffic from all the encoders as it combines them, and in real-time decides whether the bit rate is greater or lower than the transmission link capacity.
- When one encoder has a more challenging scene to compress, it requests that its output rate be allowed to rise. The hope is that one of the other encoders will have less-difficult material and will lower its output rate.

However, there is a significant probability that all the encoders could be called upon to encode a challenging scene at the same time. When this happens, the aggregate bit rate will exceed the link capacity. A conventional statistical multiplexer would discard some packets, but in the case of MPEG-2, this would be disastrous and almost guarantee poor-quality video at the output of the decoders.

Instead, the multiplexer buffers the additional packets and requests that the encoders lower their encoded bit rate. The buffered packets are delayed by only a few milliseconds, but MPEG-2 is extremely sensitive to delay variation. The multiplexer can fix this within limits; as long as the decoder pipeline does not underflow and the timestamps are adjusted to

compensate for the additional time they are buffered, the decoder continues to function normally.

Some statistical multiplexers use a technique called *look ahead statistical multiplexing* (pioneered by DiviCom—see <http://www.divi.com/>). In this technique, the material is encoded or statistics are extracted in a first pass, the information is passed to the multiplexer (while the original input video is passing through a pipeline delay), and bit rates are assigned for each encoder; so when the real encoding happens, a reasonable bit rate is already assigned. This solves some of the nasty feedback issues that can happen in less sophisticated designs.

Reencoding

Until recently, it was impossible to modify an encoded MPEG-2-bit stream in real-time. It is now possible, however, to parse the MPEG-2 syntax and modify it to reduce the bit rate by discarding some of the encoded detail. This technique was pioneered by Imedia Corporation (<http://www.imedia.com/>) and allows the feedback loop between the MPEG encoders and the multiplexer to be removed. In a reencoding (or translation) system, the multiplexer is used to combine a number of variable bit rate MPEG-2 streams. If, at any instant in time, the aggregate bit rate of all of the streams exceeds the transmission link capacity, the multiplexer will *reencode* one or more of the streams to intelligently discard information to reduce their bit rate. Unlike statistical multiplexing, where the multiplexer could not discard any bits, the multiplexer reduces the bit rate by discarding some information—for example, fine detail.

Reencoding is a very useful technique to use whenever a number of multi-program transport streams are groomed—that is, a new output multiplex is formed from program streams taken from several input multiplexes. Without some means of adapting the coded rate, re-multiplexing would result in considerable inefficiency and the output multiplex would contain fewer channels.

A second application of reencoding is in Digital Program Insertion (DPI). DPI *splices* one single-program transport program stream into another so that the viewer is unaware of the transition. It can be used to insert local advertisements into a broadcast program. Reencoding allows the inserted segment to be rate-adapted to the segment that it replaces. DPI is discussed in more detail in Chapter 15, "OpenCable Headend Interfaces."

Although reencoding techniques are extremely useful, feedback-controlled statistical multiplexing is superior from a compression-efficiency perspective when it is possible to collocate encoders and multiplexers. Hence, feedback-controlled statistical multiplexing tends to dominate at original encoding sites that include statistical multiplexing, whereas reencoding is appropriate at nodes where grooming of statistically multiplexed signals needs to be performed.

Audio Compression

Audio compression is a companion to video compression, but the techniques are very different. First, the audio signal requires much less bandwidth than the video signal. For example, a stereo audio pair sampled at 48 KHz and using 16-bit samples requires 1.536 Mbps (compared to 124 Mbps for the video signal). It would be quite possible to send uncompressed audio. Moreover, audio compression cannot achieve the same compression ratio as video; a typical rate for the stereo audio pair is 192 Kbps, an 8:1 compression ratio.

Nevertheless, it is easy to make an economic argument for audio compression. For example, in a 40 Mbps transmission link, an additional two video channels can be carried if audio compression is used (assuming 4 Mbps video).

There are two leading contenders for audio compression for entertainment quality audio: MPEG audio compression and Dolby AC-3.

MPEG-1 Layer 2 (Musicam)

MPEG-1 Layer 2 audio compression, also known as Musicam, is specified in ISO/IEC IS 13818-3. MPEG audio compression delivers near-CD quality audio using a technique called *sub-band encoding*. MPEG audio compression is used mainly in Europe and is used by most direct-to-home satellite providers in the United States.

MPEG audio compression is a two-channel system, but it can encode a Dolby Pro-Logic signal, which includes two additional channels for rear and center speakers. (This is analogous to the way in which a Dolby Pro-Logic signal is carried by a BTSC [Broadcast Television System Committee] signal as part of an NTSC transmission).

Dolby AC-3

Dolby AC-3 is a more advanced system than MPEG audio compression. AC-3 was selected as the audio compression system for digital television in North America and is specified by ATSC Standard A/52. AC-3 encodes up to six discrete channels: left, right, center, left-rear, right-rear, and sub-woofer speakers. The sub-woofer channel carries only low frequencies and is commonly referred to as a 0.1 channel because it has such a limited frequency range. Thus, AC-3 5.1 gets its name from five full channels and a 0.1 channel. In addition, AC-3 has a two-channel mode that can be used to carry stereo and Dolby Pro-Logic encoded signals.

AC-3 also uses sub-band encoding but provides a number of advantages over MPEG audio compression:

- In AC-3 5.1, surround-sound effects are reproduced much more faithfully than is possible with Dolby Pro-Logic because of the additional channels.

- AC-3 5.1 distributes the available transmission link capacity between the 5.1 channels so that more bits are used for those channels containing more information at any particular time. This method effectively makes them statistically multiplexed discrete digital channels.
- AC-3 5.1 is a consumer-grade version of the theatrical Digital Theater Sound (DTS) system, so sound tracks may be directly transferred from the theatrical release.

AC-3 has been chosen as the audio system for digital terrestrial broadcasting and for DVD in the United States. It also has been selected by OpenCable.

Other Audio Compression Algorithms

As with video, many other audio compression algorithms exist. For example, RealAudio is commonly used for sending audio over the Internet, and MP3 has caused a stir in recordable audio. MP3 uses MPEG-1 Layer 3 audio compression, which is considerably more sophisticated than MPEG-1 Layer 2. Many audio compression algorithms have special characteristics that make them suitable for very low bit rate facilities, for software decoding on a PC, and so on. However, it is unlikely that they will pose a significant threat to Dolby AC-3 encoding for entertainment-quality audio in the near future.

Data

Many applications require a data communications path from the headend to the set-top. Part 6 of the MPEG-2 standard (ISO IEC 13818-6) describes how arbitrary data packets may be segmented into MPEG-2 transport packets at the headend and reassembled at the set-top. This mechanism is the basis for the broadcast carousel (see Chapter 8, "Interactive Services").

System Information

System information (SI) is a tabular data format used by the digital receiver device to drive content navigation, tuning, and presentation. See Figure 5-1, which shows that the system information can be sent as part of the in-band digital channel. Cable systems can also transmit system information in an out-of-band channel, as described in Chapter 5, "Adding Digital Television Services to Cable Systems."

There are several standards for system information:

- ATSC A/56—Also known as SI, A/56 was adopted by the cable industry before it was made obsolete by ATSC. DVS-011 and DVS-022 are SCTE standards based on A/56; they are used for basic navigation functions. (Program guide information is sent separately in a proprietary format.)

- DVS-234—Also known as SI, DVS-234 is an SCTE-proposed standard that seeks to establish a common service information standard for out-of-band cable channels. (See the section Out-of-Band Channels in Chapter 16, "OCI-N: the Network Interface.")
- ATSC A/65—Also known as *Program and System Information Protocol (PSIP)*, A/65 has been adopted for terrestrial digital broadcast. In addition to basic navigation functions (such as channel numbering), PSIP can carry content description information that can be used by the programmer to deliver information for electronic program guides. (See the section In-Band Channels in Chapter 16.)

MPEG-2 Systems Layer

The MPEG Systems Committee has defined a systems layer, specified by ISO-IEC 13818-1. (The MPEG-2 systems layer is commonly and confusingly known as MPEG-2 Transport because it defines a transport packet for transmission purposes.) The MPEG-2 systems layer combines the various components of a digital program into a multi-program transport stream. These components include

- Compressed video
- Compressed audio
- Data
- Timing information
- System information
- Conditional access information
- Program-related data

The MPEG-2 systems layer includes the following functions:

- Timing and synchronization—The transmission of timing information in transport packets to allow the receiver to synchronize its decoding rate with the encoder
- Packetization—The segmentation and encapsulation of elementary data streams into 188-byte transport packets.
- Multiplexing—The mechanisms used to combine compressed audio and video streams (elementary streams) into a transport stream.
- Conditional access—Provision for the transmission of conditional access information in the transport stream.

Timing and Synchronization

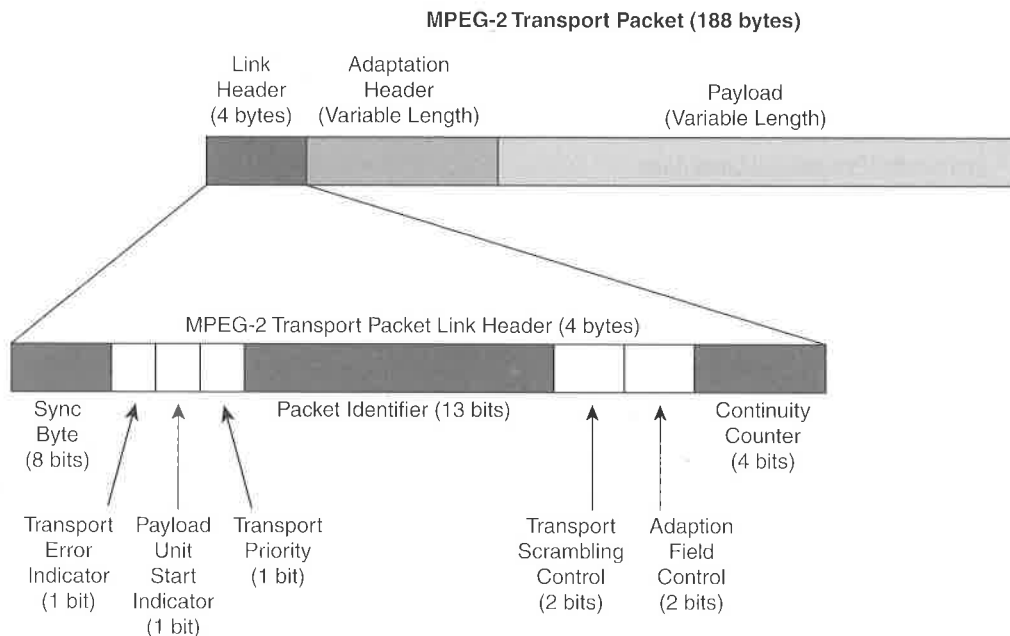
MPEG-2 transport includes timing and synchronization functions that assume a constant-delay network. The system timing information is transferred using program clock reference (PCR) timestamps, which are placed in the adaptation header of certain packets. The PCR timestamps are used by the decoder to synchronize the decoder clock very accurately to the encoder. In many designs, the decoder clock is used to synthesize the NTSC color subcarrier, which must be controlled to an extremely fine tolerance (about 30 Hz).

In addition, the decoder uses the system timing as a reference for presentation and display timestamps, which are used to synchronize audio and video components of a single program transport stream.

Packetization

The MPEG-2 systems layer is packet-based and allows great flexibility in allocating bit rate between a number of program streams. The packets are of a fixed length, and each contains 184 bytes of payload and has a fixed-length, 4-byte link header, as depicted in Figure 4-2. Within each header is a 13-bit packet identifier (PID) that identifies the stream.

Figure 4-2 MPEG-2 Transport Packet Format



MPEG-2 transport packets use an *Asynchronous Transfer Mode*; that is, there is no direct relationship between the MPEG-2 system time and the clock used by the physical link. This approach supports great flexibility in the choice of link layer (see the section “Broadband Transmission,” later in this chapter), but it does make timing recovery more complicated.

Not surprisingly, MPEG-2 packets are very similar to ATM cells:

- MPEG-2 packets are fixed-length.
- The identifier has only local significance on the link.

However, MPEG-2 packets are very different from IP packets:

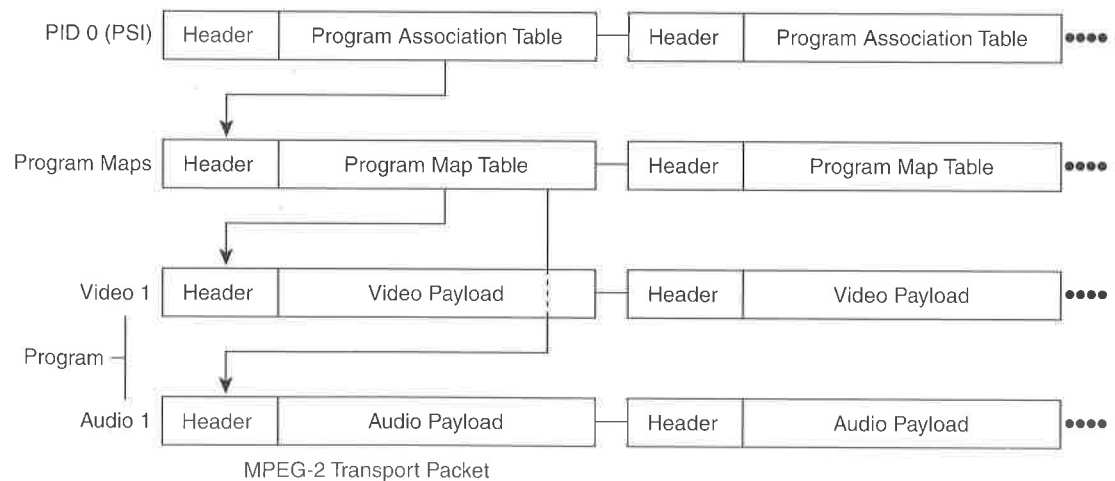
- IP packets are variable-length.
- The identifier (IP address) has global significance in the network.

Multiplexing

The MPEG-2 systems layer supports a two-level hierarchy of multiplexing: the single-program transport stream and the multi-program transport stream.

Individual program elementary streams (PESs) are packetized into MPEG-2 transport packets and multiplexed together to form a single program transport stream. An elementary stream map is included that describes the structure of the MPEG-2 multiplex; the program map table (PMT), as shown in Figure 4-3, tells the decoder which PID values to select for audio and video for that program. A program corresponds to what is traditionally called a channel—for example, HBO, CNN, and so on.

Figure 4-3 Program Association and Program Map Tables



Multiple single-program transport streams are multiplexed together to form a multi-program transport stream (or system multiplex). A program stream map is included to describe the structure of multiplex; the program association table (PAT) has an entry for each program and contains a pointer to the PMT. The PAT is always transmitted using a PID of 0.

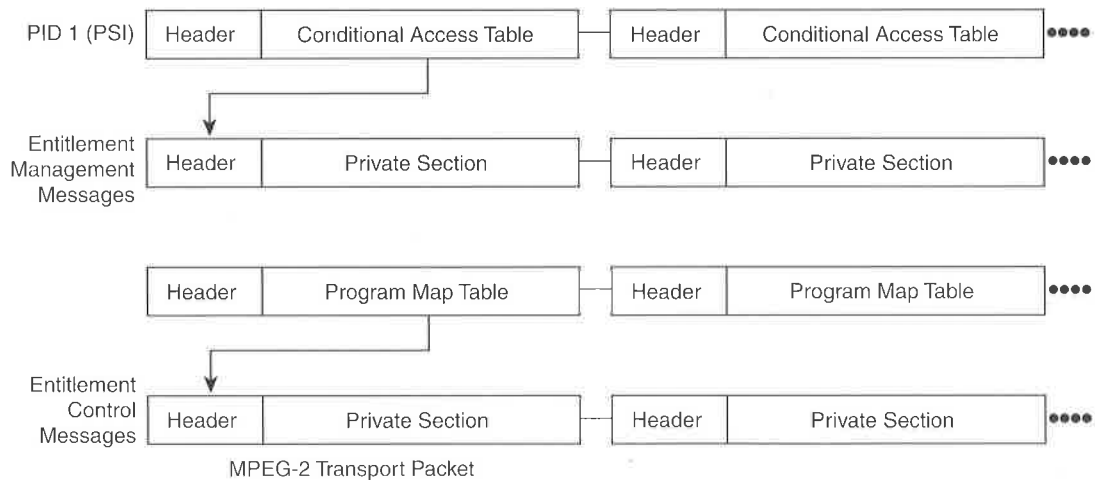
Multi-program transport streams can be combined or split into fragments by extracting the PAT from each stream and reconstructing a new PAT. This re-multiplexing operation is called *grooming*. See Chapter 15, “OpenCable Headend Interfaces,” for more details.

Conditional Access

Part of the MPEG-2 transport stream is used to carry conditional access information. Conditional access is used for security, allowing only authorized decoders to access a video stream. The MPEG-2 systems committee was careful not to specify the conditional access system; instead, it standardized a mechanism that could be used by any conditional access system.

Transport packets with a PID equal to 1 (PID 1) are used to carry the conditional access table (CAT), shown in Figure 4-4. The CAT and the PMT can support multiple conditional access messages, and this mechanism is used for simulcrypt systems in Europe and for the Harmony agreement in North America that allows for dual conditional access (see the article “Multiple Conditional Access Systems,” by Michael Adams and Tony Wasilewski, in *Communications Technology*).

Figure 4-4 Conditional Access Table



Limitations of MPEG-2 Systems Layer

The MPEG-2 systems layer was designed to support constant-delay broadcast networks. MPEG-2 transport packets can be adapted to travel over a communications network (as distinct from a communications link), but the QoS must be tightly controlled. The QoS required by MPEG can be provided by connection-oriented networks (for example, SONET or ATM networks). Connectionless, IP networks are evolving to adopt QoS and traffic engineering capabilities, driven initially by a very strong desire to make practical voice-over-IP; the net result is likely to be reasonable ways to engineer MPEG-over-IP (see *Quality of Service—Delivering QoS on the Internet and in Corporate Networks* by Paul Ferguson and Geoff Huston).

Transmission Mechanisms

There are a number of ways to transmit a multiplexed MPEG-2 transport stream. In some cases, where a dedicated synchronous transmission facility is available (such as SONET), the digital payload can simply be transmitted over a physical channel. In other cases, where it is necessary to transfer the payload over a shared network, higher-level network functions, such as routing and re-multiplexing, are supported by the network.

There are also two main categories of transmission networks: baseband and broadband. In baseband transmission, the entire physical media (twisted-pair, coaxial, laser-link, and so on) is dedicated to the transmission of a bit-stream. This makes transmission very robust but limits overall transmission rates to the maximum speed of a single transmitter or receiver. In broadband networks, each channel is modulated onto a carrier frequency in such a way as not to interfere with any of the other channels. Although broadband networks are not as inherently robust as baseband networks, they achieve a higher transmission capacity at lower cost than baseband networks.

This section covers baseband and broadband transmission as follows:

- Baseband transmission—ATM, SONET, SDH, IP, and DVB ASI provide a baseband transmission facility that can be used to carry MPEG transport streams. This section discusses each of them in turn and compares their suitability.
- Broadband transmission—QPSK, QAM, and VSB are alternative broadband modulation schemes that are commonly used to carry MPEG transport streams. This section discusses their application in cable systems.

Although baseband and broadband transmission are discussed separately, it is possible to layer multiple baseband channels onto a broadband system, and these techniques are by no means exclusive. For example, in the Orlando Full Service Network (see Chapter 11, “On-Demand Cable System Case Studies”), multiple ATM links were modulated and combined into a broadband cable system for delivery to the set-top.

Baseband Transmission

In baseband networks, the MPEG-2 transport packets must be adapted so they can be carried by the network. It was understood by the architects of the Grand Alliance system that it would be necessary to distribute MPEG-2 transport streams over a number of different network types, and considerable effort was spent in choosing an efficient mapping of MPEG-2 packets into ATM.

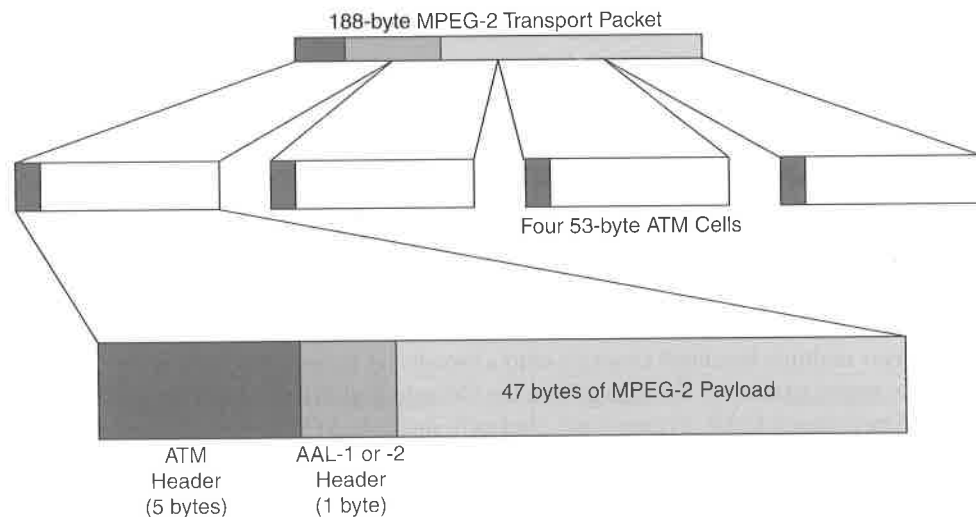
Asynchronous Transfer Mode

Asynchronous transfer mode (ATM) is a connection-oriented network protocol that can be used to build wide-area network (WAN) switched networks. ATM can provide guaranteed QoS metrics, which are sufficient to support MPEG transport streams. However, ATM adds an additional 12% overhead over native MPEG-2 transmission.

The following are characteristics of ATM:

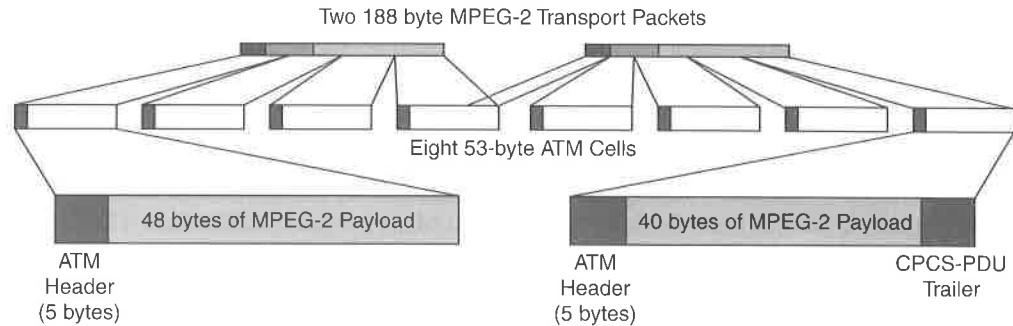
- Multiplexing structure—ATM is a data-link protocol that can carry video, audio, or data in fixed-length cells that carry a 48-byte payload. Each cell has a 5-byte header that is used to identify the connection. ATM supports bandwidth-on-demand—by varying the cell rate, connections of any required bandwidth may be created.
- MPEG-into-ATM mapping—The mapping of MPEG-2 packets into ATM packets was studied by the Grand Alliance (GA), and the MPEG-2 packet size was set at 188 bytes to facilitate mapping into ATM adaptation layer type 1 or 2, as shown in Figure 4-5.

Figure 4-5 Grand Alliance Mapping for AAL Type 1 or 2



Since that time, an alternative adaptation layer, AAL type 5, has become popular in ATM applications, as illustrated in Figure 4-6. AAL type 5 mapping has been nicknamed *straight-8* mapping because it packs two transport packets into eight ATM cells.

Figure 4-6 *Straight-8 Mapping for AAL type 5*



- Error detection and recovery—The header of each ATM cell is protected by a header error check sequence to protect the connection identifier from becoming corrupted. However, error detection of the payload is the responsibility of the ATM adaptation layer. AAL type 1 and 2 do not support error detection, whereas AAL type 5 computes a 32-bit cyclic redundancy check (CRC) across the payload. This provides a valuable error detection mechanism for data, but it is not much use for streaming (that is, video or audio) applications because there is no time to retransmit a corrupted packet.
- Timing—ATM specifies a timing mechanism for synchronous payloads called Synchronous Residual Time Stamp (SRTS). However, this mechanism is designed for telephone carrier emulation (for example, T1 or E1 emulation) and is not applicable to MPEG-2 transmission. Studies have shown that it is possible to transmit MPEG-2 across ATM networks without the need for system timestamp correction if the ATM cell delay variation is tightly controlled (see *Testing Digital Video*, by Dragos Ruiu and others). In addition, it is possible to use timing recovery mechanisms to “de-jitter” an MPEG-2 transport stream carried over ATM before delivering it back to a constant delay transmission system (such as QAM).
- Switching—The ATM cell structure was designed so that switches can be implemented entirely in hardware making multigigabit switches possible. ATM switches are extremely cost-effective for this reason. In video-on-demand applications, a switching function is required to deliver a program stream from a server to a particular set-top. Because MPEG-2 switches are not commercially available, ATM switches are often used to build video-on-demand networks (see the section Time Warner Full Service Network in Chapter 11).

- **Limitations**—ATM adaptation requires additional hardware to map MPEG-2 transport packets into cells at source and then to reassemble the cells into packets at the destination. This adaptation overhead is justified only if a true wide-area switching service is required. Few distribution applications have been deployed that require a switching function, and for broadcast systems the cost of ATM is not justified [Adams].

Synchronous Optical Networks

Synchronous Optical Network (SONET) is a North American standard specified by Bellcore for digital optical transmission. There is an equivalent European standard called the Synchronous digital hierarchy (SDH), which is specified by the International Telecommunications Union (ITU). SONET is a link-layer protocol that carries synchronous payloads in multiples of 50.112 Mbps (within a 51.84 Mbps STS-1). Similarly, SDH carries synchronous payloads in multiples of 150.336 Mbps (within a 155.52 Mbps STM-1). The SONET STS-3c is identical to the SDH STM-1.

The following are characteristics of SONET/SDH:

- **Mapping**—SONET and SDH are ideal candidates for the carriage of MPEG-2 transport streams, but no direct mapping exists for MPEG-2 into SONET or SDH payloads. An alternative approach is to map a multi-program transport stream (for example, the entire payload of a DVB ASI link) into a single ATM Virtual Channel. The ATM Virtual Channel is then mapped into a SONET payload. This approach has the advantage that only a single ATM segmentation and reassembly (SAR) process is required.
- **Error detection and recovery**—Optical networks run error-free if properly maintained, and failures are usually catastrophic due to equipment failure or a fiber cut. For this reason, SONET includes mechanisms for error monitoring and protection switching. A block check is used to monitor the error rate on each SONET link (which is usually zero). If the error rate exceeds a set threshold, a protection switch will be made to a spare link (assuming it is available and its error rate is below threshold). SONET protection switches take less than 50 milliseconds to complete but, despite this, are quite noticeable to a customer watching MPEG-2 compressed video.

Internet Protocol

IP is widely used as a data communications protocol. Recently, there has been considerable interest in using IP to carry telephony, audio, and video services. Although IP was not designed with QoS in mind, there has been considerable effort to provide QoS over IP networks (see *Quality of Service—Delivering QoS on the Internet and in Corporate Networks* by Paul Ferguson and Geoff Huston).

The following provides a brief description of IP:

- **Multiplexing structure**—IP is a network protocol that can carry arbitrary data in variable-length packets. IP supports bandwidth-on-demand—by varying the packet rate, flows of any required bandwidth may be created.

IP's variable-length packet is a disadvantage in delay-sensitive applications, because larger packets can delay shorter packets by taking considerable time to traverse a link. However, as link speed increases, this effect is less noticeable. By using a technique called packet over SONET (PoS) to map IP packets directly into high-bandwidth SONET payloads, delays due to large packets are greatly reduced. (For example, at OC-12 rates of 622 Mbps, a 4 KB packet occupies the link for only 51 microseconds.)

- **Error detection and recovery**—Like ATM, IP's approach to error detection and recovery is datacentric. That is, errors may be detected by computing a CRC-32 across the packet, and if an error is detected the Transmission Control Protocol (TCP) is used to retransmit the packet.

For streaming applications, retransmission is not useful, so user datagram protocol (UDP) is used because it has no payload error checking and requires less overhead than TCP.

- **Routing**—IP routers were first constructed from general purpose minicomputers. However, router design has evolved to the point that routers can be implemented almost entirely in hardware. By placing the packet-forwarding function in hardware, the forwarding delay can be reduced by orders of magnitude, and high-performance routers can approach ATM switches in delay performance.

However, IP routing is fundamentally very different from ATM switching in architectural terms. Classical IP routing is completely connectionless, which means that there is no state knowledge in the IP routers about the packet flow. This makes IP routing very flexible and obviates any need for connection management; however, it also means that by nature it is impossible to predict the load on any particular link or router. Thus, there is a statistical probability that a particular route may become congested, and this can interfere with a particular IP flow. The effect to the user is that video may freeze, or voice may become garbled for some period of time until the congestion clears.

- **Limitations**—IP's ability to support connectionless networks is its Achilles' heel for any application requiring QoS guarantees. Connectionless networks support dynamic reconfiguration by rerouting around congested or failed links in the network. This makes packet delay variation very difficult to control and, as we have seen, MPEG-2 is extremely sensitive to variations in delay. Moreover, IP networks currently do not provide admission control mechanisms, so the only way to obtain bandwidth guarantees is by overprovisioning.

DVB Asynchronous Serial Interface

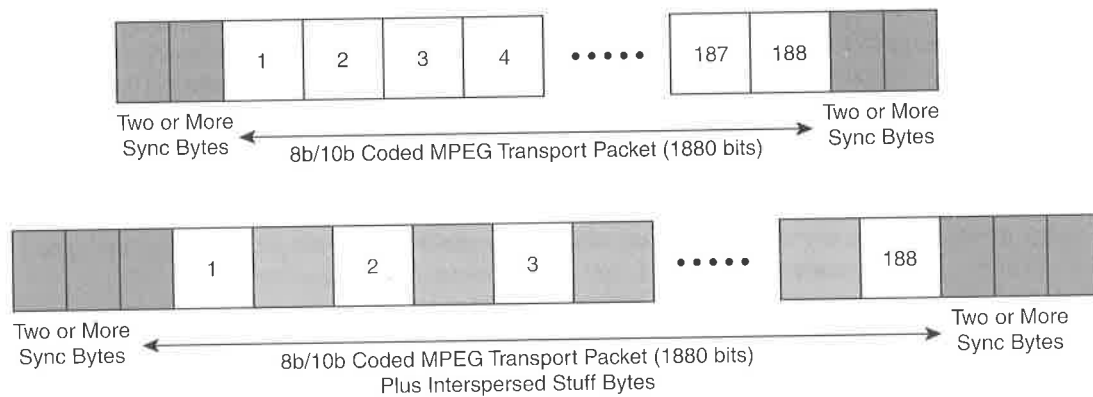
DVB asynchronous serial interface (ASI) was developed for the interconnection of professional MPEG-2 equipment and is a native baseband transmission facility for MPEG-2 transport streams. DVB ASI uses 8b/10b coding at a line rate of 270 Mbps yielding a maximum payload of 216 Mbps. DVB ASI is designed to use two physical media:

- Coaxial cable—Coaxial cable is less expensive than optical fiber and ideal for interconnecting racks of equipment in the headend. However, coaxial cable attenuation limits the reach to about 5 meters.
- Optical fiber—Optical fiber is more expensive than coaxial cable but supports considerably greater reach. Multimode fiber, which has a reach of several kilometers, is typically used, but there is no physical reason why single-mode fiber transceivers (with a reach of up to 100 kilometers) could not be used.

A brief description of DVB ASI's characteristics follows:

- Transmission format—The DVB ASI transmission format is shown in Figure 4-7. Each byte is encoded as 10 bits using 8b/10b coding, and each MPEG transport packet is preceded and trailed by at least 2 synchronization bytes. A packet may be interspersed with an arbitrary number of stuff bytes.
- Error detection and recovery—DVB ASI has no mechanisms for error detection or recovery because it is designed to be for interconnection of equipment over short distances.

Figure 4-7 DVB ASI Transmission Format



- **Timing**—DVB ASI, as its name suggests, is asynchronous. That is, there is no relationship between the line clock and the MPEG system timing. Therefore, each piece of equipment using a DVB ASI input performs timestamping using a high-fidelity local 27 MHz counter. (see the section MPEG-2 Systems Layer, later in this chapter). When MPEG-2 transport packets are output onto the ASI link, they are timestamped again to compensate for any jitter introduced by re-multiplexing.
- **Limitations**—DVB ASI is limited to the interconnection of equipment over short distances because of its lack of physical reach and error protection mechanisms.

Comparison of Baseband Transmission Alternatives

Table 4-3 compares ATM, SONET/SDH, IP, and DVB ASI for the baseband transmission of MPEG-2 transport streams. Although it is unfair, strictly speaking, to compare data link protocols with network protocols, this distinction is lost on the engineer who needs to decide how to move an MPEG-2 transport stream from one location to another.

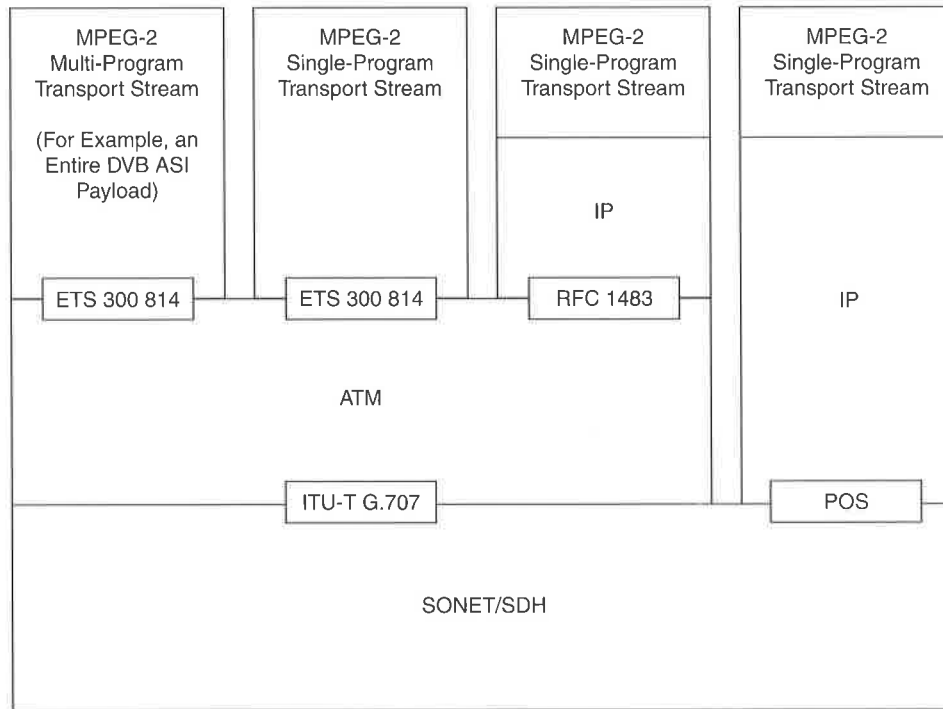
Table 4-3 Comparison of Baseband Transmission Alternatives for MPEG-2 Transport

	ATM	SONET/SDH	IP	DVB ASI
Overhead	12% (5-byte cell header)	3.3%	14% (1 MPEG-2 transport packet per UDP packet)	1% (2 sync bytes per MPEG-2 transport packet)
QoS	Admission control and connection management guarantees	Constant bit rate link	Differentiated class of service	Constant bit rate link
Advantages	Supports switching and integrates voice, video, and data	Supports long-haul optical networks; may also be used to carry ATM, IP, and DVB ASI traffic	Supports routing, ubiquitous network layer for data communications networks	Cost-effective interconnection of headend components
Disadvantages	ATM adaptation cost	Expensive Telco-class service, payloads only in multiples of 52 Mbps	Lack of proven QoS mechanisms for streaming services	Limited reach and lack of switching function

Baseband Transmission Interworking

As discussed in the previous sections on baseband transmission, it is quite common to layer one or more protocols on top of another. Figure 4-8 summarizes the protocol layerings that are, or are in the process of being, standardized.

Figure 4-8 *Standard Protocol Layering*



From left to right, Figure 4-8 illustrates the following mappings:

- An MPEG-2 multi-program transport stream is mapped into a single ATM Virtual Channel, providing transport of an entire DVB ASI payload over an ATM/SONET network.
- An MPEG-2 single-program transport stream is mapped into a single ATM Virtual Channel for transport over an ATM/SONET network.
- An MPEG-2 single-program transport stream is mapped into IP packets. The IP connection is supported by a single ATM Virtual Channel for transport over an ATM/SONET network.
- An MPEG-2 single-program transport stream is mapped into IP packets. The IP connection is supported by a direct packet over an SONET (POS) adaptation layer to provide transport over an SONET network.

In networks, simplicity is usually most cost-effective, and DVB ASI is the simplest and least expensive approach for local interconnection of equipment. The mapping of an MPEG-2 MPTS into a single ATM virtual channel is also gaining favor with cable companies that need to deliver MPEG-2 transport streams over some distance.

Broadband Transmission

Transmission in a broadband system uses modulation to separate each channel into a given frequency band. This technique is often called frequency-division multiplexing (FDM). This section discusses the three common modulation techniques used for MPEG-2 transport in North America: QPSK, QAM, and VSB.

Error correction and protection techniques are typically employed in broadband transmission systems to reduce the number of errors introduced by analog transmission; these are also discussed in this section.

Quaternary Phase Shift Keying

Quaternary phase shift keying (QPSK) modulation is very robust in the presence of noise, so QPSK is used for satellite transmission links and for control channel modulation in cable systems (see the section Out-of-Band Communications in Chapter 5).

Figure 4-9 shows how QPSK modulation is applied to a baseband signal. Two bits are encoded per baud. The 2-bit *symbol* is divided into one in-phase (I) bit and one quadrature-phase (Q) bit, which are each converted to an analog level. These levels are used to modulate the amplitude and phase of a carrier.

Figure 4-9 QPSK Modulator Block Diagram

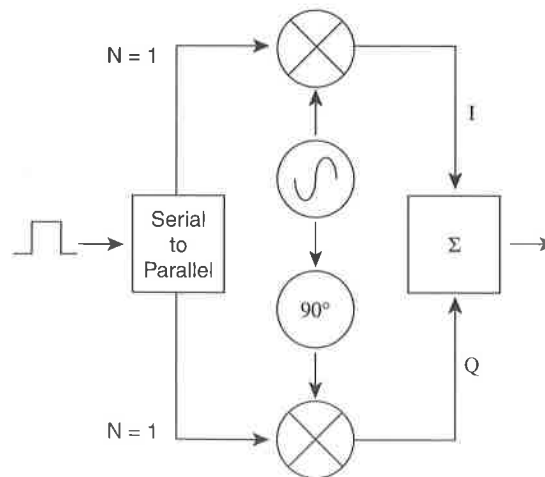
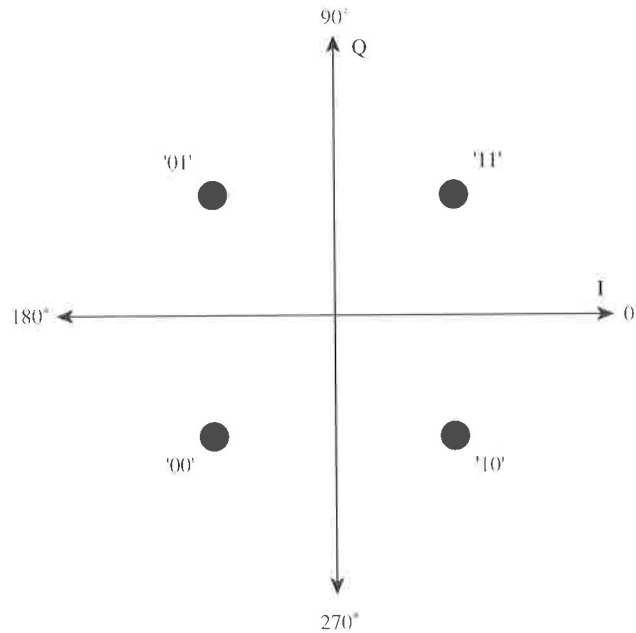


Figure 4-10 shows the constellation diagram for QPSK and the symbol mapping for each phase angle and amplitude vector. QPSK is very robust because the detector needs to detect only two levels and two phase angles to determine the symbol. (QPSK is equivalent to 4-QAM.)

Figure 4-10 QPSK Constellation Diagram



Quadrature Amplitude Modulation

Figure 4-11 shows how 64-QAM modulation is applied to a baseband signal. Six bits are encoded per baud, which is three times as efficient as QPSK. The 6-bit *symbol* is divided into three in-phase (I) bits and three quadrature-phase (Q) bits, which are each converted to an analog level. These levels are used to modulate the amplitude and phase of a carrier.

Using the North American standard for 64-QAM modulation (ITU J.83 Annex B), a payload of approximately 27 Mbps is achieved within a 6 MHz channel. This is an efficiency of 4.5 bits per baud—considerably less than the theoretical maximum of 6 bits per baud. There are two main reasons for this:

- The entire 6 MHz bandwidth cannot be used, and guard-bands need to be introduced on either side of the signal to prevent interference between adjacent channels.
- Error correction and protection mechanisms introduce some overhead (see the section Forward Error Correction, later in this chapter).

Figure 4-11 64-QAM Modulator Block Diagram

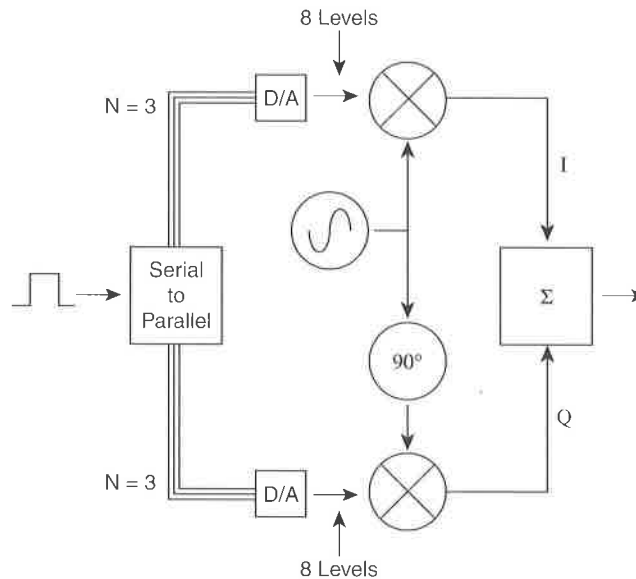


Figure 4-12 shows the constellation diagram for 64-QAM. There are 64 different phase angle and amplitude vectors and, for this reason, 64-QAM is less robust because the detector needs to differentiate between these to determine the symbol. In practice, 64-QAM requires a carrier-to-noise ratio in excess of 22 dB to work (see Chapter 4 of *Modern Cable Television Technology: Video, Voice, and Data Communications* by Walter Ciciora and others, for more details).

Figure 4-13 shows how 256-QAM modulation is applied to a baseband signal. Eight bits are encoded per baud, and this is 33% more efficient than 64-QAM. The 8-bit *symbol* is divided into four in-phase (I) bits and four quadrature-phase (Q) bits, which are each converted to an analog level. These levels are used to modulate the amplitude and phase of a carrier.

Using the North American standard for 256-QAM modulation (ITU J.83 Annex B), a payload of approximately 38.8 Mbps is achieved within a 6 MHz channel. This is an efficiency of 6.47 bits per baud—considerably less than the theoretical maximum of 8 bits per baud. A payload of 38.8 Mbps was chosen because it is sufficient to carry two HDTV channels.

Figure 4-12 64-QAM Constellation Diagram

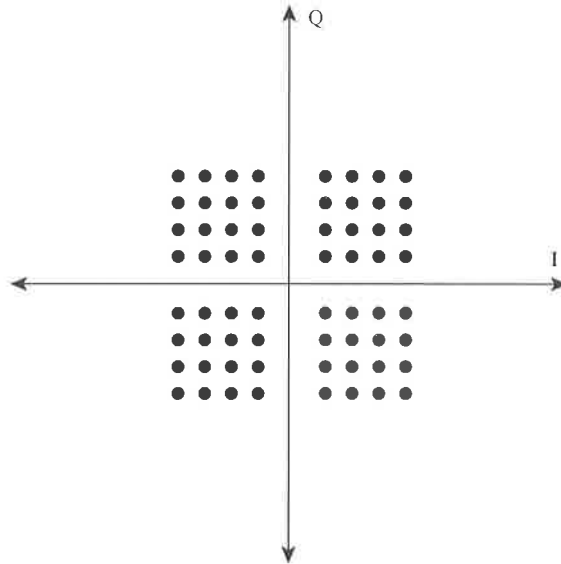


Figure 4-13 256-QAM Modulator Block Diagram

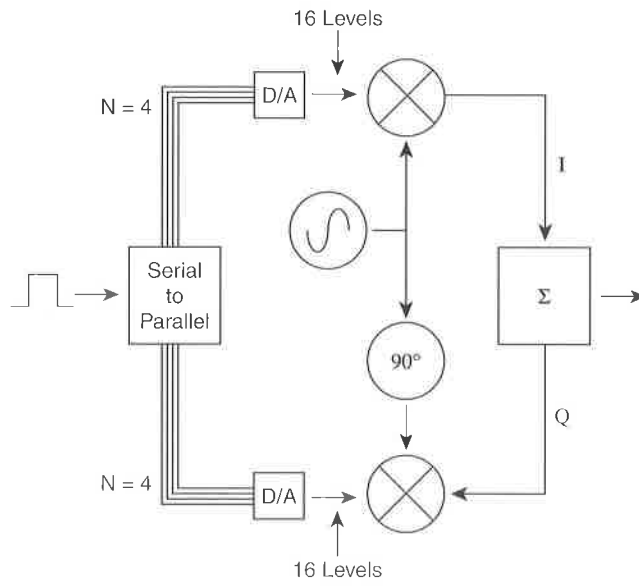
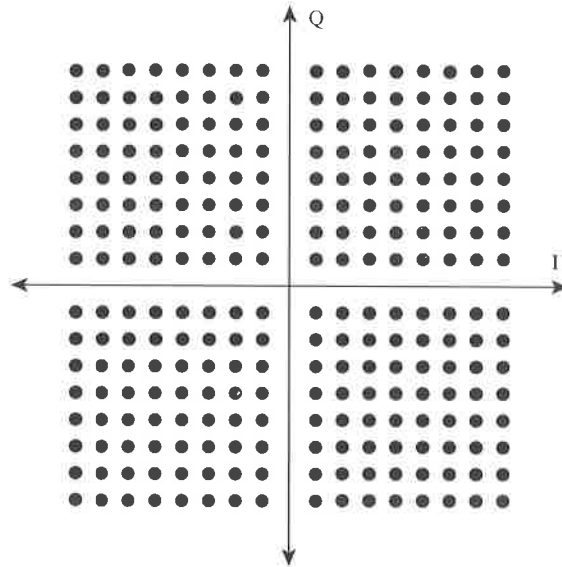


Figure 4-14 shows the constellation diagram for 256-QAM. There are 256 different phase angle and amplitude vectors, making the points on the constellation closer together; for this reason, 256-QAM is less robust than 64-QAM. In practice, 256-QAM requires a carrier-to-noise ratio in excess of 28 dB to work in practice (6 dB more than 64-QAM).

Figure 4-14 256-QAM Constellation Diagram



There has been some discussion of still higher orders of QAM modulation—512-QAM, 768-QAM, or even 1024-QAM. However, there are diminishing returns—1024-QAM increases the payload by only 25% over 256-QAM—and it is unlikely that these will be used in the near future.

Vestigial Side Band

VSB-8 modulation has been adopted for use in terrestrial digital broadcasting (see A.53 Annex D). VSB-8 has a payload of approximately 19.4 Mbps and was designed to carry a single HDTV channel. VSB-8 is a one-dimensional modulation scheme because it uses only amplitude modulation of the carrier (in contrast, QAM is two-dimensional modulation technique because it uses both I and Q components). In VSB-8 modulation, the baseband signal is coded as an 8-level value, so 3 bits are encoded per baud.

Error Correction and Protection

Analog transmission systems are subject to noise, distortion, and interference from other carriers. Therefore, error correction and protection techniques are used to maintain an acceptable error rate.

Three commonly used techniques are discussed in this section: forward error correction, interleaving, and trellis coding.

Forward Error Correction

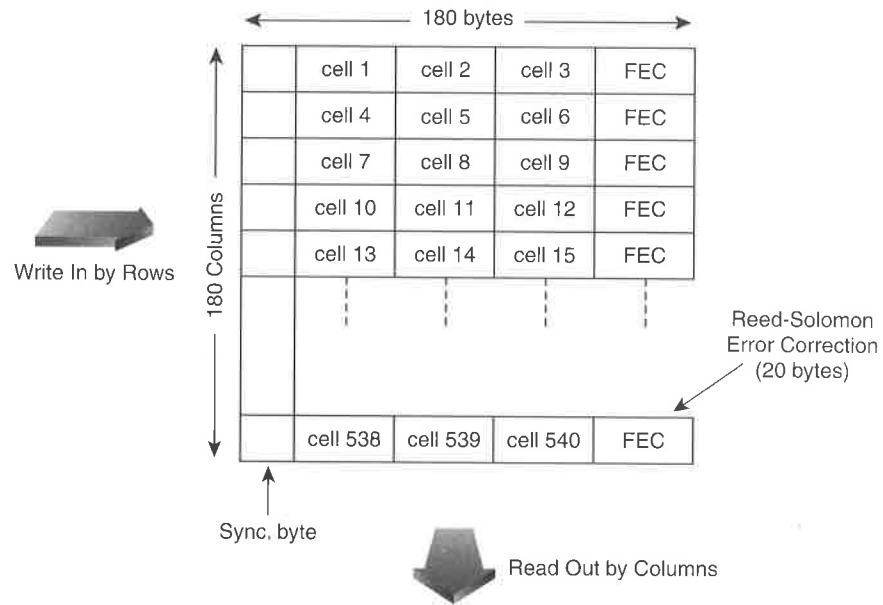
Forward error correction (FEC) uses a mathematical function to generate a check sequence across the payload data. The check sequence is transmitted with the data, and the same mathematical function is used at the receiver to check for payload errors and to correct errors. This technique is also used in error correcting code (ECC) memory.

The Reed Solomon (RS) function is used in conjunction with QPSK, QAM, and VSB modulation. RS (204,188) $t=8$ describes a scheme where 16 check bytes are generated for each MPEG-2 packet. This represents an overhead of 8.5% but allows 1- and 2-byte errors to be corrected by the receiver.

Interleaving

Noise pulses can obliterate a signal for several microseconds, and at a 256-QAM payload rate of 38.8 Mbps, this represents hundreds of bits. By itself, FEC is incapable of correcting such long error sequences.

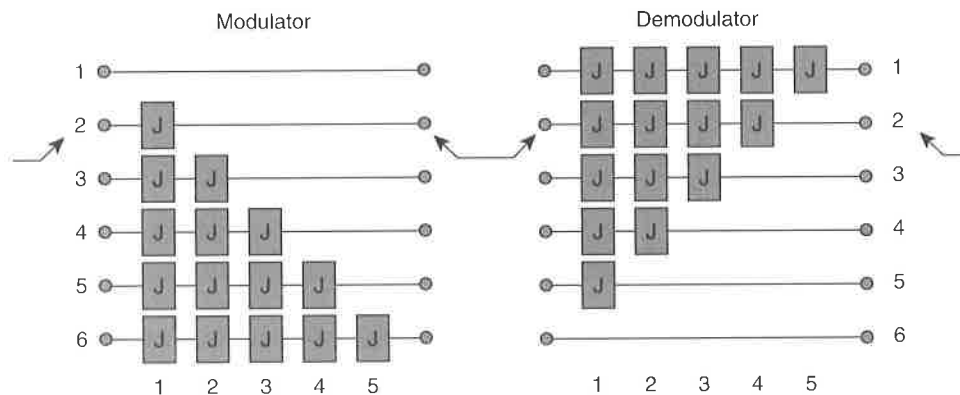
Interleaving effectively spreads the payload data over time. Figure 4-15 shows an example of a block interleaver developed by Scientific Atlanta for the Time Warner Full Service Network (see Chapter 11).

Figure 4-15 Block Interleaver Used in the Time Warner Full Service Network

In this implementation, a noise pulse affects only 1 byte in each row, a situation that can be rectified by the RS correction. The main disadvantage of interleaving is buffer memory and delay, which are 32 KB and 11.5 milliseconds, respectively, in this implementation.

Convolutional interleavers require only half the memory and introduce only half the delay of block interleavers. For this reason, both ITU J.83 Annex B (QAM) and ATSC A/53 Annex D (VSB) use convolutional interleaving. Figure 4-16 shows an example of a convolutional interleaver with an interleave depth of six. The blocks (labeled J) buffer the payload and operate as a shift register. The interleaver (at the modulator) and the deinterleaver (at the demodulator) are synchronized so that the payload is reassembled in its original form.

The effect of interleaving is to distribute the errors due to a noise burst over a period of time so that the errored bits are no longer adjacent, which makes FEC more effective.

Figure 4-16 *Example of a Convolutional Interleaver*

Trellis Coding

As noise is introduced into a QAM or VSB signal, it perturbs the points on the constellation diagram so that ultimately they overlap and the decoder sees the wrong symbol. Trellis coding adds some redundancy and uses sophisticated mathematics at the receiver to determine the best fit of the constellation to a symbol. The trellis coding specified by ITU J.83 Annex B QAM uses a 14/15 coding rate to improve noise immunity by approximately 2 dB. VSB-8, which is designed for more challenging broadcast applications, specifies a 2/3 trellis coding rate, improving noise immunity but reducing payload rate by 33%.

Summary

This chapter introduced a number of important techniques and concepts:

- Video and audio compression algorithms and how they are used to increase the channel capacity of a cable system
- The MPEG-2 systems layer and the multiplexing of multiple program elementary streams into a single physical channel
- Baseband transmission mechanisms for compressed audio and video streams, including ATM, SONET, IP, and DVB ASI
- Broadband transmission mechanisms for compressed audio and video streams, including QPSK, QAM, and VSB modulation

References

Books

Ciciora, Walter, James Farmer, and David Large. *Modern Cable Television Technology; Video, Voice, and Data Communications*. San Francisco, CA: Morgan Kaufmann Publishers Inc., 1999.

Evans, Brian. *Understanding Digital TV—The Route to HDTV*. New York, NY: IEEE Press, 1995.

Ferguson, Paul, and Geoff Huston. *Quality of Service—Delivering QoS on the Internet and in Corporate Networks*. New York, NY: John Wiley and Sons, 1998.

Gibson, Jerry D., Toby Berger, Tom Lookabaugh, Dave Lindberg, and Richard L. Baker. *Digital Compression for Multimedia: Principles and Standards*. San Francisco, CA: Morgan Kaufmann Publishers, Inc., 1998.

Hodge, Winston William. *Interactive Television—A Comprehensive Guide to Multimedia Technologies*. New York, NY: McGraw-Hill, 1994.

Ruiu, Dragos *et al.* *Testing Digital Video*. Hewlett-Packard, 1997.

Whitaker, Jerry C. *DTV: The Revolution in Electronic Imaging*. New York, NY: McGraw-Hill, 1998.

Periodicals

Adams, Michael. "ATM and MPEG-2 in Cable TV Networks, Parts 1 and 2." *Communications Technology*, December 1995 and February 1996.

Adams, Michael, and Tony Wasilewski. "Multiple Conditional Access Systems." *Communications Technology*, April 1997.

Standards

ATSC Standard A/52, *Digital Audio Compression (AC-3)*, 1995.

ATSC Standard A/53, *Digital Television Standard*, 1995.

ATSC Document A/54, *Guide to the Use of the ATSC Digital Television Standard*, 1995.

NOTE The ATSC was formed to establish voluntary technical standards for advanced television systems, including digital high definition television (HDTV). The ATSC is supported by its members, who are subject to certain qualification requirements. ATSC documents are available at the ATSC Web site (see the section "Internet Resources").

Bellcore GR-253-Core Issue 2. *Synchronous Optical Network Standard*. 1995.

ISO/IEC IS 13818-1, International Standard. *MPEG-2 Systems*. 1994.

ISO/IEC IS 13818-2, International Standard. *MPEG-2 Video*. 1994.

ISO/IEC IS 13818-3, International Standard. *MPEG-2 Audio*. 1994.

ITU G-series, *Synchronous Digital Hierarchy Standard*.

Internet Resources

The Advanced Television Systems Committee (ATSC)

<http://www.atsc.org/>

Digital TV: A Cringely Crash Course

<http://www.pbs.org/opb/crashcourse/>

Dolby Web site of technical information on audio compression and related matters

<http://www.dolby.com/tech/>

DVB Web reference

<http://www.dvb.org/>

The Grand Alliance HDTV Specification 2.0

<http://www.sarnoff.com/Papers/hdtv.html>

Adding Digital Television Services to Cable Systems

Cable operators are adding many types of digital services to their systems, including digital television, cable modem services, and telephony. This chapter focuses on digital services delivered through the television receiver. Such services might include broadcast, interactive, and on-demand services, but they have some important characteristics in common:

- The service fits into existing television viewing habits. This is sometimes called the *lean-back* model to contrast it with the *lean-forward* model of personal computer use. For example, the audience expects to be entertained, educated, and informed—viewing often coexists with other social activities, such as conversation, drinking, and eating.
- Video and audio quality are important to the service, to captivate the viewer despite other demands on his or her attention. (In the entertainment world, these are called *production values* and actually extend far beyond technical measures of quality, such as signal-to-noise ratio and video bandwidth.)

For convenience, I will refer to all these services generically as *digital television* to suggest a much broader category than digital picture and sound.

The first part of this chapter discusses the technical and market drivers for digital television. The second part of the chapter concentrates on the necessary enhancements to the cable network to support digital television services, which can be summarized as

- Hybrid fiber coaxial (HFC) upgrade (see Chapter 2, "Analog Cable Technologies")—Strictly speaking, it is not necessary to upgrade the cable system to carry digital television services. However, the additional spectrum required to carry both analog and digital television is a welcome *side effect* of an HFC upgrade.
- Broadband transmission (see Chapter 4, "Digital Technologies")—All cable systems are broadband analog networks, which means that any digital signals must be modulated onto a radio-frequency (RF) carrier before they can be carried by the cable system.

- Out-of-band communications—Cable systems were originally conceived as one-way broadcast networks. Adding a forward out-of-band channel, which is routed (to the distribution hub or fiber node), provides significant operational advantages. Adding a reverse out-of-band channel completes the two-way signaling path, effectively placing each digital set-top on a local area network (LAN).
- Data communications overlay—Interconnecting all the distribution hub LANs to the headend equipment with a data communications overlay network completes the infrastructure required for digital television and also enables interactive and on-demand services (see Part II, “Interactive and On-Demand Services”).

Drivers for Digital Television

Digital channels have some major technical advantages over analog channels:

- Each additional digital channel can carry from 10 to 14 services in the same bandwidth that is occupied by a single analog channel (see the section Video Compression in Chapter 4).
- Digital channels are less sensitive to noise and distortion compared to analog channels, allowing the cable operator to run the digital channels at lower power levels than the analog channels to reduce laser loading and overall distortion (see the section Broadband Transmission in Chapter 4).
- Digital channels can carry data as easily as they carry video and audio. This means that new services can be added relatively easily to the payload (see Chapter 8, “Interactive Services”).
- Premium digital channels are secured using digital cryptographic techniques that are much harder to break than analog scrambling (see Chapter 6, “The Digital Set-Top Converter”).

However, the customer requires a relatively expensive digital set-top to receive digital channels (a capital outlay by the cable operator of about two to three times the cost of an analog set-top converter). Moreover, the customer cannot receive digital channels with an analog cable ready receiver. (See Chapter 18, “OCI-C2: the Security Interface,” for a discussion of cable ready *digital* television.)

This section addresses the drivers for digital television by addressing the following:

- Channel expansion—Digital television squeezes many more channels into a given amount of cable spectrum than analog.
- DBS competition—Matching DBS channel line-ups and providing the all-important “digital picture and sound” sticker.
- High definition television (HDTV)—Compressed digital formats that provide the only way to send HDTV in North America.

- Consolidation with other digital services—Sharing the network infrastructure necessary for digital television with cable modem and telephony services.
- RF return traffic—The aggregation of reverse traffic.
- Business communications—Opportunities to reduce existing data communications costs by using a shared enterprise network.
- Network management—The increased need for network management to provide a highly reliable cable system.

Channel Expansion

It is a fundamental truth that no cable operator can ever have too much bandwidth or too many channels. New channels appear each year (for example, the Radio City Music Hall channel, BBC America, Food Network, Lifetime, OVATION, International Channel, Nick-At-Nite TV Land, Outdoor Channel, Discovery Kids, and Game Show Network, to name just a few), and this trend shows no sign of slowing down. Because each additional analog channel requires an additional 6 MHz of valuable spectrum, cable operators spend significant time and energy considering which channels to carry. (If you visit the Web site of a new channel, you will probably be asked to help lobby for carriage on your local cable system.) In New York, Manhattan Cable already carries more than 100 analog channels.

By using digital channels to carry new services, the pressure on cable system spectrum is reduced. At some future time, it will be possible to collapse existing analog channels into digital channels.

Direct Broadcast Satellite Competition

DirecTV, EchoStar, and others use “*digital* picture and sound quality” to market their services. In fact, poorly compressed digital video can be markedly inferior to analog (see the section Video Compression in Chapter 4). Digital video compression allows the operator to precisely control the amount of quality granted to each video channel without regard to the intervening transmission. This gives a digital operator the flexibility to offer some channels at higher quality than analog and some at lower quality—a flexibility that is quite valuable in terms of maximizing revenue. Because marketing is often a game of “spec”-manship, it is important for cable operators to match the specifications claimed by digital broadcast from satellite (DBS) operators.

High Definition Television

High definition television (HDTV) offers a resolution of up to 1,920 horizontal pixels by 1,080 vertical lines. Multiplying these numbers gives a pixel count of 2,073,600 pixels per frame. This is exactly six times as many pixels as a standard-definition (or full-resolution) picture with 720 horizontal pixels and 480 vertical lines.

HDTV offers a stunning increase in picture quality, and at normal viewing distances is almost as good as the human eye; for this reason, an HDTV picture has been described as “like looking through a window.”

Because of its high bandwidth requirements, digital compression provides the only practical way to deliver HDTV.

HDTV Bandwidth Requirements

Without compression, an HDTV picture would require a bit rate of about 746 Mbps (assuming 8-bit, 4:2:0 sampling—see the section Picture Resolution in Chapter 4). If this signal was analog-modulated with the same efficiency as NTSC, it would require six times the 4.2 MHz video bandwidth of NTSC, or 25.2 MHz. In practice, the European HDMAC-60 system requires 24 MHz and the Japanese Multiple Sub-Nyquist Sampling Encoding (MUSE) system requires 9 or 12 MHz. Each of these systems is designed to use the entire bandwidth of a satellite transponder.

There is insufficient radio frequency spectrum to carry many analog HDTV channels, even on upgraded cable systems. In fact, HDTV is so bandwidth-hungry that no North American standard exists for analog HDTV transmission, and digital television provides the only form of HDTV carriage.

Using MPEG-2 main profile at high level (MP@HL) compression, the bit rate of an HDTV signal can be reduced to 18.8 Mbps, an MPEG compression ratio of nearly 40:1. (This is similar to a typical 31:1 compression ratio for SDTV—see the section Video Compression in Chapter 4.)

The North American standard (ITU J.83 Annex B) for QAM-256 modulation is designed with HDTV in mind. The QAM-256 transport payload is 38.48 Mbps, enough to carry two HDTV channels. Similarly, the VSB-8 (which is the North American terrestrial broadcast modulation standard defined by ATSC A/53) has a transport payload of 19.24 Mbps, enough to carry a single HDTV channel. (See the section Broadband Transmission in Chapter 4.)

Consolidation with Other Digital Services

Cable operators have offered many new service offerings based on digital technology—for example, local-area network (LAN) interconnection, telephony, cable modem services, meter reading, and power management. Some of these services cannot justify their own network infrastructure but generate significant revenue if they ride on a network infrastructure deployed for digital television.

Radio Frequency Return Traffic

Many cable operators have deployed advanced analog converters and are already struggling with an efficient way to aggregate radio-frequency (RF) return traffic and transport it back to the headend controller. Terminating the RF return at the hub and using a digital network to transport the return traffic to the headend is an attractive solution.

Business Communications

Nearly all cable operators have some existing leased or dial-up lines that are used to connect to billing providers and other services. The connection charges on these lines are an ongoing expense, and the ability to run *legacy* data communications, such as IBM's Systems Network Architecture (SNA), over a shared network infrastructure is an important consideration. Like any other business, cable operators are adopting enterprise network approaches to consolidate their internal communications needs to interconnect computers systems, LANs, and PBXs. Some of the larger cable systems span multiple area codes, and operators can save long-distance charges by using their own network infrastructure to carry voice calls.

Network Management

As digital services develop, cable operators have to address a new area: network management. New services being offered require higher availability (for example, telephony), and it is no longer sufficient to react to outages by truck rolls and to maintain headend equipment by responding to customer phone calls. A more proactive approach to network management often requires real-time monitoring of hub equipment, power supplies, and, in some cases, amplifiers. The network management traffic is digital in nature and often uses the Simple Network Management Protocol (SNMP) and requires a digital network overlay on top of the existing analog network.

Transmission of Digital Television

Chapter 2 describes how hybrid fiber coaxial (HFC) upgrades are used to expand the number of analog channels in a cable system. There are two techniques that can be used to transmit a digital television payload to the home:

- The digital payload can be carried all the way to the home in baseband digital form. This technique has been used experimentally (see *Residential Broadband* by George Abe, Chapter 5, which describes Heathrow fiber-to-the-home, NHK fiber-to-the-curb systems). Telcos are forced to deploy fiber closer to the home than cable companies because of the limited transmission distance that is achievable, and over the existing

twisted-pair wiring to the home. Some telcos are still considering fiber-to-the-home, but it will remain uneconomical for some time because it is so expensive to deploy fiber to each home and then to illuminate each fiber with a separate laser.

- Broadband transmission can be used to convert the digital payload into analog form so that an existing analog network can carry it. However, the original digital payload must be recovered (demodulated) by a digital set-top before it can be used.

Not surprisingly, cable operators have chosen to use broadband transmission over their existing analog networks instead of building a separate digital overlay network. This allows the cable operator to maximize the benefits of an HFC upgrade. (See "Digital and Analog Can and Will Coexist," by Walter Ciciora.) Alternatively, some cable operators are using digital transmission to extend their channel lineups in un-upgraded systems (for example, AT&T BIS HITS).

Meanwhile some regional bell operating companies are still experimenting with digital overlay networks, and others are adopting a similar approach to cable (for example, Ameritech) or have deployed MMDS (for example, Bell South).

In practice, a digital set-top is necessary (see Chapter 6) for other purposes than demodulation. Even so, broadband transmission brings its own set of challenges (see the section Broadband Transmission, in Chapter 4).

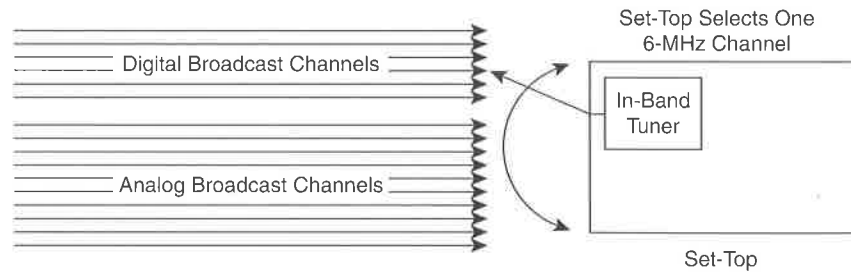
Figure 5-1 illustrates a hybrid analog/digital channel lineup. The set-top converter tunes to any *one* 6 MHz channel at a time with an *in-band* tuner, effectively performing a switching function (more on this in the section QAM Matrix, in Chapter 10, "On-Demand Services"). Although it is possible to build a set-top converter with multiple tuners, most set-tops are single-tuner designs to minimize cost. Using a single tuner to select either an analog or a digital channel causes the set-top to have two distinct operational modes:

- Tuned to analog—The set-top receives a single analog channel and associated VBI data.
- Tuned to digital—The set-top receives a digital program multiplex that includes multiple video and audio and data channels.

Although a single-tuner set-top has access to many in-band digital channels, each with significant data capacity, in-band digital channels cannot be used for reliable data communications because

- The customer can tune the set-top to an analog channel at any time for an arbitrary period, reducing the in-band payload to a trickle (about 9,600 bps per VBI line).
- The customer can rapidly tune the set-top between digital channels, interrupting the data flow and causing the set-top to miss messages.

What is needed is a *dedicated* channel that is always available to support data communications with the set-top; this is the rationale for an *out-of-band* channel.

Figure 5-1 *Set-Top Tuning Model*

Out-of-Band Data Communications

The out-of-band (OOB) channel provides the foundation for a reliable data communications service between the headend and the set-top. This section describes

- Drivers—The motivation for a dedicated OOB channel
- Out-of-band architectures—A summary of the alternatives for the OOB channel
- Forward OOB channel—The forward path (from headend to set-top)
- Reverse OOB channel—The reverse path (from set-top to headend)
- SCTE DVS-178—A description of the OOB channel format developed by General Instrument
- SCTE DVS-167—A description of the OOB channel format adopted by Scientific Atlanta from the DAVIC OOB specification
- DOCSIS cable modem—Applying the DOCSIS standard for in-band data communication to provide an OOB channel for digital television services
- Out-of-band evolution—How the OOB channel evolves to provide a two-way, connectionless datagram service

Drivers

The “one-way” system described so far is sufficient for broadcast services. However, increased digital television revenues can be generated by offering the following:

- Impulse pay-per-view (IPPV)
- Video-on-demand (VOD)
- Interactive enhancements to television
- Standalone interactive services

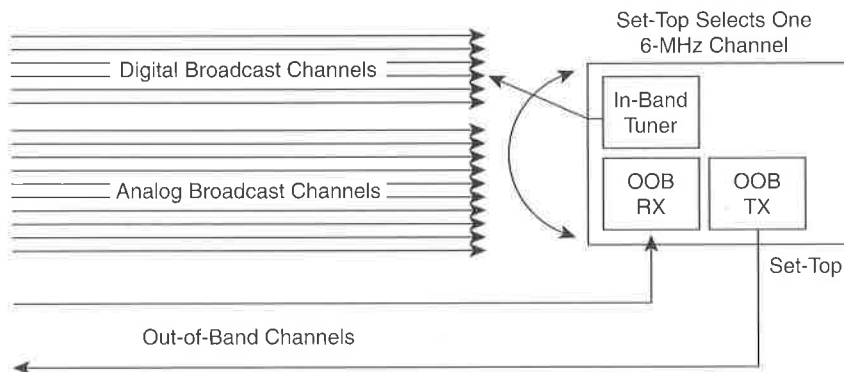
All these services require reliable data communications to the digital set-top.

Out-of-Band Architectures

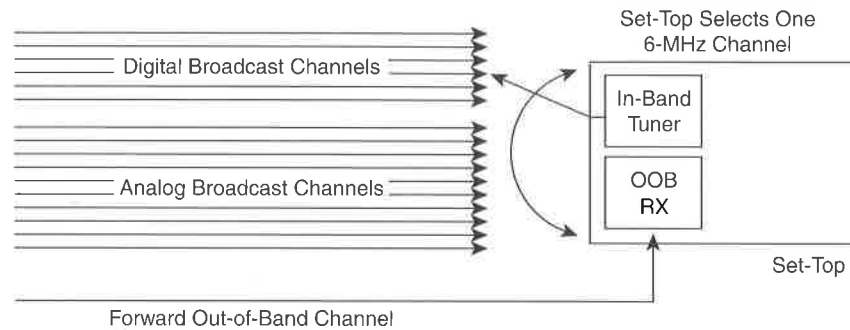
OOB signaling is by no means unique to the cable industry. The ISDN "D" channel is a good example of an out-of-band communications channel that is separated from the *bearer* channels by time-division multiplexing. Other examples include Frame Relay and ATM networks, which use a special link identifier to differentiate signaling from payload traffic. In Signaling System 7 (SS7), telephone systems use entirely different circuits to enable disassociated signaling for increased reliability and to support features such as toll-free services.

In broadband cable systems, frequency-division multiplexing (FDM) is used to separate the out-of-band channels from in-band channels. In Figure 5-2, a dedicated out-of-band receiver (OOB RX) allows the set-top to receive data at any time, independent of which analog or digital channel is selected by the in-band tuner. The *forward OOB channel* supports the flow of data from the headend to the set-top. A cable system with only a forward OOB channel allows one-way, addressed messages to be delivered to the set-top.

Figure 5-2 Forward OOB Channel



In upgraded cable systems, it is common to activate the return path (see the section Return Path Activation in Chapter 2), and these systems can support two-way out-of-band data communications. Advanced analog converters have established the viability of a reverse signaling path for IPPV. Figure 5-3 illustrates the addition of a dedicated out-of-band transmitter (OOB TX) to the digital set-top, allowing it to transmit data on a shared out-of-band carrier frequency.

Figure 5-3 *Reverse OOB Channel*

Because the *reverse OOB channel* is shared by a number of set-tops, a media access control (MAC) protocol is required. The MAC protocol is similar to that used by shared-media Ethernet networks. (One of the first shared-media networks in the world was the Aloha network built in Hawaii. The Aloha network introduced a contention resolution mechanism to allow a number of radio transmitters to share the same radio frequency channel.)

The forward and return out-of-band channels are discussed separately in the following sections, but it is important to think of these channels as a single bidirectional data communications channel.

Forward OOB Channel

The forward OOB channel delivers system messages to the set-top. System messages convey conditional access entitlement information, system information, program-guide information, emergency alert system information, and management commands to the set-top.

System messages can also be delivered in-band, as part of a 6 MHz channel, and this method is still used by some cable systems. However, in-band delivery has some significant disadvantages:

- To guarantee delivery, system messages must be simulcast in every 6 MHz channel because the set-top can be tuned to any channel. Simulcasting consumes a considerable amount of system bandwidth and requires message insertion equipment for every channel. (See the section 8600X in Chapter 3, “The Analog Set-Top Converter,” for an example of the operational impacts when system messages are *not* simulcast).
- An analog NTSC channel has only a very limited capacity (about 9,600 bps per line) to carry data in the vertical blanking interval.

- Broadcast channels, whether they are in-band or out-of-band, repeatedly transmit a circular queue of system messages intended for every set-top in the cable system. (The queue is circular because there is no way to acknowledge that a message has been received by a set-top in a one-way system.) In large systems, this causes considerable queuing delay due to the volume of system messages, often leading to an operational nightmare (see the section Limitations in Chapter 3).
- Digital channels provide a considerable increase in data capacity over analog channels (by a factor of 100). However, system messages must be delivered regardless of whether the set-top is tuned to an analog or digital channel, so it is impossible to take advantage of the increased payload of digital channels. (This problem can be solved, at considerable cost, by building set-tops with separate analog and digital tuners.)

The last point is interesting, because it illustrates why direct broadcast satellite systems have no OOB channel. In DBS systems, all channels are digital, and each one can carry 6 to 12 services. System messages are simulcast on each digital channel at a rate of several hundred kilobits per second, an overhead of approximately 1%.

Not only do cable system designers perceive considerable value in the OOB channel, but they also extend its use in digital systems. Early OOB channel rates were quite limited by the cost and availability of high-speed modulation techniques, but inexpensive silicon is now available that supports multimegabit modulation rates. With higher-rate OOB channels, the system messages occupy only about 10% of the OOB channel capacity. The remaining 90% is earmarked for additional services, including extended program guides, email, and network games (see Chapter 8 for more information on these services).

Reverse OOB Channel

Like the forward OOB channel, the reverse OOB channel was first developed for advanced analog set-tops to support impulse pay-per-view (IPPV) services. These implementations (see the sections 8600X and CFT-2200, in Chapter 3) use very robust modulation, and each set-top can transmit only when polled by the headend controller. For this reason, their use is limited to the periodic collection of IPPV purchases (stored in the set-top until it is polled by the headend controller).

The Orlando Full Service Network (see Chapter 11, "On-Demand Cable System Case Studies") pioneered the use of a high-speed bidirectional OOB channel with a MAC function to allow the set-top to transmit spontaneously without waiting for a poll.

There are at least three digital out-of-band channel options available to the cable system designer: DVS-178, DVS-167, and DOCSIS cable modem.

SCTE DVS-178

DVS-178 was developed by General Instrument and is published as a contribution to the Society of Cable Telecommunications Engineers (SCTE) Digital Video Standards (DVS) subcommittee. DVS-178 was the first OOB channel format deployed and supports millions of first generation DCT-1000 digital set-tops. DVS-178 is designed for polled operation to support IPPV services.

DVS-178 uses a 2.048 Mbps forward channel and a 256 Kbps return channel. The forward and reverse channels employ QPSK modulation for a good balance of robustness and efficiency. The forward channel has a pass-band of 1.8 MHz and is centered at 72.75, 75.25, or 104.2 MHz. The return channel frequency range is from 8.096 to 40.160 MHz in 192 KHz steps. The forward channel uses the MPEG-2 transport stream format to carry system messages, and the return channel uses the ATM cell format. Media access control uses polling and Aloha. (See Table 5-1.)

SCTE DVS-167

The Digital Audio Visual Council (DAVIC) developed an out-of-band communications mechanism as part of the DAVIC systems specification. DVS-167 is extracted verbatim from the DAVIC 1.2, Part 8, "Lower Layer Protocols and Physical Interfaces," Section 7.8, "Passband Bi-directional PHY on Coax." DVS-167 was submitted by Scientific Atlanta as a contribution to the SCTE Digital Video Standards subcommittee. DVS-167 is used by DAVIC-compliant cable set-tops, including the Scientific Atlanta Explorer series, the Pioneer Voyager, and the Philips MediaOne set-top. An advantage of DVS-167 is that it was the first OOB channel standard, allowing multiple vendors to work toward common data communications, and has an efficient MAC layer that can support interactive services.

NOTE

The European Digital Video Broadcast (DVB) standard has also adopted DAVIC OOB signaling.

DVS-167 specifies a 1.544 Mbps or 3.088 Mbps forward OOB channel and a 256 Kbps, 1.544 Mbps, or 3.088 Mbps reverse OOB channel. The forward and reverse channels employ QPSK modulation for a good balance of robustness and efficiency. The forward channel has a pass-band of 1 or 2 MHz and is tunable from 70 to 130 MHz in 250 KHz steps. The return channel is tunable from 8 to 26.5 MHz in 50 KHz steps. (See Table 5-1.)

Table 5-1 *Comparison of DVS-178 and DVS-167*

	DVS-178	DVS-167
Forward OOB Channel	2.048 Mbps	1.544 or 3.088 Mbps
Modulation	QPSK, differential coding for 90° phase invariance	Differentially encoded QPSK
Pass-band	1.8 MHz	1 MHz or 2 MHz
Frequency range	72.75, 75.25, or 104.2 MHz	70–130 MHz
Frequency step size	N/A	250 KHz
Forward error correction	96,94 Reed-Solomon block code, T=1, 8-bit symbols	55,53 Reed Solomon block code, T=1, 8-bit symbols
Framing	Locked to MPEG-TS, two FEC blocks per MPEG packet	Signaling Link Extended Superframe (SL-ESF)
Interleaving	Convolutional (96,8)	Convolutional (55,5)
Reverse OOB Channel	256 Kbps	256 Kbps, 1.544, or 3.088 Mbps
Modulation	Differentially encoded QPSK	Differentially encoded QPSK
Pass-band	192 KHz	200 KHz, 1 MHz, or 2 MHz
Frequency range	8.096–40.160 MHz	8–26.5 MHz
Frequency step size	192 KHz	50 KHz
RF output power range	+24 dBmV–+60 dBmV	+25 dBmV–+53 dBmV

dBmV = decibels with respect to 1 millivolt in a 75-ohm system

The forward channel and reverse channel use a standard ATM cell format. The reverse channel is time-division multiplexed into fixed-length *slots* used by the Media Access Control layer. The DAVIC OOB MAC has three modes:

- *Contention mode* uses slotted-Aloha to opportunistically transmit in a similar way to Ethernet. Collision detection notification, carried by dedicated bits in the forward channel, provides the set-top with an indication of a reverse channel collision with another set-top. A reverse channel collision occurs when two or more set-tops attempt to transmit into the same time slot. Collision detection allows the set-tops to back off and retransmit after a random delay. Contention mode works well for short messages and in low traffic conditions because the probability of collisions is low. Additional modes are included for efficient operation for larger messages and higher traffic conditions.

- *Reservation mode* allows a set-top to request a guaranteed transmit opportunity for larger messages (longer than about 100 bytes). A request is sent in a single slot and a grant message is returned, which enumerates slots that can be used to transmit the message. Thus, there is no possibility of collision in reservation mode, because only one set-top is allocated any given slot—but the slot payload is wasted if the set-top does not use it.
- *Contentionless mode* guarantees a certain number of slots to a connection and is designed for delay-sensitive applications, such as telephony, video, and network games. The contentionless bandwidth must be established at connection set-up time, and any bandwidth that is not used by the set-top is wasted (that is, it cannot be used by another set-top). There is no possibility of collision in this mode.

DOCSIS Cable Modem

The DOCSIS cable modem standard was designed as an in-band mechanism for data transport. (See Chapter 4 of *Modern Cable Television Technology; Video, Voice, and Data Communications* by Walter Ciciora and others.) However, with the addition of a second tuner, the DOCSIS standard can be used to perform all the functions (and more) of DVS-167 and DVS-178. DOCSIS supports 64- or 256-QAM modulation for the forward (downstream) channel and QPSK or 16-QAM modulation for the reverse (upstream) channel, providing better spectral efficiency than DVS-167 or DVS-178. DOCSIS also supports much higher data rates than DVS-167 or DVS-178; the forward data rates are 30.343 or 42.884 Mbps, and the reverse data rates are 0.32, 0.64, 1.28, 2.56, 5.12, or 10.24 Mbps. There are a number of advantages and disadvantages to using DOCSIS as an out-of-band standard, which are summarized in Table 5-2.

Table 5-2 *The Pros and Cons of DOCSIS*

Advantages	Disadvantages
The DOCSIS standard is well established for use in cable modems.	The addition of the 6 MHz tuner and DOCSIS silicon increases the cost of a digital set-top.
Many cable systems are adding support in the hubs to terminate DOCSIS channels.	Some operators have chosen to treat television and cable modem services as separate businesses.
The DOCSIS forward channel offers considerably higher data rates than DVS-167 and DVS-178.	The DOCSIS forward channel consumes an entire 6 MHz channel (this is not a problem if the DOCSIS channel can be shared with cable modems).

Out-of-Band Evolution

OOB signaling has evolved to provide a real-time, two-way data-link between the set-top and termination equipment in the cable system. As with any data communications network, higher-layer protocols are required to support network functions, such as address management, message routing, network management, and the like.

The OOB data channel can use the TCP/IP protocol suite to avoid reinventing new protocols. TCP/IP provides connectionless services to the digital set-top over the OOB channels. Using an IP-based, connectionless network model has a number of advantages:

- System messages can be sent to individual set-tops without the overhead of establishing a connection. This is very important, because there might be thousands of digital set-tops, making connection management difficult.
- Data communications equipment, such as Ethernet hubs and IP routers, are available from multiple vendors. This equipment also provides a convenient way to aggregate return traffic from many hubs.
- Interactive applications (see Chapter 10) are frequently developed on LAN-based systems and then adapted to cable systems. TCP/IP provides a unifying software layer that supports this approach. In fact, interactive cable systems have been compared to giant LANs (see *Digital Cable Systems Really Are Giant LANs* by Hal Benner and Bill Wall).

Out-of-Band Channel Termination

RF modulators and demodulators terminate the OOB channels in the network. These RF modulators and demodulators do not have a generic name, so I will refer to them as the out-of-band termination system (OOBTS). The OOBTS provides functions equivalent to those implemented by the DOCSIS cable modem termination system (CMTS). In fact, the OOBTS is a DOCSIS CMTS, where DOCSIS is used for OOB signaling.

In early implementations, the OOBTS was located at the headend. However, this arrangement has some limitations:

- Large systems are difficult to support because the forward OOB channel becomes congested with system messages for tens, or hundreds, or thousands of set-tops.
- The return traffic must be brought back to the headend as an RF signal, making it difficult to maintain correct input levels to the OOBTS (see Chapter 14 of *Modern Cable Television Technology; Video, Voice, and Data Communications* by Walter Ciciora and others).

It is becoming common to place the OOBTS at the distribution hub. This approach provides better mechanisms for

- Hub-level addressing
- Return traffic aggregation
- Media access control (MAC)

Hub-Level Addressing

Any OOB channel has only limited traffic capacity, so it will eventually suffer message congestion as the cable system increases in size. However, the system can be divided into smaller subscriber groups, each served by its own out-of-band channel. Distribution hubs provide this scalability in the hybrid fiber coax architecture (see Chapter 2). Placing the OOBTS at the distribution hub allows the creation of a routed architecture in which only messages addressed to a set-top in a particular hub are inserted into the out-of-band channel for that hub. Moreover, if additional capacity is required, the distribution hub can be subdivided into still smaller subscriber groups, each with its own out-of-band channel.

At some level of scaling, physical limitations constrain the minimum size of the subscriber group, because the output of the laser transmitter is optically split to serve several fiber nodes. Adding more laser transmitters would solve the problem. Alternatively, additional out-of-band carrier frequencies can be allocated.

Return Traffic Aggregation

Combining return paths in a cable system also combines any noise and ingress from the cable drops. In practice, this funneling effect limits the number of fiber nodes that can be combined to somewhere between 8 and 40, depending on the modulation scheme. For QPSK modulation (as used by DVS-167, DVS-178, and DOCSIS), the return from between 8 and 16 nodes can be combined (depending on how well noise and ingress from the plant is managed). This means that more reverse OOB channels are needed than forward OOB channels. For this reason, DVS-167, DVS-178, and DOCSIS are all designed to support multiple reverse channels per forward channel.

After the return channels are terminated at the hub, they can be multiplexed into a single data channel for transport to the headend using off-the-shelf data communications technologies (for example, SONET, ATM, or IP networking).

Shared Media Access Control

Shared media access control (MAC) is the science of sharing a single channel between a number of traffic sources so that

- Each source (in this case the set-top transmitter) has a fair chance to access the shared media channel.
- Each source obtains a fair share of the available shared media capacity.
- The shared media channel can be used at close to its maximum capacity under all offered loads.

This set of requirements is difficult to achieve because each traffic source acts independently. Shared MAC protocols rely heavily on negative feedback—each source is designed to back off

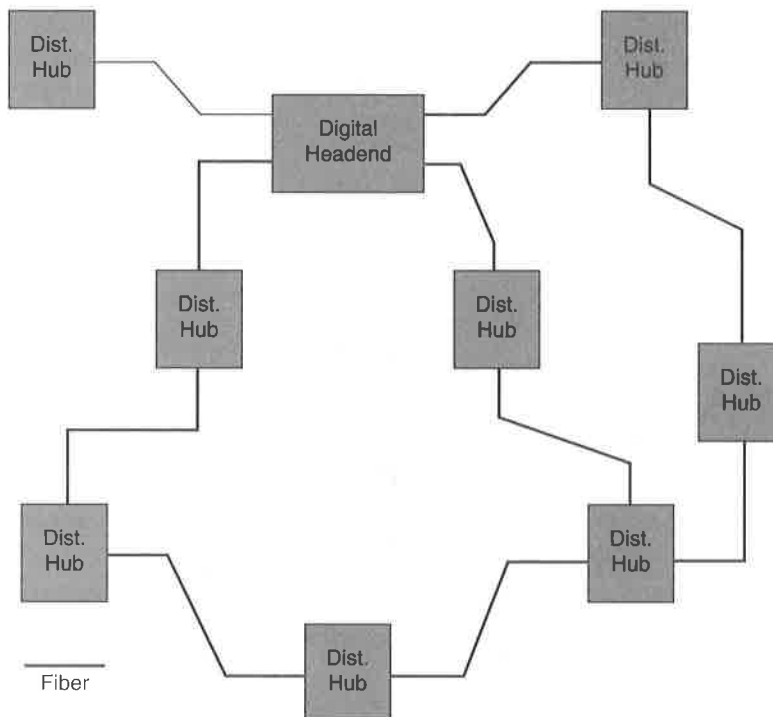
when another source is active. In Ethernet networks, *carrier-sense* is used to detect whether any other source is currently transmitting. However, this technique cannot be used in cable systems because of the return amplifier configuration. Therefore, the source needs feedback from the OOBTS to tell it when another source is transmitting. To provide this feedback signal, the OOBTS continuously sends feedback messages in the forward OOB channel to all sources.

Control theory states that too much delay can cause instability in any feedback loop. In practice, this means that the round-trip delay from the source to the OOBTS must be constrained. Placing the OOBTS at the distribution hub keeps round-trip delay down to an acceptable level.

Headend-to-Distribution Hub Interconnection

When the OOBTS is placed at the distribution hub, a mechanism is required to transport the OOB channel traffic between the headend and the distribution hubs. Figure 5-4 illustrates a typical cable system topology. A digital headend is connected to the distribution hubs by fiber rings in most cases (except for the distribution hub at the upper-left of Figure 5-4, which is connected by a fiber spur). The headend controller (located in the digital headend) requires a data communications service to each distribution hub.

Figure 5-4 Typical Cable System Topology



A number of off-the-shelf data communications solutions are available, which can provide cost-effective headend-to-hub interconnection using the fiber facilities:

- LAN extension products—Point-to-point Ethernet bridge extenders that use fiber facilities.
- SONET—Synchronous Optical Networks were originally developed for high-capacity telephone trunking applications using fiber facilities.
- ATM networks—Asynchronous transfer mode uses SONET facilities to build a connection-oriented network that supports data, video, and voice traffic.
- IP networks—Internet Protocol uses a ATM or SONET facilities to build a connectionless data communications network.

LAN Extension Products

A number of companies build products that can bridge LAN segments over a single fiber strand. These products are relatively inexpensive and provide a plug-and-play solution for connecting 10/100BASE-T devices over distances of up to about 60 fiber miles.

The major drawback of these products is that they provide only point-to-point connections, and this can consume a lot of fiber in cable systems with many hubs. At the headend, a fiber from each hub must be terminated.

SONETs

A SONET ring architecture matches typical cable system topologies quite well. However, SONET has a number of disadvantages for this application:

- The bit rates involved do not justify the cost of a SONET infrastructure in most cases.
- SONET does not perform a traffic aggregation function—it merely *back-hauls* traffic to a central location where a switch or router can perform that function.

In large cable systems, especially those that cover a considerable geographical area, SONET is sometimes justified as a foundation for ATM or IP Networks.

ATM Networks

ATM switches are available with SONET interfaces that support up to 29 dB of optical loss budget. Assuming a figure of 0.35 dB loss per fiber kilometer, these provide a reach of up to 80 kilometers (or 50 miles). Each SONET connection requires a fiber-pair (one fiber per direction).

An ATM network has a number of advantages over point-to-point solutions:

- Ring topologies can be supported that are *self-healing* so that a single fiber cut does not cause an outage.
- ATM provides several traffic classes with different *quality of service* parameters. This allows the ATM network to be shared between the out-of-band communications system and many other traffic types (for example cable modem, telephone, video transport, and so on).
- ATM provides traffic management tools, including rate control and admission control, that can be used to police traffic and prevent oversubscription of network bandwidth.

ATM solutions, however, require a greater capital investment than point-to-point solutions and also require more planning and provisioning.

IP Networks

IP routers are also available with SONET interfaces that support up to 29 dB of optical loss budget. Assuming a figure of 0.35 dB loss per fiber kilometer, these provide a reach of up to 80 kilometers (or 50 miles). Each SONET interface requires a fiber-pair (one fiber per direction).

An IP network provides certain advantages over point-to-point solutions:

- IP networks support arbitrary topologies by means of self-learning routing algorithms. They can discover and use alternative routes so that a single fiber cut does not cause an outage.
- IP provides several traffic classes with different *class of service* parameters. This allows the IP network to be shared between the out-of-band communications system and many other traffic types, for example cable modem, network management, and so on.
- IP provides traffic management tools in the Transport Control Protocol (TCP) that are designed to react to network loss and reduce offered load. Considerable work is also being done on *differentiated classes of service*, which shows considerable promise (see *Quality of Service* by Paul Ferguson and Geoff Huston).

However, IP solutions typically require a greater capital investment than point-to-point solutions and also require more planning and provisioning.

Summary

This chapter discussed how digital television services are added to analog cable systems and introduced the following concepts:

- The drivers for digital television—Including DBS competition, high-definition television, and consolidation with other digital services
- Transmission of digital television—The rationale for broadband transmission of digital payloads over cable systems

- Out-of-band architectures—The need for a reliable data communications network to support advanced digital television services and its implementation, using the forward and reverse out-of-band channels
- Out-of-band channel termination—The advantages of terminating and aggregating the data communication channels (including OOB and DOCSIS cable modem channels) at the distribution hub
- Headend-to-hub interconnection—The various data networking alternatives for interconnecting the headend and hub, including LAN extension, SONET, ATM, and IP

References

Books

Abe, George. *Residential Broadband*. Indianapolis, IN: Cisco Press, 1997.

Ciciora, Walter, James Farmer, and David Large. *Modern Cable Television Technology; Video, Voice, and Data Communications*. San Francisco, CA: Morgan Kaufmann Publishers, Inc., 1999.

Ferguson, Paul, and Geoff Huston. *Quality of Service: Delivering QoS on the Internet and in Corporate Networks*. New York, NY: John Wiley and Sons, 1998.

Periodicals

Benner, Hal, and Bill Wall. "Digital Cable Systems Really Are Giant LANs." *Communications Technology*, October 1998.

Ciciora, Walter. "Digital and Analog Can and Will Coexist." *Communications Technology*, January 1999.

Internet Resources

ASTC standards Web site

<http://www.atsc.org/stan&rps.html>.

Hi-Vision overview available at the NHK Web site

<http://www.nhk.or.jp/hi-vision/hivi-e.html>.

The Digital Set-Top Converter

Digital set-top converters are evolving rapidly, and it is impossible to predict what capabilities they will possess in a few years' time. However, they can be divided into two major categories:

- Basic digital set-tops perform tuning, demodulation, conditional access, and decoding functions. This class of set-top provides a digital equivalent of the advanced analog set-top converter, providing similar electronic program guide and impulse pay-per-view functions for broadcast digital services.
- Advanced digital set-top converters perform all the functions of a basic digital set-top, and provide support for two-way, real-time data communications over the cable plant and support for more advanced applications.

Although this chapter focuses on the *basic* digital set-top, it also introduces some of the functional extensions of advanced digital set-tops. Cable operators are deploying digital set-tops for two main reasons:

- Channel expansion—Digital compression can greatly expand the number of channels in a cable system by delivering 6 to 10 programs in a single 6 MHz channel. For example, TCI (now AT&T BIS) has pursued a digital channel expansion strategy known as Head-End in The Sky (HITS) to avoid the cost of HFC upgrades.
- Competitive response to direct broadcast satellite (DBS)—DBS providers chose a digital system because it provides more channels than analog. They have aggressively marketed the "digital picture and sound quality" of their service. Although an upgraded hybrid fiber coax (HFC) system can rival the quality of a digital system, analog is beginning to sound old-fashioned to many customers. Thus, cable operators are forced to offer digital television to keep up with DBS competition.

There are many reasons to deploy more advanced digital set-tops, and these are described in Part II, "Interactive and On-Demand Services," but the basic digital set-top provides similar broadcast services to the advanced analog set-top—only digital.

Cable Environment

In contrast to analog cable, where fewer than 50% of cable customers have a set-top, a digital set-top is essential to provide digital services. In fact, a digital set-top converts the digital television picture into an analog NTSC picture to allow customers to use their existing analog television receivers and VCRs.

NOTE

This situation will start to change in July 2000 with the availability of digital navigation devices. Customers will be able to buy digital set-tops or cable ready digital televisions that are capable of receiving premium services. See Chapter 18, "OCI-C2: the Security Interface."

The digital set-top has evolved directly from the advanced analog set-top (see Chapter 2, "Analog Cable Technologies"). In fact, to be successfully deployed in cable systems, the digital set-top is actually a hybrid analog/digital set-top. That is, the set-top must be able to support all existing analog services, as well as the new digital services. There are a number of reasons for this:

- The digital set-top is deployed side-by-side with existing analog set-tops (often in the same house), so cable customers naturally expect the digital set-top to provide all the familiar (analog) services as well as new (digital) services.
- This perception is reinforced by most service tier structures that bundle all analog services into the new digital tier. In other words, the customer is never offered a *digital-only* package.
- The digital set-top is considerably more expensive than an analog set-top, so the cable operator naturally expects it to perform all the functions of an analog set-top in addition to its new digital functions.

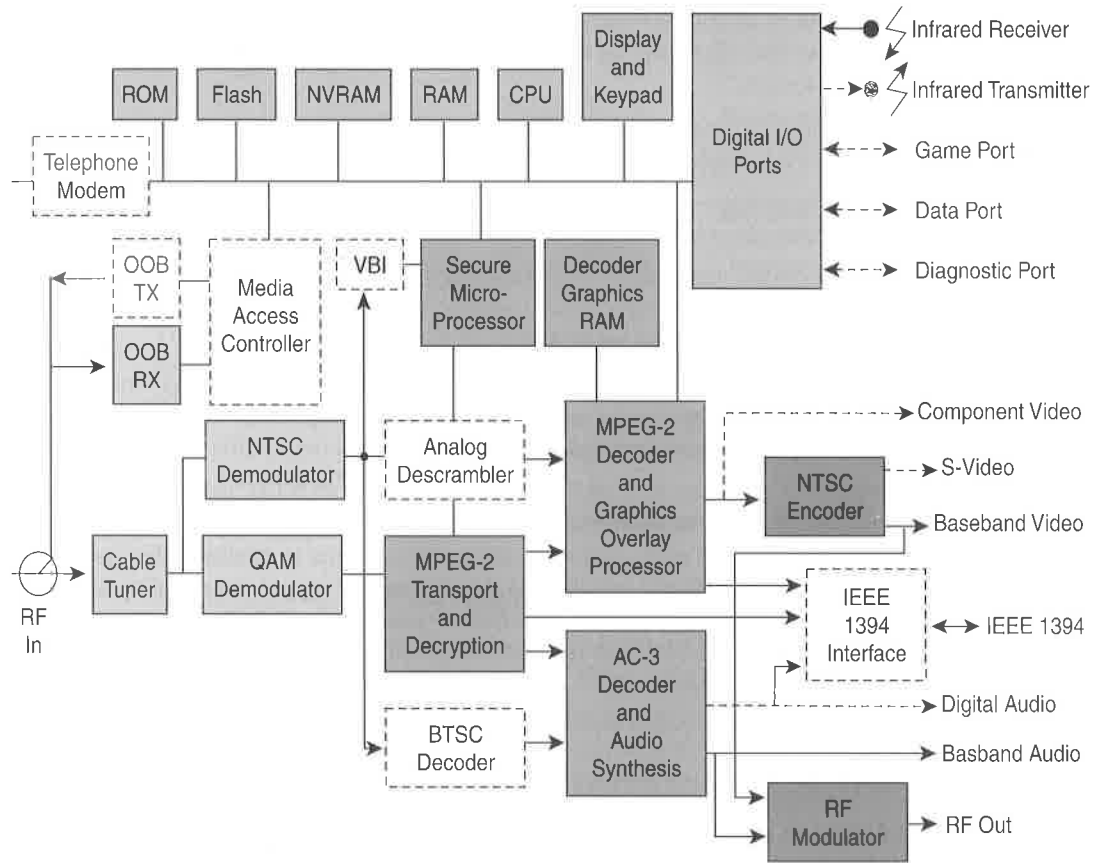
Although it is true that a digital-only set-top costs less to build than a hybrid analog/digital set-top, deploying a digital-only set-top requires that all existing analog services are *simulcast* in digital form. This approach not only requires more bandwidth but also involves significant capital investment at the headend for MPEG-2 encoders and multiplexers (see the section Headend Equipment in Chapter 7, "Digital Broadcast Case Studies," for more details).

For these reasons, this chapter focuses on the hybrid analog/digital set-top, which will be referred to as the *digital set-top*. (Please refer to Chapter 3, "The Analog Set-Top Converter," for information on the analog functions of a hybrid analog/digital set-top.)

Overview

Figure 6-1 shows the block diagram of the digital set-top. Although there are a number of different implementation approaches, this block diagram describes most digital set-tops. Optional blocks are shown using broken lines.

Figure 6-1 Block Diagram for a Typical Digital Set-Top



The next sections consider Figure 6-1 in detail after breaking it down into its major subsystems:

- The cable network interface (CNI)—All set-tops have a 75 ohm, F-connector, which is the interface between the cable system and the set-top (see Chapter 16, "OCI-N: The Network Interface"). The cable network interface circuitry in the set-top includes

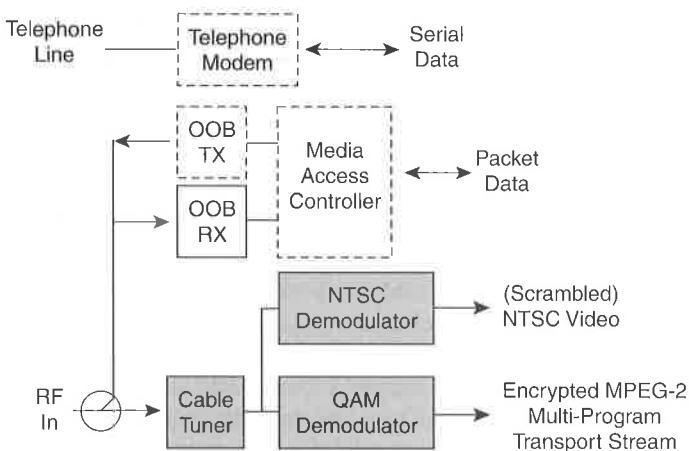
the cable tuner, NTSC demodulator, QAM demodulator, and an out-of-band (OOB) receiver. Optionally, an OOB transmitter, a media access controller, and a VBI decoder might also be included. When the cable system does not support an RF return path, a telephone modem is an option (in place of the OOB transmitter).

- **Conditional access**—Conditional access is the means by which access to specific services is granted to the customer, based on payment for those services. A digital decryption circuit, in conjunction with a secure microprocessor, performs the equivalent of analog de-scrambling for digital services. Optionally, digital set-tops might also include an analog de-scrambler that shares the same secure microprocessor.
- **Video and graphics processing**—This subsystem is responsible for MPEG-2 decoding and on-screen display generation and usually has some dedicated RAM.
- **Audio processing**—This subsystem includes an AC-3 decoder and, optionally, a BTSC stereo decoder and some means for audio synthesis.
- **Microprocessor subsystem**—The microprocessor subsystem is the brains of the set-top, and includes a microprocessor together with ROM, Flash, NVRAM, RAM memory, an LED display, and keypad.
- **Inputs**—User control signals are received via an infrared (IR) receiver from an IR remote control or (optionally) an IR keyboard. Other (optional) inputs include a data port, a diagnostic port, and a game port.
- **Outputs**—All set-tops have an F-connector output for an NTSC channel modulated at broadcast channel 3 or channel 4. In addition, the set-top has *baseband* (or *composite*) video and baseband audio outputs to allow the customer to realize higher-quality picture and sound than is possible via the channel 3/4 output. A host of optional outputs are designed to provide even higher-quality signals, including component video (for example, YPrPb or RGB), S-Video, digital audio (S/PDIF) and the IEEE-1394 digital interface.

The Cable Network Interface

Figure 6-2 shows the *cable network interface (CNI)* for a digital set-top. It is a superset of the interface for an analog set-top (see Chapter 3). Whereas in the analog world there are standards (NTSC, BTSC and EIA 542) that have been established over many years of practice, in the digital world most standards have been developed by standards committees.

Figure 6-2 Cable Network Interface



The digital modulation scheme is chosen to fit into the same 6 MHz frequency slots as specified by EIA 542, originally for analog channels. This approach has two major advantages:

- Digital and analog channels can be mixed within the same system without any loss in spectrum efficiency.
- A single 6 MHz tuner is used to tune both types of channel, thereby reducing set-top cost.

The standard aspects of the cable network interface are

- **ATSC Digital Television Standard (A/53)**—This standard describes the overall system characteristics of the U.S. Advanced Television System.
- **RF Interface Specification for Television Receiving Devices and Cable Television Systems (EIA 23)**—This standard provides RF performance recommendations for all receiving devices that might be directly connected to a cable television system, including the digital set-top.
- **Cable Television Channel Identification Plan (EIA 542)**—This standard specifies the frequencies of all cable channels. (See Chapter 8 of *Modern Cable Television Technology; Video, Voice, and Data Communications* by Walter Ciciora and others.)
- **Digital Transmission Standard for Cable Television (SCTE DVS-031)**—This standard specifies the modulation (64-QAM or 256-QAM), forward error correction, and framing of the digital payload (see Chapter 4, "Digital Technologies").
- **Draft Digital Video Service Multiplex and Transport System Standard for Cable Television (originally DVS-093, now updated as DVS-241)**—This standard references the MPEG-2 systems layer (see Chapter 4).

Another aspect of the cable network interface is called system (or service) information (SI). This provides information on channel numbering and allows basic navigation of the cable channels. There are currently two leading proposals for system information:

- Out-of-band—Service Information for Out-of-Band Digital Cable Television (SCTE DVS-234). This standard is based on ATSC A/56 and is used by both DCT series and Pegasus digital set-tops (see Chapter 7).
- In-band—Program and System Information Protocol for Terrestrial Broadcast and Cable (SCTE DVS-097). This standard is based on ATSC A/65 and is the terrestrial broadcast standard for system information.

The proprietary aspects of the cable network interface are the OOB data communication channels. A proprietary OOB channel format developed by General Instrument has been deployed in North America (see the section DVS-178 in Chapter 5, "Adding Digital Television Services to Cable Systems"). The DAVIC standard for OOB channels has also been deployed in North America by several vendors, including Scientific Atlanta, DiviCom, Pioneer, and Philips (see the section DVS-167 in Chapter 5).

The set-top components that terminate the cable network interface are described in the next sections. When compared to the analog design, the major addition is the QAM demodulator.

Cable Input

The cable input is a 75 ohm female F-connector. It is the only radio frequency input to the shielded set-top unit. Shielding is important to prevent off-air television broadcasts from interfering with the cable channels.

Tuner

Tuning requirements for digital set-tops are fundamentally the same as for analog, except that a 860-MHz tuner is usually specified because digital channels are commonly placed in the 550 to 860 MHz range above the highest analog channels. Digital tuner performance is similar to analog tuner performance, with the exception that phase noise must be tightly controlled to support the higher symbol rate used by 256-QAM modulation.

QAM Demodulator

The block labeled QAM Demodulator in Figure 6-2 is actually responsible for a number of related transmission functions:

- Adaptive equalization of the received signal to compensate for reflections introduced by the cable plant

- 64-QAM or 256-QAM demodulation according to the modulation type of the channel tuned
- Trellis code interpolation
- De-interleaving
- Forward error correction (FEC)

The output of the demodulator is a baseband digital stream at 26.97035 or 38.81070 Mbps for 64-QAM or 256-QAM modulation, respectively. (See the section Broadband Transmission in Chapter 4.)

NTSC Demodulator

The NTSC demodulator is a standard component from the analog set-top. In addition, many digital set-tops include an analog-to-digital (A/D) converter to digitize the demodulated video so that it can be manipulated by the graphics engine (see the section Analog Video Digitization later in this chapter).

Out-of-Band Channel Termination

There are two alternative out-of-band channel specifications currently being used in North America:

- SCTE DVS 167, as specified by General Instrument
- SCTE DVS 178, as specified by Scientific Atlanta and based on the DAVIC OOB

Please refer to the section Out-of-Band Data Communications in Chapter 5 for a detailed description of the out-of-band channel.

Out-of-Band Receiver

The out-of-band channel can be one-way or two-way. An out-of-band receiver is required in either case, and all digital set-tops include an out-of-band receiver. The out-of-band channel is of greatly increased capacity—in the order of megabits per second—compared to analog systems.

Out-of-Band Transmitter

An out-of-band transmitter is required for two-way operation of the out-of-band channel. Two-way signaling supports services such as impulse pay-per-view, video-on-demand, and interactive applications.

An out-of-band transmitter is optional for set-tops deployed in a one-way cable plant, although if an upgrade to two-way is planned it makes good sense to deploy a set-top with an OOB transmitter.

Media Access Control

The *Media Access Control (MAC)* function supports the out-of-band transmitter and allows a number of set-tops to share a common return path. The shared media access control can be quite sophisticated (see the section SCTE DVS-167 in Chapter 5).

Telephone Modem

The telephone modem provides an alternative to the out-of-band transmitter in a one-way plant. The modem is a standard data modem and usually supports a data rate from 300 to 9600 bps. The telephone modem is used to periodically call the headend modem bank to report purchase information. The call time is usually set to 2 or 3 a.m., so a separate phone line is not required. Unfortunately, this means that no interactive services can be supported, due to the lack of two-way, real-time data communications.

Transport Processing

Transport processing is required to break the MPEG-2 system multiplex into the specific packetized elementary streams (PES) that are required for the selected service (see the section MPEG-2 Systems Layer in Chapter 4). Transport processing is linked into the conditional access system because different services in a multiplex are typically encrypted using different working keys.

Transport processing can be broken down into a number of steps:

- Filter packet identifier (PID) 0 and process the MPEG-2 program association table (PAT) contained in it to discover what program transport streams are contained in the system multiplex.
- Obtain the program map PID for the desired program transport stream. Filter the program map PID and process the program map table (PMT) contained in it.
- Filter the packetized elementary stream PIDs for the selected program transport stream and route them to the appropriate video or audio decoders via the payload decryption engines (see Figure 6-3).
- Filter the entitlement control message (ECM) PID for the selected program stream and pass the ECMs to the secure microprocessor. (The PAT and PMT are described in the section Multiplexing in Chapter 4.)

Conditional Access System

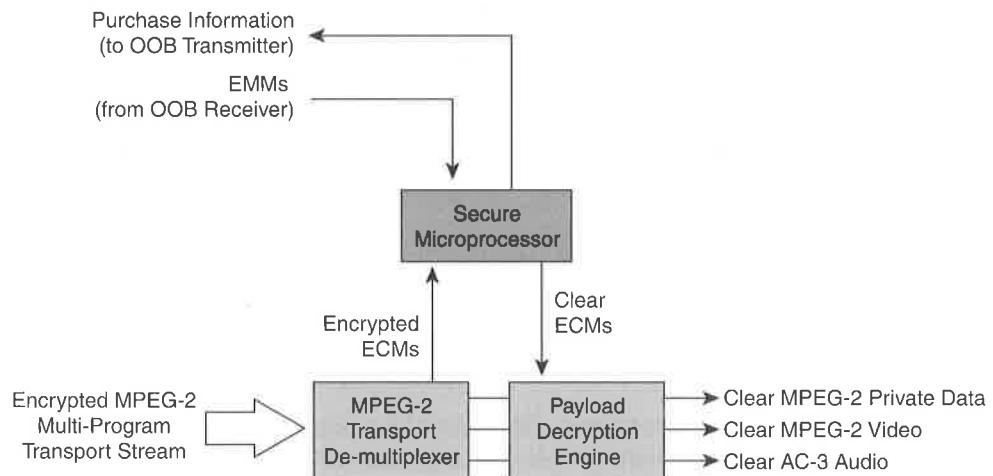
The conditional access system provides the means by which access to specific services is granted to the customer based on payment for those services. The conditional access system uses embedded security components in the set-top—the MPEG-2 transport and decryption block and the secure microprocessor, shown in Figure 6-1. An optional analog de-scrambler provides backward-compatibility for scrambled analog channels.

Every aspect of the digital conditional access system is highly proprietary. It is common for set-top purchase contracts to include remedies in the case of security breach, which might include upgrading or replacing security elements. Thus, the set-top vendor has every incentive to make the conditional access system as obscure and complicated as possible in an effort to defeat signal theft.

Conditional access systems from General Instrument (Digicipher II), Scientific Atlanta (PowerKEY), and SECA (MediaGuard) have been deployed by the cable industry in North America. Nevertheless, all these digital conditional access systems have common design strategies that originate in the work done by the MPEG Systems Committee. The MPEG-2 transport stream is used to carry conditional access information, and the MPEG-2 systems layer is designed to support any conditional access system.

Figure 6-3 illustrates the basic principles of operation of the digital conditional access system.

Figure 6-3 *Conditional Access System*



The following describes the operation of the digital conditional access system:

- The MPEG transport de-multiplexer separates the transport stream into separate streams based on the packet identifier (PID), and selects the video, audio, and data packetized elementary streams for the selected service.
- The entitlement control messages (ECMs), which carry the encryption keys for the service are sent to the secure microprocessor. The ECMs also are encrypted; they are deciphered by the secure microprocessor before being sent to the payload decryption engine.
- The payload decryption engine decipheres the packetized elementary streams using the encryption keys supplied by the secure microprocessor.

The encryption keys are changed frequently (every few seconds) to reduce the chance that they can be discovered. The secure microprocessor can decipher only encryption keys for authorized services.

Digital Decryption

In a digital set-top, conditional access is achieved by using digital encryption to *scramble* the MPEG-2 transport stream. The payload decryption engine (refer to Figure 6-3) is responsible for deciphering the MPEG payload. Three digital encryption ciphers are commonly used by different systems:

- The *Harmony* Cipher—The Harmony Cipher was developed by General Instrument and is used in the Digicipher II conditional access system. It is based on the Data Encryption Standard (DES) cipher called cipher block chaining (CBC). The Harmony Agreement between General Instrument and Scientific Atlanta, allows PowerKEY to use the same Harmony Cipher as Digicipher II.
- The DVB common scrambling algorithm (CSA)—CSA is part of the simulcrypt system and is a closely held cipher that is licensed by ETSI to *bona fide* conditional access vendors.
- DES electronic code book (ECB)—The DES ECB cipher is supported by the PowerKEY conditional access system. It is a published ANSI standard.

Secure Microprocessor

The secure microprocessor provides the brains of the conditional access system and provides a number of important conditional access system functions:

- The secure microprocessor allows selective access to premium services under the control of the headend controller.
- The secure microprocessor receives entitlement management messages (EMM) from the out-of-band channel, as shown in Figure 6-3. The EMMs are encrypted and the secure microprocessor decrypts them to recover a *session* key.

- The secure microprocessor decrypts each ECM using the session key to recover a *working* key. The working key is sent to the payload decryption engine.
- The secure microprocessor stores purchase information until the headend controller collects it (via the out-of-band channel or telephone modem).

The secure microprocessor is actually a single-chip, secure microcontroller. The secure microprocessor is fabricated using a semiconductor process that makes it difficult to analyze the program and data contained in it. This physical security is essential because the secure microprocessor contains the secrets that secure the conditional access system. If a pirate can copy (or *clone*) the secure microprocessor, the condition access system is compromised.

Security Upgrade

All digital set-tops contain some means of upgrading the conditional access system without the need for a visit to the customer's home. The most common method of security upgrade is achieved by inserting a smart card. The smart card contains a secure microprocessor and is mailed to the customer. When the smart card is inserted into a slot in the set-top, it replaces the embedded secure microprocessor. When all customers have received and installed the smart card, the conditional access system is upgraded so that only set-tops with the smart card continue to work.

Security Replacement

Replacing the secure microprocessor by means of a smart card is effective in most security breaches, but if the means of payload encryption is compromised, smart card replacement will not close the breach.

Security replacement also supports portability—the ability to move the set-top from one system to another. This is the basis for retail set-top portability and is discussed in more detail in Chapter 18.

Analog De-scrambling

In a digital set-top, analog de-scrambling is an option. Two approaches are currently being considered by the cable industry:

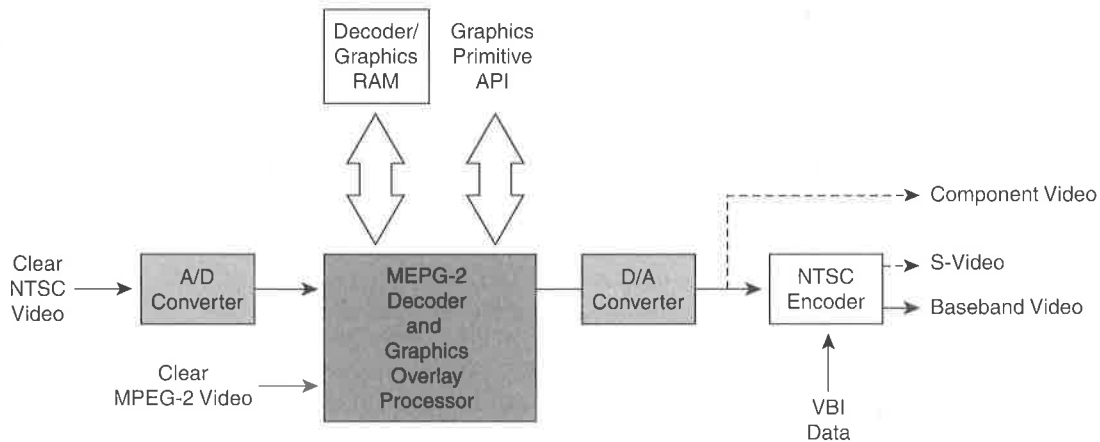
- Include an analog de-scrambler in the digital set-top. The disadvantage with this approach is that analog scrambling is not that effective (see Chapter 3) and that each analog scrambling system is highly proprietary, effectively preventing new vendors from building an analog de-scrambling portion of the hybrid analog/digital set-top.

- Remove the need for an analog de-scrambler by simulcasting all analog-scrambled channels in digital form. This approach requires additional investment at the headend and consumes more cable spectrum. However, it reduces set-top cost and provides a migration path away from analog scrambling. (There also is some early indication from the FCC that this obviates the requirement for a separate analog de-scrambling module—see Chapter 18.)

Video and Graphics Processing

The video and graphics processing system is shown in Figure 6-4. This system processes both analog and digital video inputs, which are described in the following sections.

Figure 6-4 *Video and Graphics Processing System*



MPEG-2 Video Decoding

The video decoder is compliant to MPEG-2 main profile at main level (MP@ML), which supports all standard definition digital television programs. It is common to reduce the horizontal resolution to three-quarters or one-half of the full resolution supported by MP@ML to increase the compression ratio (see the section Picture Resolution in Chapter 4).

As high definition television signals become more common, and as main profile at high level (MP@HL) decoders become more affordable, digital set-tops will include the option of a high-definition decoder. However, the only output that currently supports an uncompressed high-definition video is the component video output. This is specified by EIA 770.3 (see the section Component Video in Chapter 17, "OCI-C2: the Consumer Interface").

Analog Video Processing

Two alternative approaches to analog video processing are commonly used in digital set-tops:

- Pass the analog video through to the set-top output in exactly the same way as an analog set-top (see Chapter 3).
- Convert the analog video into digital video and process it in the same way as the output of the MPEG-2 decoder.

Each approach has its advantages and disadvantages and will be described in the following sections.

Analog Video Pass-Through

The analog video pass-through approach uses the same video compositing approach as the advanced analog set-top (described in Chapter 3). This approach has significant disadvantages in a digital set-top:

- The on-screen-display does not support *alpha-blending*, making the on-screen display more intrusive because it completely obscures the video. (Alpha-blending is a technique that makes the graphics overlay transparent so that the video is visible through the overlay).
- Transformations of the analog video are not supported—for example video scaling and frame capture. These features are commonly used in advanced guides to maintain context in the program guide, using, for example, quarter-screen video.
- If the set-top offers S-Video or component video outputs, only digital channels will be present on these outputs because analog channels are not converted to these formats. This forces the customer to manually switch the television receiver input back to the baseband composite video or the RF input to watch analog channels.

These limitations cause the user interface designer to make a difficult choice: either to limit the digital user interface to the same level as the analog user interface, or to design the digital and analog user interface with a different look and feel.

Analog Video Digitization

Figure 6-4 illustrates the approach typical in second-generation digital set-top designs. The analog video is digitized, and each field is captured in the graphics memory. This approach allows digital transformations of the video (such as video scaling) with an appropriate graphics engine. The advantages of analog video digitization are that the user interface can be *seamless* across analog and digital programs. Moreover, the analog output is present on S-Video and component video outputs.

However, there are some disadvantages to this approach:

- It is more costly (although use of VLSI can mitigate this).
- If the analog video is of poor quality when captured, frame synchronization can be a problem.
- The process of A/D conversion and subsequent D/A conversion will inevitably cause some degradation in analog video quality due to sampling errors.

Nevertheless, these issues can be properly managed, and all advanced digital set-top designs perform analog video digitization.

On-Screen Display

The *on-screen display (OSD)* generates a graphical overlay on the video output. In digital set-tops, the on-screen display is generally more sophisticated than its analog forbears; it can support scalable fonts, graphics acceleration, and more colors. In addition *alpha-blending* is usually supported. (Alpha-blending is a technique that makes the graphics overlay transparent so that the video is visible through the overlay.)

The on-screen display is usually implemented into the same chip as the MPEG-2 decoder. The on-screen display usually includes a graphics processor, or accelerator, that offloads some of the graphics processing from the main processor. Graphics operations are described by means of a graphics application programming interface (API) that describes graphical operations as a series of *graphics primitives*. For example, a graphics primitive might instruct the graphics processor to draw a circle or a line or to display some text with a certain font, color, and transparency.

The resulting graphical image is combined with the video signal using alpha-blending. This allows the pixel value in each video line to be composed from the video input signal and the graphics image in memory, based on its x-y coordinate.

The combined signal is digitally filtered to prevent illegal transitions of the video signal. The best example of filtering is called *anti-alias filtering*, where adjacent field lines are filtered to prevent flickering.

The graphics processor can also support digital transformations of the video. This requires special hardware support or an extremely powerful processor because the chosen transformation must be applied to each field of video in less than one-sixtieth of a second. Common video transformations include

- Video scaling—The video is enlarged or scaled down in size. This can be used for picture-in-graphics or picture-in-picture composition.

- Field/frame capture—The video is captured in freeze-frame. This might be used in a multiprogram guide where multiple panels are used to show images captured from a series of available programs.
- Video warping—The video is spacially distorted according to a mathematical formula. For example, the video might be wrapped around the surface of a cylinder.

As in the advanced analog set-top, the main purpose of the on-screen display is to support a menu-driven user interface and an electronic program guide. However, the addition of video transformations and graphics acceleration in the advanced digital set-top mean that the look and feel of these applications can be much higher in quality. In addition, all these features can be useful for interactive applications, such as interactive advertising (see Chapter 8, "Interactive Services").

Digital-to-Analog Conversion

The output of the standard definition MPEG decoder is in CCIR-601 digital format. A digital-to-analog (D/A) converter transforms the digital video into a component analog signal, which has three components, as specified by EIA 770.1:

- Luminance (Y)—This component is the sum of all colors (R,G,B) and represents the brightness of the picture according to this equation:

$$Y = 0.299R + 0.587G + 0.114B$$

- Color difference (Pb)—This component is a color difference signal formed by subtracting the luminance signal from the blue signal according to this equation:

$$Pb = B - Y = -0.299R - 0.587G + 0.886B$$

- Color difference (Pr)—This component is a color difference signal formed by subtracting the luminance signal from the red signal.

$$Pr = R - Y = 0.701R - 0.587G - 0.114B$$

Some high-end television receivers accept a component video signal according to EIA 770.1, and this is an optional output from the set-top. However, most television receivers accept only an NTSC input.

NTSC Encoding

All television receivers accept an NTSC input, and the component video signal must be encoded according to the NTSC standard.

The NTSC encoder also accepts a data input to modulate onto the VBI lines. The most important VBI line is line 21 field 1, which is used to carry closed-captioning. It is an FCC regulation that closed-captioning must be present at the output of the set-top box if

closed-captioning was transmitted with the program. There are two ways in which closed-captioning can be received by the set-top:

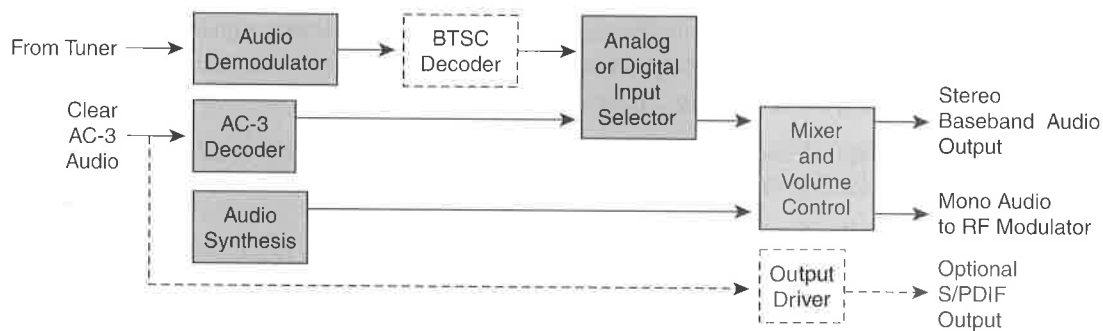
- An analog video signal can contain closed-captioning in line 21, field 1. Only the active video lines are digitized, so line 21 is lost during this process. The solution is that the set-top decodes line 21 and reinserts it at the NTSC encoder.
- A digital video signal can contain closed-captioning in the video user data (VUD), which is embedded in the video program elementary stream (PES). The set-top decodes the VUD to recover closed-captioning and inserts it in line 21 using the NTSC encoder.

As shown in Figure 6-4, the NTSC encoder generates two outputs: an S-Video and a composite output. The S-Video output uses two wires to carry luminance and chrominance signals separately. The composite (or baseband) signal is formed by modulating the chrominance information onto the luminance signal.

Audio Processing

Figure 6-5 shows the audio processing path for the digital set-top. This subsystem has to deal with both analog and digital audio inputs.

Figure 6-5 *Audio Processing Subsystem*



There are three sources of audio:

- Analog audio—The analog audio signal can be encoded in stereo using the BTSC standard and can include a secondary audio program.
- Digital audio—The digital audio signal can be encoded in 2 or 5.1 channel Dolby AC-3.
- Digitally synthesized audio—This audio signal provides the application programmer with the ability to generate sounds locally on the set-top.

The audio processing subsystem is responsible for decoding the audio inputs and selecting or mixing them appropriately.

Dolby AC-3 Audio Decoding

The Dolby AC-3 (also known as *Dolby Digital*) decoder provides multichannel digital audio from mono through 5.1 channels. Dolby AC-3 has been adopted by North American cable operators as the audio format of choice. (An alternative audio standard, MPEG-1 audio, is losing favor but is still supported by many set-tops; see the section Audio Compression in Chapter 4.)

Most digital set-tops provide only a 2-channel baseband audio output (stereo left and right channels), and the Dolby AC-3 decoder is designed to convert a 5.1 channel input signal into a 2-channel Dolby Pro-Logic encoded output.

BTSC Stereo Decoding

BTSC decoding is optional in some digital set-tops, but there are a number of reasons why BTSC decoding is a candidate for serious consideration:

- Seamless audio output—It is very expensive to BTSC *encode* the multichannel audio contained in digital channels, and BTSC encoding would cause a significant quality loss. Therefore, the customer is generally encouraged to use a baseband audio connection to the television receiver (or a separate audio amplifier). In this environment, a BTSC decoder is required to provide *seamless* stereo support for analog channels.
- High-end audio support—Including a BTSC decoder provides the customer with a convenient source of stereo baseband audio to supply home theater equipment. (Currently, the hi-fi VCR provides this function for many customers, but only for unscrambled analog channels.)

Audio Synthesis

The digital set-top usually includes some mechanisms for generating sounds. These sounds must be mixed with the audio signal at an appropriate volume level. Some of the common audio synthesis techniques include wave-files (a common PC format) and MIDI synthesis, which can be used to generate simple musical effects.

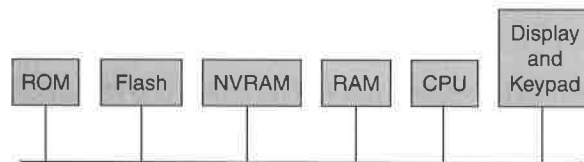
Audio synthesis is inexpensive to include in a large ASIC and is usually included for game support. Focus studies show that many customers disable clicks or beeps associated with the user interface if they are given the option of doing so.

Microprocessor Subsystem

The microprocessor subsystem in a digital set-top is similar to the one in an advanced analog set-top (see Chapter 3). (See Figure 6-6.) However, in a digital set-top, the microprocessor usually has considerably higher performance to support the additional processing requirements of a digital set-top:

- Each of the new digital components (QAM demodulator, MPEG-2 decoder, AC-3 decoder, and so on) requires a device driver that executes on the microprocessor.
- Higher CPU performance is needed to take advantage of additional graphics features.
- More applications are envisioned to justify the additional cost of a digital set-top.

Figure 6-6 *Microprocessor Subsystem*



The microprocessor subsystem supports the same features as an advanced analog set-top:

- Menu-driven interface—The set-top contains software that supports a very flexible user interface that, in combination with the on-screen display, allows greater ease-of-use and more complex features to be offered to the customer.
- Electronic program guide (EPG)—The EPG allows the customer to scan the program offerings using the on-screen display, usually by means of a table of channels (the rows of the table) and times (the columns of the table).
- Software download—In most digital set-tops, the software can be downloaded over the network so that additional features (or bug-fixes) can be distributed to set-tops that have been deployed.

In advanced digital set-tops, *application sharing* must be supported to either download applications on request or to maintain multiple applications in set-top memory. Operating system support for application sharing becomes very important (see Chapter 8).

The microprocessor subsystem has a number of components:

- The central processing unit
- The memory subsystem
- The display and keypad

Central Processing Unit

The CPU is typically a 32-bit microprocessor. By trailing the microprocessor price-performance curve of the processors commonly used in personal computers, the digital set-top limits the cost of the CPU to just a few dollars. Another approach is to license a microprocessor core and to include it on the same chip used for other set-top functions. For example, a microprocessor core is commonly included in the transport ASIC.

The CPU performance of a basic, first-generation digital set-top is similar to an advanced analog set-top at 3 to 10 million instructions per second (mips). This is adequate for a limited set of embedded applications as long as they are coded efficiently (for example, in C or C++).

The CPU performance of advanced digital set-tops must be considerably higher than that of a basic digital set-top. A minimum performance of about 50 mips is required, although processors of 100 to 200 mips will soon be available at reasonable cost. Improved CPU performance is required by more sophisticated applications and by interpreted environments (such as JAVA) to enable application portability.

Memory Subsystem

The memory subsystem contains a number of different types of memory, each having a specific purpose. In more recent designs that use a microcontroller, part of the memory might be contained on the same chip as the microprocessor.

The different types of memory are

- Read-only memory (ROM)
- Flash memory
- Nonvolatile random access memory (NVRAM)
- Random access memory (RAM)

Read-Only Memory

Read-Only Memory (ROM) contains the firmware for the set-top. ROM is the least expensive type of memory, and it is used for storing program instructions and data that are never changed. Because "never" is a long time, programs are usually not committed to ROM until they are thoroughly debugged, which requires exhaustively thorough testing of the program. This testing takes time, so early versions of the set-top might use programmable read-only memory (PROM) to speed the development cycle. When the code is stable, it will typically be stored as masked-ROM because it is the cheapest type of ROM. (In the masked-ROM process, the program is lithographically etched into the chip as the final metal connection layer.)

As the software complexity of a digital set-top design increases, ROM is used less and less because it limits flexibility. ROM is now relegated to basic set-top bootstrap functions that are very unlikely to change.

Flash Memory

Flash memory is also referred to as electrically erasable programmable read-only memory (EEPROM). Flash memory can be erased and rewritten thousands of times, so it is ideal for storing program code and data that might need to be changed due to bug-fixes or other enhancements. Flash memory is typically two to three times more expensive than ROM.

Most digital set-tops support field upgrade of their functions, and this is achieved by downloading new program code and data over the network. Such field upgrades can be very useful to the cable operator but are also fraught with peril! If the download fails, the set-top might be left in an undefined state or might fall back to a basic set of functions. In addition, sometimes customer preferences, which are stored in the set-top, are lost.

One disadvantage of certain types of flash memory is that it cannot support the access speed required by the microprocessor. In this case, the contents of the flash can be copied to RAM before execution, or the microprocessor can be programmed to insert wait cycles. The first solution requires additional RAM, whereas the second solution can seriously affect performance.

Nonvolatile Random-Access Memory

Nonvolatile Random-Access Memory (NVRAM) is used to store any program variables or user data that must survive when the set-top loses AC power. There are two common techniques for implementing NVRAM:

- The use of NVRAM chips—These chips are very reliable, but they are expensive and available only in limited sizes (2 KB is used in some digital set-top designs).
- The use of battery-backed RAM—This is standard RAM that is powered by a small on-chip battery. The RAM size can be many times larger than NVRAM chips, but if the battery fails it must be replaced.

Random Access Memory

Random Access Memory (RAM) is read/write memory that retains its contents only while it is powered. RAM is used to store data and program code. You are probably familiar with the dramatic reduction in the price of RAM for personal computers. The same trends make digital set-tops with 4 to 8 MB of RAM economical.

The main uses for RAM in a digital set-top are

- MPEG-2 decode—Main profile at main level (MP@ML) MPEG decoding requires 2 MB of RAM. A high-definition, main profile at high level (MP@HL) decoder requires approximately 10 MB of RAM; however, some decoders use memory compression techniques to reduce this to 8 MB or less.)
- Graphics processing—A basic rule of thumb is that the more sophisticated the graphics processing, the more memory it requires. Graphics systems are often run in lower resolution and reduced color depth (less bits are allocated to describe the color of each pixel) to reduce memory requirements.
- Operating system—The operating system requires a relatively small amount of memory for its internal data structures and I/O buffers.
- Applications—Each application requires a certain amount of memory, which is dependent on the type of application and its design. If multiple applications are active simultaneously, the digital set-top must have sufficient RAM for their total memory requirements. (This is different from a PC, where applications can be swapped out of memory and onto disk if necessary.)

Display and Keypad

In a digital set-top, customers use the on-screen display and the remote control for most of their interaction with the set-top. However, remote controls have a way of slipping down the back of the couch or exhausting their batteries, so a keypad is usually included on the set-top itself for this eventuality.

For convenience, the on-screen display is supplemented by a 4-digit display that can be used to display the channel number or the time.

RF Modulator

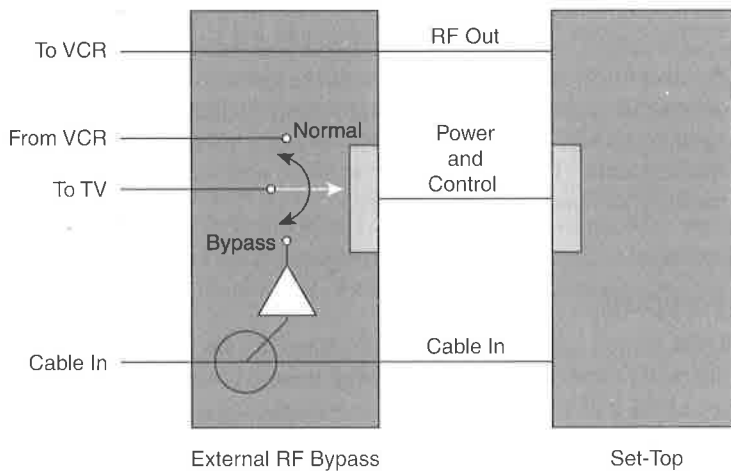
The RF modulator takes the composite NTSC video and audio carriers and places them on channel 3 or channel 4 (although another channel might be available as an option) for output to the television receiver. Channel 3 or 4 selection might be by a switch on the set-top or by software control. Channel 3 or 4 is selected according to which is the *weakest* broadcast signal—that is, which will cause least interference at the television receiver input.

RF Bypass Switch

The RF bypass switch is actually a coaxial relay that returns to the bypass setting when the set-top is in standby. In this position, the RF bypass switch allows all cable channels to pass through the set-top unchanged to the television receiver. (This behavior was mandated by the 1992 Cable Act and later included in the Code of Federal Regulations by the FCC.)

An RF bypass option (see Figure 6-7) allows the customer to connect a VCR to the digital set-top in such a way as to provide recording of all channels.

Figure 6-7 RF Bypass Option



The bypass switch has two positions:

- Normal—The television receives the RF output of the set-top (via the optional VCR loop-through). The customer can view television programs and is able to record and play back using the VCR.
- Bypass—The television receives the entire cable spectrum from the RF input. The customer can tune unscrambled analog television programs with the television receiver.

Inputs

The digital set-top has the following inputs:

- Infrared receiver—An infrared receiver is mounted on the front of the set-top box to allow line-of-sight operation with an infrared remote control. All digital set-tops are supplied with a *universal* remote control, which can also control the television and VCR.

- Diagnostic port—It is quite common for the set-top to have a serial data port for diagnostic purposes. This port is typically used during development of the set-top firmware and can be completely disabled in a production unit for security reasons.
- Data port—A data port on the set-top can be used for various expansion purposes. The most common of these is to connect two digital set-tops to make them function as a dual-tuner set-top for watch-and-record applications (see the section Data Port in Chapter 3).
- Game port—The optional game port provides a way to connect a joystick or game controller to the set-top.

Outputs

The digital set-top has the following outputs:

- RF channel 3/4 output
- Baseband video
- Baseband audio
- Infrared transmitter
- Component video
- IEEE-1394

The RF output is always present and uses the standard 75 ohm F-connector. Baseband audio and video outputs are almost always provided because they provide a significant increase in quality over the RF channel 3/4 output.

All other outputs are optional.

RF Output

An RF output provides the simplest and the most common method of connection to a television receiver or VCR. The RF output carries a clear NTSC signal modulated onto channel 3 or channel 4.

Baseband Video

A baseband video output is provided for connection to the *video-in* input of a television receiver or VCR. This signal is sometimes called a *composite signal* because it contains the baseband luminance information and the modulated chrominance information according to the NTSC standard. The RCA-style connector is the *de facto* standard for this signal.

Baseband Audio

A baseband audio output is provided for connection to the *audio-in* input of a television, VCR, or home-theater. This is a stereo output when an optional BTSC decoder is included in the set-top.

S-Video

An optional S-Video (or Super-Video) output provides superior video bandwidth when compared with baseband NTSC. The S-Video output uses two wires to carry luminance and chrominance signals separately. The physical connection is provided by a 4-pin DIN connector.

Component Video

An optional component video output provides the highest-quality analog signal. Component video is now becoming more common on high-end television receivers, and the Electronic Industries Association (EIA) has published a standard, EIA 770.1, for component video interconnection. The physical connection is made using 3 RCA connectors. (Refer to the section Digital-to-Analog Conversion earlier in this chapter.)

However, there is no agreed-upon standard for copy protection of a component video output (although MacroVision Corporation has proposed a proprietary mechanism for copy protection).

IEEE 1394

IEEE-1394 or FireWire is a recent addition to PCs and will soon be available as an input to high-end television receivers. IEEE-1394 is a high-performance serial interface that supports isochronous and asynchronous transfer modes. It is intended as the de facto standard for digital interconnection of multimedia devices in the home (refer to Figure 6-1).

IEEE-1394 was originally introduced without any means for copy protection, and this has effectively limited its usefulness for anything but connecting video camcorders to display devices. The movie studios are very concerned about copy protection for a couple of reasons:

- The signal is transferred over the IEEE-1394 interface as an MPEG-2 transport stream carrying video and audio in compressed digital form. This makes it simple to store, to replicate, and to transmit the information, essentially providing anyone with his or her own pirate video store.

- The IEEE-1394 is becoming widely available on PCs. The PC is a flexible tool that can be made to perform almost any operation with the appropriate (or inappropriate) software. Thus, when a movie is available in digital form at a PC interface, it is simple to store it, manipulate it, and transmit it to other PC users (via the Internet).

To address these concerns, a Copy Protection Technical Working Group was formed to find copy protection mechanisms that could be adopted by the consumer electronics and computer industries. (See Chapter 17.)

Digital Audio

Most digital set-top designs include only a stereo, baseband output, which does not support the full capabilities of Dolby AC-3. S/PDIF is a *de facto* digital serial interface that was developed for digital transmission of the compressed AC-3 audio. Many DVD players support this output. S/PDIF (pronounced *spidiff*) stands for Sony/Philips digital interface.

S/PDIF is used to connect the digital audio signal to an external Dolby AC-3 decoder, which is usually packaged with a high-quality, 6-channel power amplifier. This arrangement provides the audiophile with the full capabilities of the Dolby AC-3 system.

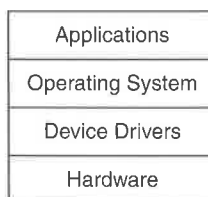
Infrared Transmitter

An optional infrared transmitter is used to control a VCR. (This is sometimes called an *IR blaster*.) If correctly configured, this can simplify the watch-and-record function considerably (see Chapter 3).

Software

The software architecture of a basic digital set-top is designed to support a small number of embedded applications. In this environment, the software architecture shown in Figure 6-8 is quite adequate.

Figure 6-8 *Software Architecture*



The software can be divided into three main categories:

- Device drivers
- Operating system
- Applications

Device Drivers

Device drivers are programs that work directly with the hardware devices and are often written in assembly language for optimum performance. The device drivers provide a hardware abstraction layer to the operating system.

Operating System

The set-top operating system is quite simple because only a few functions are required:

- Task scheduling—This provides the basic mechanism for ensuring that applications are run in sequence so that each has a turn to use the processor.
- Timers—Timers are used to trigger task execution at some future time and are used by unattended recording and sleep timer applications.
- Interrupt handling—The operating system dispatches interrupts to the appropriate device driver.

A number of commercial operating systems have been developed that satisfy these requirements, including pSOS, VRTX, and VxWorks. It is also quite simple to develop an operating system with these functions in-house.

The operating system is simple because all other functions—for example, memory management—are done by the application.

Applications

The set-top applications perform all the desired functions of the set-top and are typically coded in C or assembly language. The applications developer is responsible for ensuring that the application does not corrupt its own or other applications data, because there is no memory management in the operating system or hardware. Therefore, the applications are *trusted* and must be tested thoroughly; if an application fails, it will cause the set-top to crash or hang up.

Many digital set-top applications are not portable; that is, they cannot be readily moved to a different set-top platform because they rely too much on the specifics of the underlying operating system and hardware.

Limitations

A basic digital set-top has some of the same limitations as an advanced analog set-top: Any new applications must be ported to the proprietary operating system and hardware platform. This requires an agreement between the applications developer and the set-top provider. As the number of applications grow, and as third-party application developers develop more applications, the lack of application portability becomes a significant issue.

Summary

This chapter covered the hardware and software characteristics of a basic digital set-top and some of the extended features of the advanced digital set-top.

The basic digital set-top solves few of the problems associated with analog set-tops. The basic digital set-top has fewer proprietary interfaces, but still supports only a limited number of embedded applications. It cannot always interoperate well with other video equipment, such as television receivers, VCRs, and home-theater equipment, in the customer's home.

The advanced digital set-top can include hardware support for IEEE 1394, high definition television and digital audio (S/PDIF). Its additional CPU performance, graphics functions, and memory footprint allow it to function as an applications platform for interactive and on-demand services (see Part II).

References

Books

Ciciora, Walter, James Farmer, and David Large. *Modern Cable Television Technology; Video, Voice, and Data Communications*. San Francisco, CA: Morgan Kaufmann Publishers Inc., 1999.

Standards

EIA-770.1, "Analog 525 Line Component Video Interface—3 Channels." CEMA R.4 Video Systems Committee.

EIA-770.2, "Standard Definition TV Analog Component Video Interface." CEMA R.4 Video Systems Committee.

EIA-770.3, "High Definition TV Analog Component Video Interface." CEMA R.4 Video Systems Committee.

SCTE DVS-031, "Digital Transmission Standard for Cable Television." Society of Cable Telecommunication Engineers Digital Video Subcommittee.

SCTE DVS-093, "DVS-241 Draft Digital Video Service Multiplex and Transport System Standard for Cable Television." Society of Cable Telecommunication Engineers Digital Video Subcommittee.

SCTE DVS-234, "Service Information for Out-of-Band Digital Cable Television." Society of Cable Telecommunication Engineers Digital Video Subcommittee.

Internet Resource

Apple's FireWire Web site

<http://www.developer.apple.com/hardware/FireWire>

Digital Broadcast Case Studies

This chapter considers the implementation of digital broadcast cable systems. The two predominant systems in North America are

- **DigiCable**—General Instrument developed the DigiCable system to meet TCI's requirements. DigiCable has been deployed by multiple cable operators in North America, including TCI (now AT&T BIS), Comcast, and Shaw.
- **Pegasus**—Scientific Atlanta developed its Digital Broadband Delivery System (DBDS) in response to Time Warner Cable's Pegasus system requirements. Multiple cable operators in North America, including Time Warner Cable, Comcast, Cox, Marcus, and Rogers Cable, have adopted it.

This chapter is organized into three main sections:

- **Architectural comparison**—This section compares and contrasts the two system architectures and describes the different cable operator requirements for digital broadcast.
- **DigiCable**—This section describes the DigiCable system and its components in detail.
- **Pegasus**—This section describes the Pegasus system and its components in detail.

Architectural Comparison

DigiCable and Pegasus systems are very different in their implementation but share many common architectural features:

- **Satellite distribution to headends**—Digital programming is compressed at an uplink and delivered in compressed digital form to the headend location via satellite.
- **Headend to subscriber distribution**—QAM modulation is used for broadband transmission of the digital programming over an analog cable network.

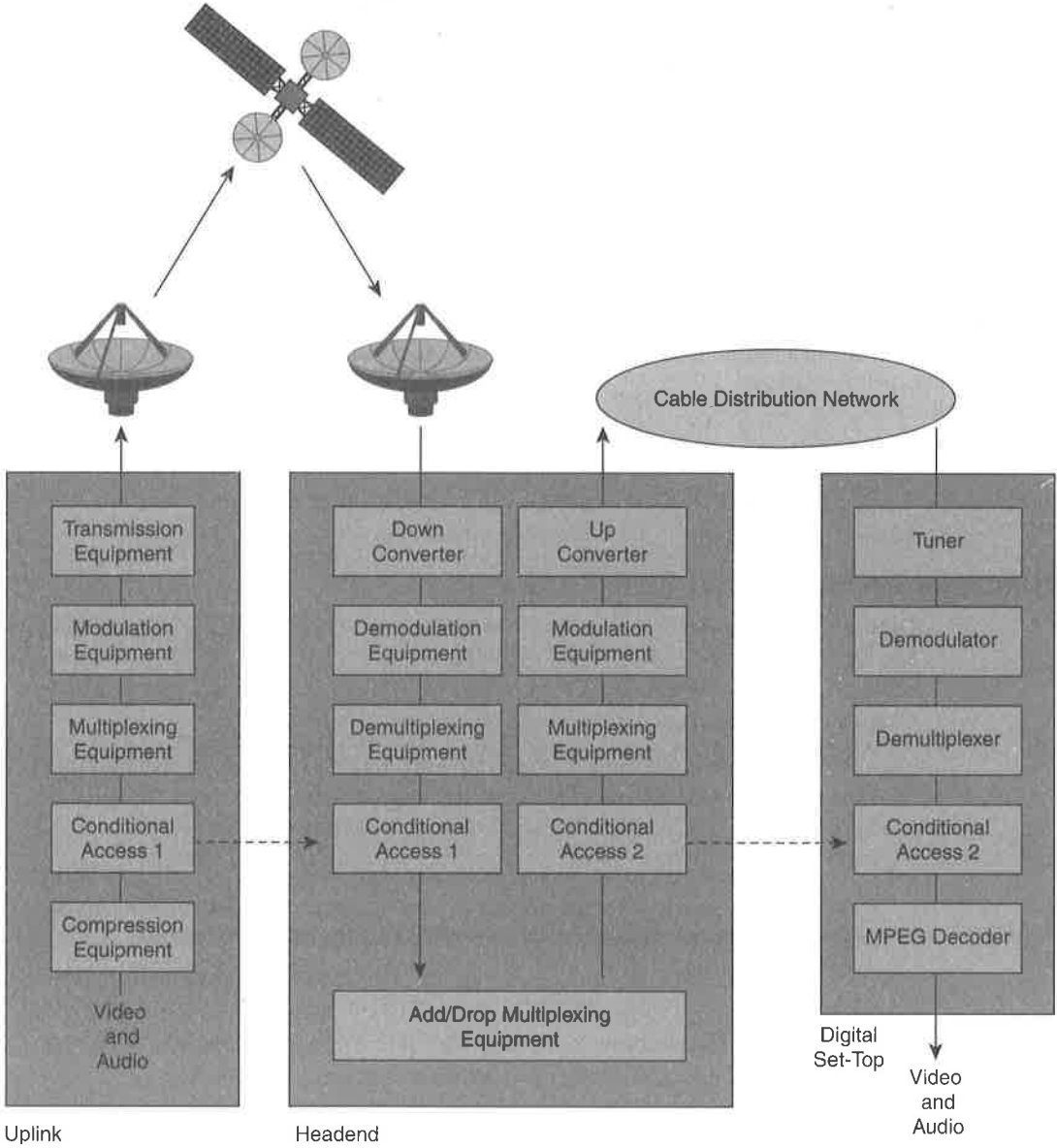
- A *split* security model—The conditional access system used to secure the satellite link operates independently from the conditional access system used to secure the cable system.
- An out-of-band data channel—A separate QPSK carrier is used to deliver common system information associated with all in-band channels.

Satellite Distribution to Headends

In both systems, a satellite link distributes MPEG-2 compressed digital channels to cable headends, as shown in Figure 7-1. The up-link equipment generates an MPEG-2 multi-program transport stream (MPTS) for each satellite transponder. The processing for each transponder is as follows:

- A real-time encoder compresses the video and audio content, generating an MPEG-2 single program transport stream (SPTS) for each program. (The content may be received from an analog or digital satellite feed, or played back from a video tape or digital file server.)
- Each SPTS is secured using a conditional access system (CA₁). General Instrument's Digicipher II is by far the most common system in use in North America, and Scientific Atlanta's PowerVu is also used.
- Multiplexing equipment combines a number of SPTSs to generate an MPEG-2 multi-program transport stream (MPTS). The MPTS bit rate is chosen to fill the entire payload of a satellite transponder, which varies from between about 27 and 44 Mbps, depending on transponder bandwidth and forward error correction coding.
- The modulation equipment applies forward error correction to the MPTS and is responsible for QPSK modulation. The output is at L-band.
- The transmission equipment up-converts each L-band carrier to the satellite transmission band (C-band or Ku-band). Typically, 24 carriers are amplified, combined, and fed to the dish via a wave-guide.

Figure 7-1 *Satellite Distribution to Headends*



At each headend, the processing of the digital payload is fundamentally the same in DigiCable and Pegasus systems:

- The satellite signal is received and down-converted to L-band by a low-noise block (LNB) converter. (See Chapter 6 of *Modern Cable Television Technology; Video, Voice, and Data Communications* by Walter Ciciora and others.) The satellite modulation is QPSK (see the section Broadband Transmission section in Chapter 4, "Digital Technologies").
- Each L-band carrier is tuned and demodulated to recover the MPEG-2 MPTS by the demodulation equipment.
- The de-multiplexing equipment separates the MPTS into its component SPTS.
- The conditional access system (CA₁) decrypts each SPTS. Each SPTS is secured using a different key so that each SPTS can be authorized separately.
- The SPTS may be *groomed* by an Add/Drop Multiplexer at the headend to build a new system multiplex (or MPTS). In many systems, the MPTSs constructed at the up-link are designed to be passed through the headend intact, so grooming is unnecessary, thereby reducing headend equipment costs.

Headend-to-Subscriber Distribution

DigiCable and Pegasus systems use the same mechanisms to distribute broadcast digital services to subscribers, using QAM modulation to carry MPEG-2 MPTS over the analog cable distribution network to a digital set-top (see the section Digital Transmission in Chapter 5, "Adding Digital Television Services to Cable Systems").

The North American cable industry agreed to use the ITU J.83 Annex B standard for QAM modulation developed by General Instrument. The headend-to-set-top processing is shown in Figure 7-1:

- Each SPTS is secured using a conditional access system (CA₂). The DigiCable system uses General Instrument's Digicipher II conditional access system. The Pegasus system uses Scientific Atlanta's PowerKEY conditional access system.
- Multiplexing equipment combines a number of SPTSs to generate an MPEG-2 multi-program transport stream (MPTS). The MPTS bit rate is chosen to fill the entire payload of a QAM channel, which is 26.97035 Mbps for a 64-QAM channel or 38.81070 Mbps for a 256-QAM channel.
- The modulation equipment converts the MPEG-2 MPTS into a 6 MHz 64-QAM or 256-QAM channel.
- The up-converter places the QAM channel on the desired frequency.

The QAM channels are combined with existing analog NTSC channels and sent over the cable distribution network. In the customer's home, a hybrid digital/analog set-top is responsible for the final processing steps in the signal path (see Chapter 6, "The Digital Set-Top Converter," for more details):

- The chosen cable channel is selected by a tuner, which feeds an intermediate frequency to a QAM demodulator.
- The demodulator recovers the MPEG-2 MPTS.
- The de-multiplexer selects a single SPTS from the MPTS.
- The conditional access system (CA₂) decrypts the SPTS.
- The audio and video elementary streams in the SPTS are decoded and fed to the television receiver.

The signal path from the up-link to the customer has a great many processing elements, but the original SPTS encoded at the up-link is unchanged when it arrives at the set-top decoder. The great advantage of a digital system is that so many processing steps are possible while maintaining the integrity of every bit in the SPTS. Thus, the picture quality from a decoder placed at the up-link is identical to the picture quality from a set-top in the customer's home (assuming that the transmission system is properly maintained).

Split Security Model

The split security model separates the conditional access system for the satellite link from the conditional access for the cable system:

The digital channels are secured over the satellite delivery system by CA₁.

The digital channels are secured over the cable system by CA₂. The DigiCable system uses Digicipher II conditional access for the satellite link (CA₁) *and* for the cable system (CA₂). The Pegasus system uses Digicipher II *or* PowerVu conditional access for the satellite link (CA₁) and PowerKEY conditional access for the cable system (CA₂).

The split security model maintains complete independence between CA₁ and CA₂. This approach has significant operational benefits for the content provider and the cable operator:

- The content provider need only authorize equipment at the cable headend. This enables the content provider to manage *its* customer, the cable operator.
- Cable operators can choose their cable conditional access system independently of the satellite conditional access system.

- The cable operator can insert locally encoded channels into the channel lineup and secure them with the cable conditional access system. (This is particularly important for video-on-demand services. See the section Conditional Access in Chapter 10, "On-Demand Services").
- A security breach of either the CA₁ or CA₂ conditional access systems does not affect the integrity of the other conditional access system.

Out-of-Band Channel

Both the DigiCable and Pegasus systems rely on an out-of-band data channel. The out-of-band channel delivers common system information associated with all in-band channels to the digital set-tops:

- Entitlement management messages (EMM)—These messages are addressed to individual set-tops and carry secure authorization instructions from the conditional access system in the headend to the set-top.
- Service information (SI)—These messages provide the set-top with information to support channel navigation.
- Program guide information—These messages provide the program guide information to the EPG application in the set-top.
- Emergency alert system (EAS) messages—In response to an EAS message, the set-top displays a text message, plays an audio message, or force-tunes to an alert channel.

Central Versus Local Subscriber Management

The DigiCable system is designed to provide either central or local subscriber management. The Pegasus system is designed to provide only local subscriber management.

In central subscriber management, the satellite distribution system is used to control each cable headend by using part of a satellite transponder as a data communications link to carry authorization and control messages. In local subscriber management, a local headend controller provides all the management function for the cable system.

In both cases, the satellite distribution also provides a way of obtaining digital programming services from the content provider. However, in the centrally managed system, the programming and the management functions may be provided as a *bundled* package. In a locally managed system, multiple sources of digital programming may be combined, or *groomed*, as desired by the cable system operator.

DigiCable

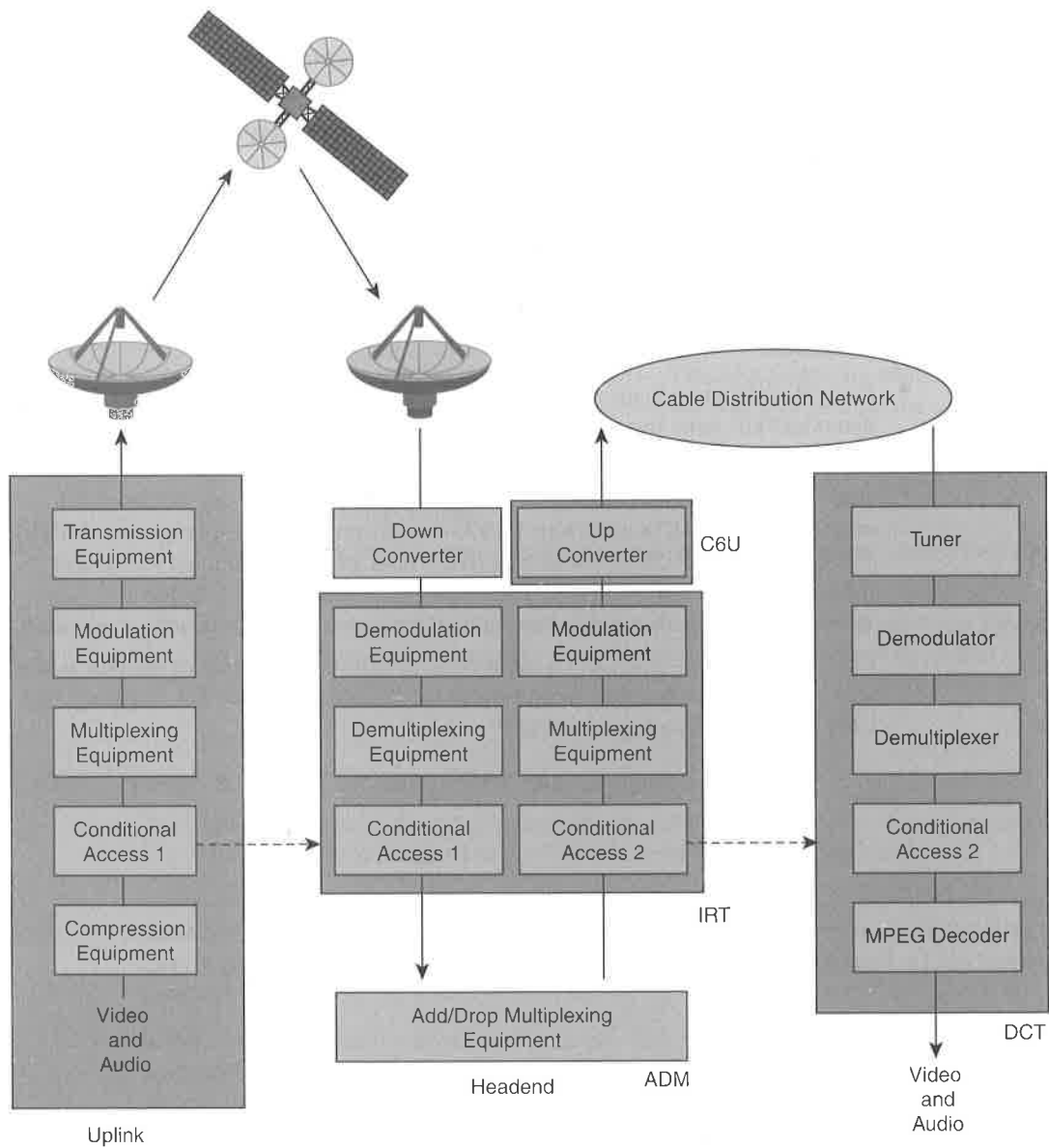
The DigiCable system is supplied by General Instrument. It provides an end-to-end solution for the satellite and cable system distribution networks, including

- Up-link equipment, including MPEG-2 encoding, conditional access, and multiplexing equipment. (This equipment is not described in detail in this chapter.)
- Headend equipment, including satellite receiver, conditional access, and cable transmission equipment.
- Subscriber equipment, a range of digital set-top converters provided by GI, from the DCT-1000 to the DCT-5000 series Digital Communications Terminal (DCT).

Figure 7-2 illustrates how the headend processing functions are packaged in the DigiCable implementation:

- Integrated Receiver Transcoder (IRT)—The IRT efficiently packages the demodulation, conditional access, and modulation functions into a single rack-mounted chassis.
- Add-Drop Multiplexer (ADM)—The ADM can be used to groom the baseband outputs of several IRTs into a new MPEG-2 multi-program transport stream (MPTS). The output of the ADM is fed into an IRT, which applies conditional access and modulation to the MPTS. (Note: The ADM is not widely deployed due to incompatibilities with statistical multiplexing widely used by programmer providers.)
- C6U up-converter—The C6U is required to up-convert the output of the IRT, which is at a 44 MHz intermediate frequency, to the desired QAM channel frequency. The C6U is a dual-channel up-converter.

Figure 7-2 DigiCable Headend Processing Functions



Head-End In The Sky Model

The DigiCable system was designed to meet TCI's (now AT&T BIS) requirement for centralized subscriber management. The Head-End In The Sky (HITS) model uses the satellite distribution system to control each cable headend from the National Authorization Service (NAS) in Littleton, Colorado. The NAS controller supports interfaces to the various NAS customer billing systems. It also supports multiple digital service providers, including HBO, Showtime, and Music Choice.

NOTE

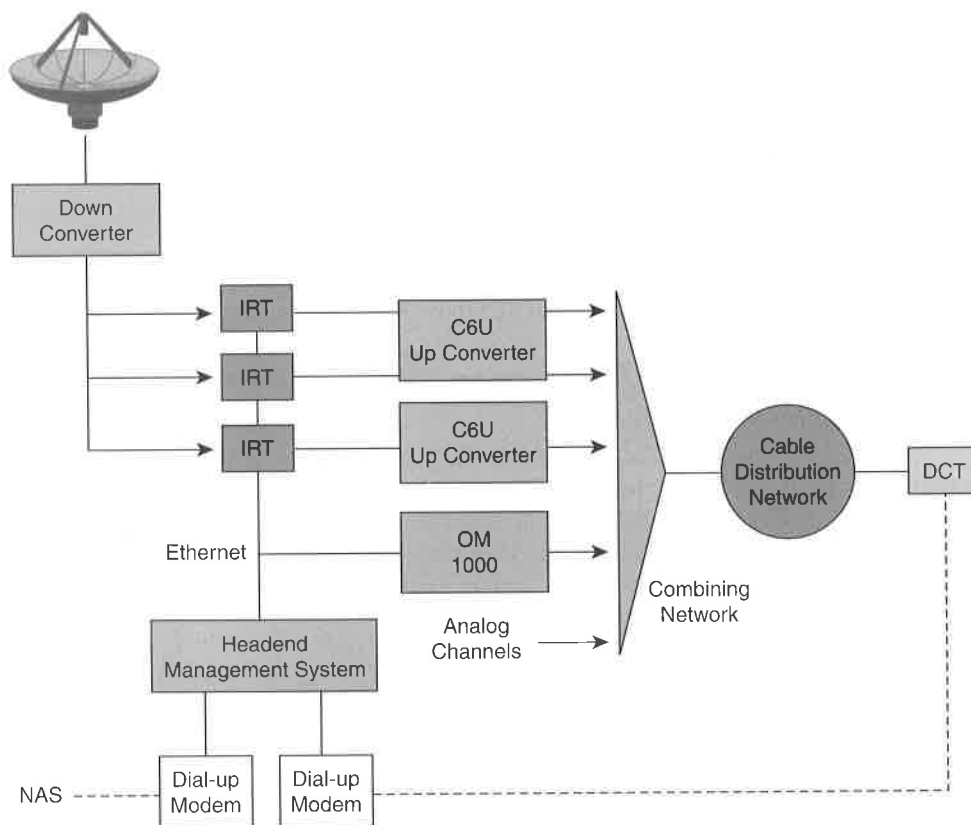
TCI transferred ownership of the National Authorization Service from HITS to General Instrument in June of 1998, though TCI (now AT&T BIS) continues to use the service as before.

Figure 7-3 shows a typical, *three-pack*, HITS configuration. In this configuration, the satellite feed is received by three IRTs, and each IRT is tuned to a different transponder frequency. The IRT provides demodulation, satellite conditional access, cable conditional access, and QAM modulation for an MPTS containing up to 12 programs. The IRT outputs are fed to C6U up-converters and combined in the main cable system feed.

Figure 7-3 shows the other headend components:

- The OM-1000 is an out-of-band modulator that generates a 2.048 Mbps, out-of-band control channel that is received by all digital set-tops.
- The Headend Management System (HMS) is an unattended workstation computer connected by an ethernet to the other components in the headend. Its main function is to collect IPPV purchase records from the digital set-tops (DCTs). It also provides remote management of the headend components.
- The dial-up modems provide a return path from the digital set-tops to the controller, and also provide a mechanism for connection back to the NAS.

Figure 7-3 HITS Architecture



One IRT receives the control channel from the NAS. This 1.5 Mbps data stream is carried as MPEG-2 private data within the MPTS. The IRT recognizes the data stream and outputs it as user data protocol (UDP) packets over an ethernet connection. The NAS control channel carries information to manage both the headend and the subscribers. In each case, these messages arrive over the Ethernet connection from the IRT receiving the NAS control channel:

- The IRTs are managed by control messages addressed to them from the NAS. These messages include satellite conditional access (CA_1) authorization messages to the IRT decryption section as well as cable system conditional access (CA_2) messages to the encryption section of the IRT.
- The OM-1000 is managed by control messages addressed to it from the NAS.

- The Headend Management System (HMS) is managed by control messages addressed to it from the NAS.
- The digital set-tops receive control messages forwarded by the OM-1000 over the out-of-band data channel. These messages include conditional access (CA₂) authorization messages, system information tables, and program guide data messages.

Channel Expansion

The system described in Figure 7-3 shows how the HITS architecture expands the channel lineup of a small, un-upgraded cable system. The three IRTs add about 30 digital channels and 20 music channels to the channel lineup at a cost of only 18 MHz of spectrum. The system is cost-effective for small systems with only a few thousand subscribers because of the small number of highly integrated headend components.

Multiple Channel Lineups

The HITS system provides a single, satellite feed to a large number of cable systems. The program guide data provides information about the analog and digital channels. Each cable system may potentially have a different analog channel lineup, and a mechanism is required to support this environment. The solution is a feature in the service information protocol (SCTE DVS-147, since updated to DVS-234) called multiple Virtual Channel Tables (VCT).

DVS-147 supports up to 4,000 different VCTs, and each VCT is sent in the NAS control channel with a different multicast address. Every digital set-top is provisioned with a multicast address according to its geographical location, and uses this address to filter the proper VCT from all of the VCTs carried in the out-of-band control channel.

The *superset* program guide data is repeatedly broadcast in carousel fashion to all HITS systems. Each set-top recognizes its *own* VCT and uses the pointers contained in it to construct its own program guide from the superset program guide data.

Remote Purchase Record Collection and Forwarding

The National Authorization Service provides a store-and-forward approach to purchase record collection that manages the limited return bandwidth in a telephone return environment. In addition, the NAS now supports purchase collection via the cable upstream path, with upstream signals terminating at an RPD in the headend. The RPD is interfaced to the Headend Management System (HMS).

Each cable headend has an HMS (refer to Figure 7-3). The HMS is programmed to collect and forward purchase records from the digital set-tops:

- The purchase record collection system sends an addressed *poll* message to each set-top in turn over the out-of-band data channel at a scheduled time.
- The digital set-top responds to the poll by retrieving the IPPV purchase records from its secure microprocessor and sends them to the purchase record collection system using a dial-up telephone modem (or via the cable upstream path where the return has been activated).
- Periodically, the HMS dials up the NAS system and uploads the aggregate purchase records to it.
- The HMS sends an addressed message to each set-top in turn to clear the collected purchase records from the secure microprocessor after the purchase records are safely received and acknowledged by the NAS.

Split Security Model

In the HITS and NAS system, Digicipher II provides both the satellite conditional access system (CA₁) and the cable system conditional access system (CA₂), and the Digicipher II controllers are located at the NAS. Nevertheless, CA₁ and CA₂ are managed as completely independent security systems so that if a pirate discovers a session key for CA₂, it would be useless for CA₁.

Local Subscriber Management

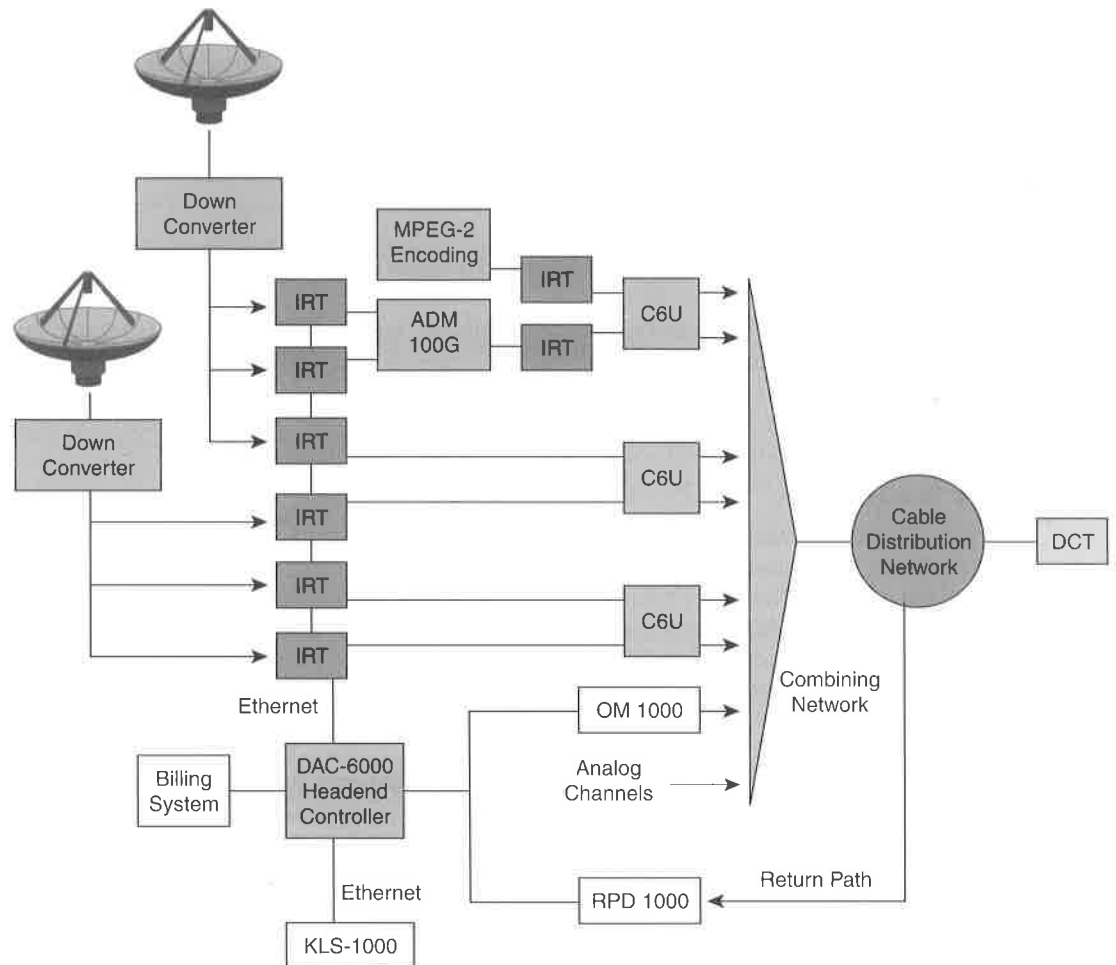
The DigiCable architecture also supports a local subscriber management model. Figure 7-4 shows a diagram of an upgraded cable system that is locally managed.

A locally managed system requires a number of additional components in the headend to provide the functions that were centralized in the HITS model:

- The headend controller, which is called a Digital Access Controller (DAC)-6000 (previously called the ACC-4000D)—This controller is responsible for management and provisioning of all the headend components and digital set-tops. The DAC-6000 also provides the interface to the billing system for service provisioning and IPPV purchase reporting.
- The KLS-1000 key server—This device provides encryption services to the headend controller to support the Digicipher II conditional access system (CA₂).
- MPEG-2 encoding—Local analog channels may be encoded and multiplexed together for transmission over a QAM channel.

- The ADM-1000G add/drop multiplexer—The ADM-1000 is used to groom MPTSs received from multiple sources into a single MPTS for transmission over a QAM channel.

Figure 7-4 *Local Subscriber Management Model*



In Figure 7-4, the upgraded system provides an RF return path, and the digital set-tops may use it for return signaling. The DCT set-top has an optional RF return module called the StarVue RF module. The RPD-1000 receives a 256 Kbps QPSK return signal from the DCT set-top (as specified by DVS-178, see the section SCTE DVS-178 in Chapter 5). The RPD-1000 terminates and aggregates up to six return paths into a single ethernet output.

Finally, the billing system may be located on-premise. Alternatively, if a billing service bureau is used, a wide-area network (WAN) link is used to connect the headend controller to the billing system.

Channel Expansion

The system described in Figure 7-4 shows how a locally managed system uses DigiCable to expand the channel lineup of a larger, upgraded cable system. The number of IRTs is limited only by the available spectrum for digital. The locally controlled DigiCable system is cost-effective for larger systems with more than about 50 thousand subscribers.

Split Security Model

In the locally managed system, Digicipher II is used for both the satellite conditional access system (CA₁) and the cable system conditional access system (CA₁). The CA₁ controllers are located at the content provider up-link locations, and the CA₂ controller is located in the headend. CA₁ and CA₂ are managed as completely independent security systems; if a pirate discovers a session key for CA₂, it would be useless for CA₁.

DigiCable Summary

The DigiCable system uses satellite distribution to the headend and provides for either central or local subscriber management. Its design supports TCI's HITS and the NAS architecture by sending a 1.5 Mbps control channel over a satellite transponder from the NAS. The DigiCable system also supports a local controller architecture.

Pegasus

The Pegasus system is designed to meet TWC's requirements for local subscriber management. Headends are clustered into cable *divisions* that are independently managed and operated. Each division requires a local headend control system and uses the satellite distribution system as a convenient way of obtaining program content.

The Pegasus architecture is defined by Time Warner Cable and incorporates many of the lessons learned from the Time Warner Full Service Network. Pegasus is a phased system with two major phases:

- Phase 1.0 Digital Broadcast
- Phase 2.0 On-Demand Services

This chapter focuses on Pegasus Phase 1.0. On-demand services are discussed in Chapter 10. Pegasus also has minor phases, including Phase 1.1, which introduces interactive services to a broadcast system; these services are discussed in Chapter 9, "Interactive Cable System Case Studies."

Pegasus Phase 1.0 Goals

The Pegasus architecture introduces a number of important technical and business concepts. These are articulated in the Pegasus Phase 1.0 goals:

- Support interactive and broadcast digital services in an integrated manner—Pegasus is conceived as a broadcast digital system that grows to support interactive and on-demand services. (These services are examined in detail in Part II, "Interactive and On-Demand Services.")
- Multiple access control based on robust digital cryptography—In analog conditional access systems, the actual scrambling method is proprietary. In a digital conditional access system, the payload encryption algorithm can be standardized without reducing the overall security of the system. (See "Multiple Conditional Access Systems," by Michael Adams and Tony Wasilewski.)
- Support many different content providers—Competition between content providers is essential to a cable operator to ensure that the cost of programming remains reasonable. Pegasus was designed to use digital programming services from all available content providers.
- Efficient use of forward and reverse bandwidth—Bandwidth (or spectrum) in a cable system is a limited resource, and much of a cable operator's decision-making revolves around making the best use of cable spectrum. This includes the choice of modulation, topology, channel allocation, system design, and many other factors. The Pegasus architecture formalizes many of the *rule-of-thumb* design rules that have evolved for a hybrid fiber coax (HFC) cable plant.
- Critical-mass, third-party applications platform—One of the most important concepts in the Pegasus architecture is the creation of an applications platform. The set-top hardware and operating system supports the development of *third-party* applications. The set-top becomes an attractive platform to applications developers only when it is deployed in sufficient numbers—that is, when it reaches *critical mass*. Lessons from other industries, namely the game console and personal computer industry, suggest that this critical mass is about one million units.

This last goal is worth examining in more detail. Historically, the set-top has been single-sourced from the same vendor that is responsible for building the headend components. This inevitably leads to closed, proprietary interfaces, as engineers within the same company iterate their way through the network interface design process. To ensure that the Pegasus initiative achieved its goal of a critical-mass applications platform, Time Warner Cable issued a set of requirements as the Pegasus RFP.

Pegasus Request for Proposal

Time Warner Cable issued the Pegasus Request for Proposal (RFP) in March 1996. The introduction included the following passage about the applications platform:

A major goal of the Pegasus strategy has been to provide an application platform that lowers the barrier to entry for applications into the market for interactive services delivered over cable. As such the Pegasus set-top terminal needs to offer a level of platform independence to leverage the software tools, developer experience, and content that exists in the market today. The first step in achieving this platform independence was achieved at the hardware level by introducing multiple vendors for the set-top terminal hardware.

In fact, persuading multiple set-top vendors to build digital set-tops to a set of open standards is probably the most difficult part of the Pegasus program. There are many reasons why set-top vendors are reluctant to build to open standards:

- It is more difficult than building a proprietary set-top—There are many standards to choose from and each has to be interpreted carefully so that all implementations are interoperable.
- Standards are constructed by a committee and are often complicated due to compromise agreements—As someone once said, "A camel is a horse designed by a committee."
- Standards can limit the innovation of a company by spelling out too clearly how to do something.
- Standards are never quite finished—DirecTV learned this when it launched its service with a different transport packet length than the one that was ultimately agreed upon by the MPEG-2 Systems Committee.

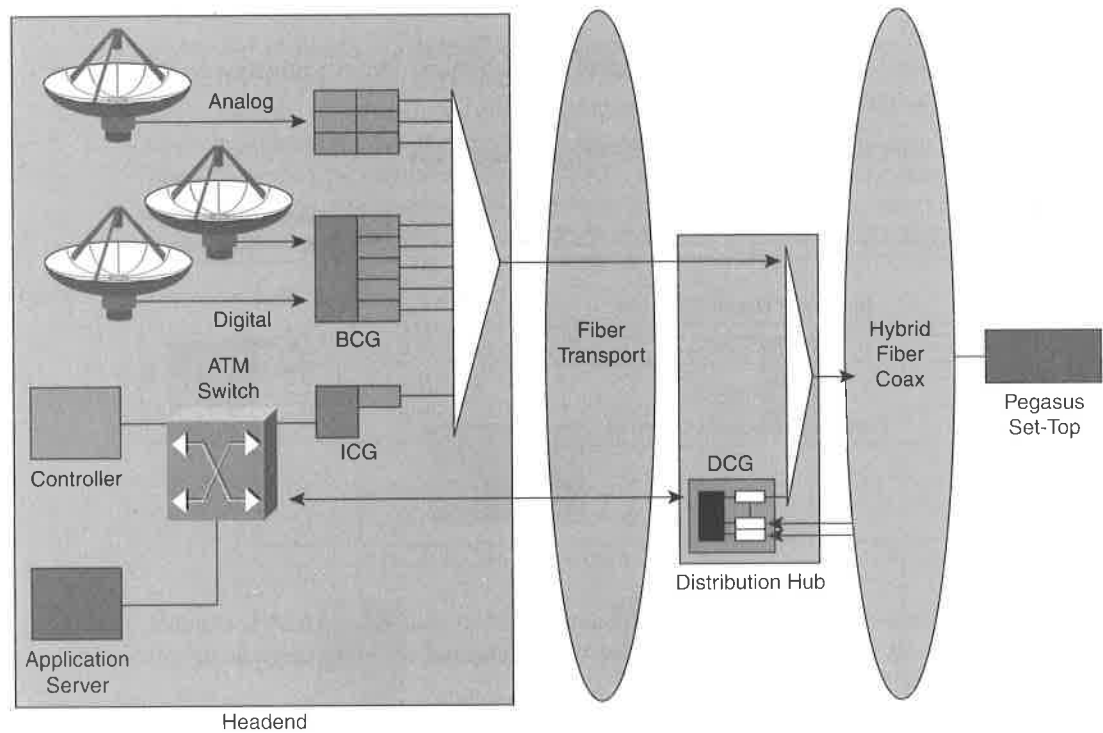
Nevertheless, by 1996 there were many established standards for the transmission and coding of digital television (see the section The Cable Network Interface in Chapter 6). The Pegasus RFP leverages these standards to define the Pegasus network interface, which supports a family of digital set-tops from multiple vendors.

The Pegasus RFP is available on the Web at <http://timewarner.com/rfp>.

Pegasus Architecture

Figure 7-5 illustrates the Pegasus architecture.

Figure 7-5 *Pegasus Architecture*



The major components are defined as *gateways* (the networking terminology that describes a communications device that connects two dissimilar networks):

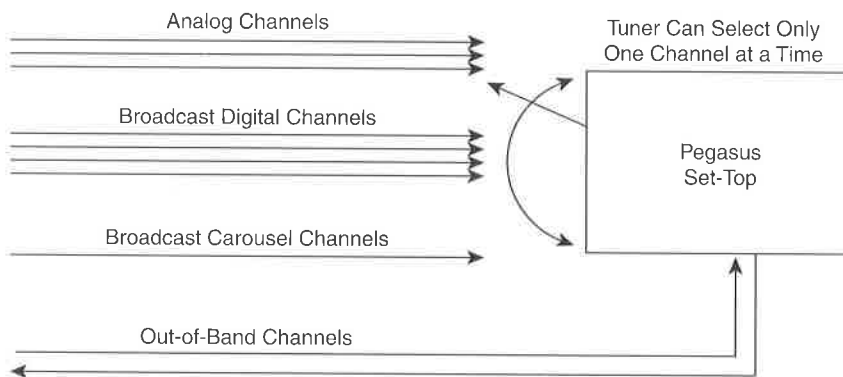
- **Broadcast cable gateway (BCG)**—The BCG takes digital programming from the satellite distribution network and converts it into a form suitable for transmission over a cable distribution network.
- **Interactive cable gateway (ICG)**—The ICG takes data from an asynchronous transfer mode (ATM) network, reassembles the ATM cells, and converts it into MPEG-2 private data format. This allows data communications to the set-top using standard internet protocols, without the need for an ATM adaptation function in every set-top.

- Data channel gateway (DCG)—The DCG takes data from an ATM network and converts it into the DAVIC out-of-band channel format. The DCG function is distributed to the distribution hub locations of the cable system to allow traffic growth to be accommodated.

Digital Channels

Figure 7-6 shows the logical channel connections to a set-top provided by the Pegasus architecture.

Figure 7-6 *Set-Top Channel Connections Within Pegasus Architecture*



There are three types of in-band channel, each modulated as a 6 MHz channel. The set-top has only a single tuner, so only one type of channel can be received at any given time:

- Analog broadcast channels—These channels contain in-the-clear NTSC television channels.
- Digital broadcast channels—These channels are QAM-modulated channels, and each contains an MPTS.
- Broadcast carousel channels—These channels are also QAM-modulated, but they contain a data carousel that is used to provide application data and code to the set-top.

There are two types of out-of-band channel. In contrast to the in-band channels, the out-of-band channel is received by all Pegasus set-tops all of the time:

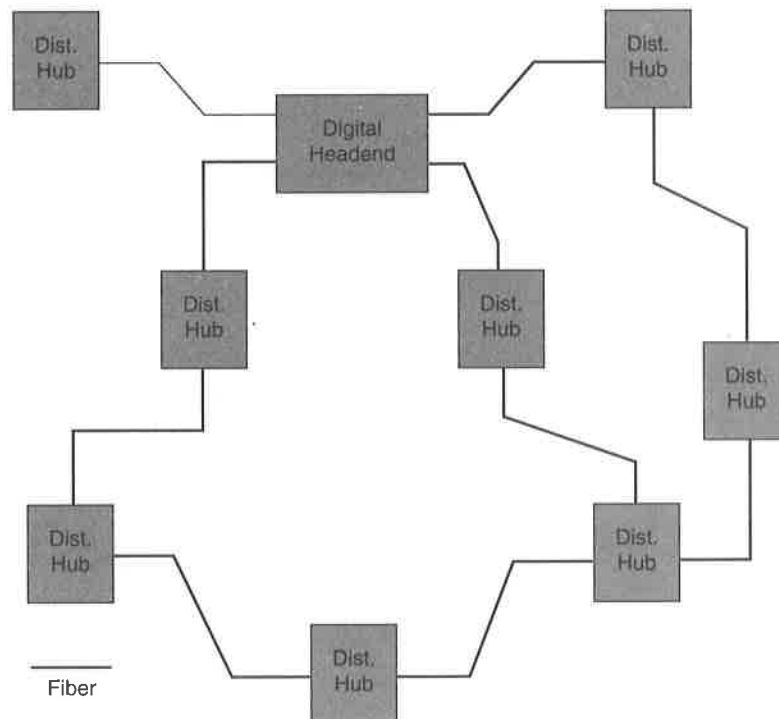
- Forward Data Channel (FDC)
- Reverse Data Channel (RDC)

The Pegasus architecture specifies a standard TCP-IP network to connect the digital set-tops to the headend. This network is owned and managed by the cable operator, not by the cable vendor.

Hub Interconnection

Figure 7-7 shows an example of an upgraded cable system. The system is organized into a single digital headend and a number of distribution hubs. The distribution hubs are interconnected into rings using bundles of fiber. Some fibers are used for analog distribution of the broadband signal out to the fiber node (see Chapter 2, "Analog Cable Technologies"); other fibers may be used to support a SONET connection between the headend ATM network and the data channel gateways in the hubs. (The DCG component is located at the hub for reasons explained in the section Out-of-Band Termination in Chapter 5.)

Figure 7-7 *Upgraded Cable System*



Network Management

The Pegasus RFP specified requirements for remote network management of all components in the system, whether they are located in the headend, the distribution hub, or the customer's home.

Digital Broadband Delivery System

The Digital Broadband Delivery System (DBDS) is Scientific Atlanta's implementation of the Pegasus system.

Figure 7-8 illustrates how the headend processing functions are packaged in the DBDS implementation:

- **Integrated Receiver Decoder (IRD)**—The IRD packages the satellite demodulation and conditional access functions into a single rack-mounted chassis.
- **Broadband Integrated Gateway (BIG)**—The BIG can be used to groom the baseband outputs of several IRDs into a new MPEG-2 multi-program transport stream (MPTS). The output of the BIG is fed into a CA QAM, which applies conditional access and modulation to the MPTS.
- **Condition Access Quadrature Amplitude Modulator (CA QAM)**—The CA QAM applies conditional access and modulation to the MPTS. The CA-QAM has a built in up-converter to place the QAM channel on the desired cable frequency.

So far, there are two suppliers of Pegasus set-tops: Scientific Atlanta (Explorer-series set-top) and Pioneer (Voyager-series set-top).

Headend Equipment

The DBDS is designed to support a local subscriber management model. Figure 7-9 shows a diagram of a locally managed cable system with an HFC upgrade.

Figure 7-8 DBDS Headend Processing

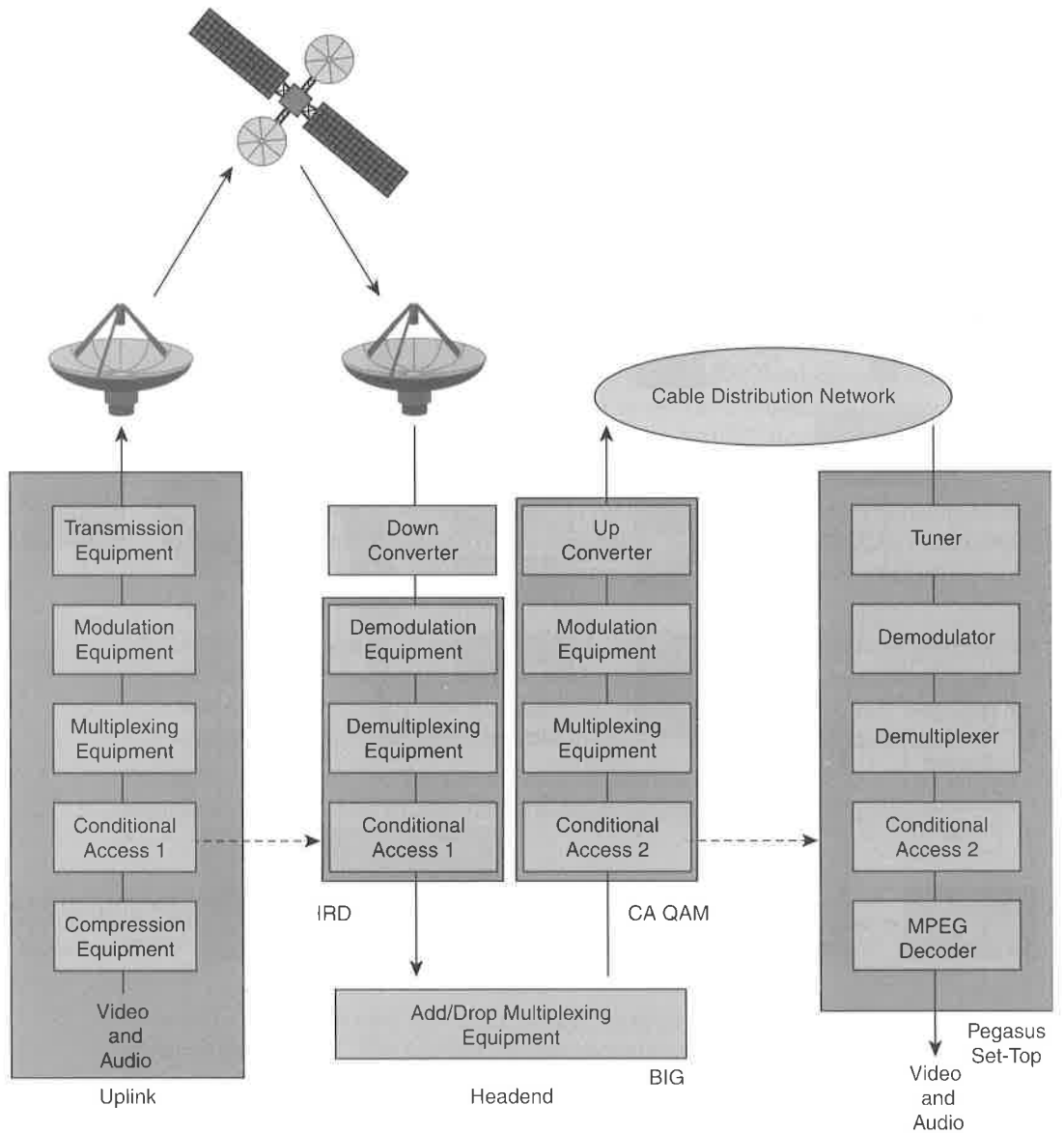
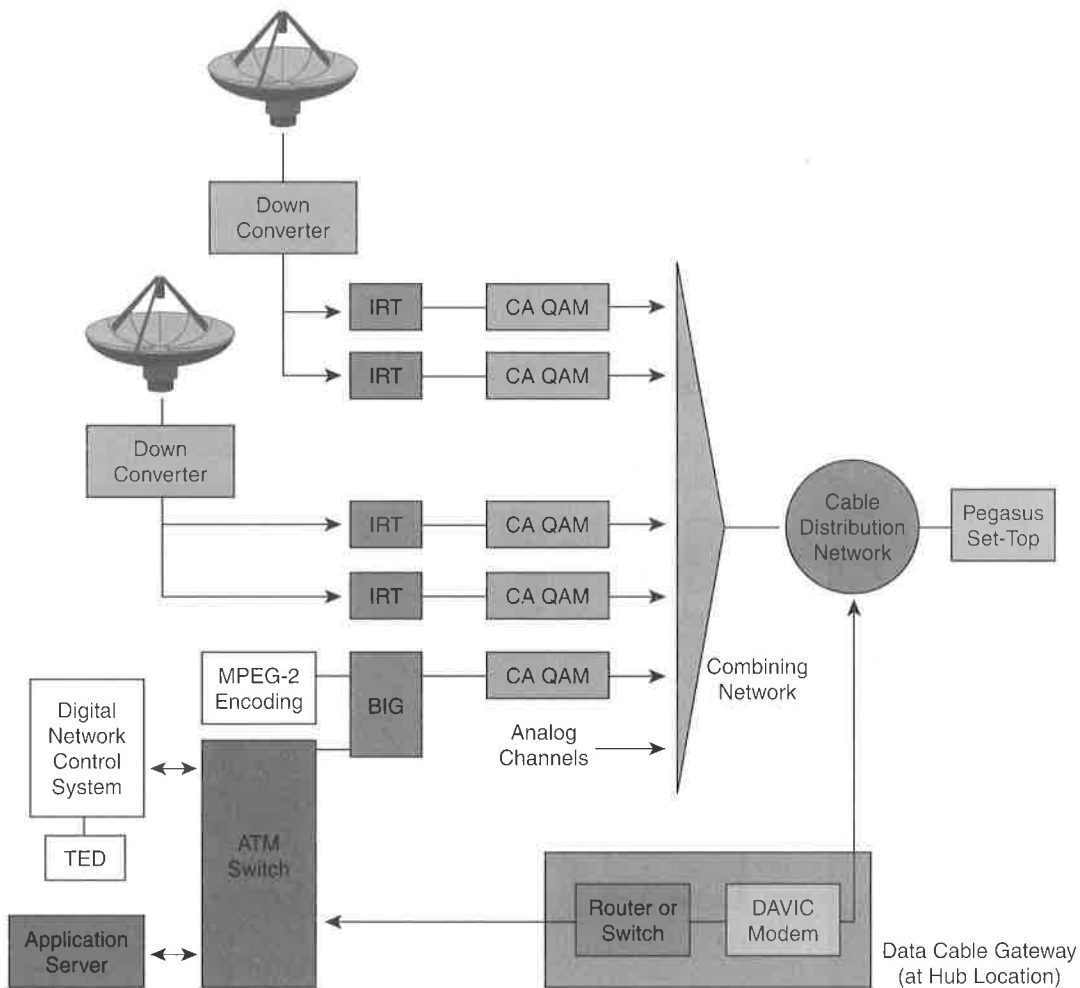


Figure 7-9 HFC Upgrade on a Locally Managed Cable System



The locally managed model includes the following:

- Digital programming—Digital programming is received from satellite using either an Integrated Receiver Transcoder (IRT) for Digicipher II secured channels or an Integrated Receiver Decoder (IRD) for PowerVu secured channels.
- CA-QAM—If no grooming is required, the output of the IRT or IRD is fed directly into the CA QAM. The CA QAM applies PowerKEY conditional access and QAM modulation to the MPTS.

- The headend controller, or Digital Network Control System (DNCS)—The DNCS is responsible for management and provisioning of all the headend components and the Pegasus set-tops. The DNCS also provides the interface to the billing system for service provisioning and IPPV purchase reporting. The DNCS also supports an integrated network management system that is based on the Cabletron Spectrum tool-set.
- The Transaction Encryption Device (TED)—This device provides encryption services to the DNCS to support the PowerKEY conditional access system.
- MPEG-2 encoding—Local analog channels may be encoded and multiplexed together for transmission over a QAM channel.
- The Broadband Integrated Gateway (BIG) add/drop multiplexer—The BIG is used to groom MPTSs received from multiple sources into a single MPTS for transmission over a QAM channel and may also be used to multiplex local encoded channels into an MPTS. The BIG also provides an ATM-to-MPEG-2 adaptation function, which allows private data to be injected into the MPEG-2 MPTS from the application server or DNCS.
- Application server—The application server is a workstation class system that is used to support the server part of applications that are deployed in the DBDS. For example, an applications server supports the delivery of program guide data to client applications in the digital set-top.
- ATM switch—An ATM switch is used to interconnect the headend components and the hub components of the DBDS. ATM was chosen because it is a cost-effective networking technology that is available from multiple vendors. It also provides the quality of service guarantees required to deliver MPEG-2 transport streams.

The billing system may be located on-premise. Alternatively, if a billing service bureau is used, a WAN link is used to connect the headend controller to the billing system.

Hub Equipment

Figure 7-9 shows the DBDS equipment that is located at the hub locations in the cable system.

A router or switch provides a SONET optical interface for interconnection to the ATM switch at the headend. This equipment supports multiple DAVIC modems at the hub as the out-of-band communications traffic increases over time.

The DAVIC modem supports a 1.5 Mbps full-duplex communication to many thousands of Pegasus set-tops.

DBDS Summary

Scientific Atlanta's digital broadband delivery system uses satellite distribution to the headend and provides local subscriber management. Its design supports Time Warner Cable's Pegasus architecture.

Summary

This chapter described the two predominant choices available to a North American cable operator.

The DigiCable system was designed as a solution to the TCI HITS architecture and defines only proprietary interfaces. However, a number of these interfaces have now been published by General Instrument as part of the SCTE Digital Video Standards (DVS) subcommittee work. (Part of the SCTE DVS charter is to publish standards that reflect existing practice in the cable industry.)

The Pegasus architecture was designed as an open system and makes use of many international standards (predominantly from MPEG and DAVIC). Scientific Atlanta publishes any new standards that are developed in SCTE DVS.

References

Book

Ciciora, Walter, James Farmer, and David Large. *Modern Cable Television Technology; Video, Voice, and Data Communications*. San Francisco, CA: Morgan Kaufmann Publishers, Inc., 1999.

Periodical

Adams, Michael and Tony Wasilewski. "Multiple Conditional Access Systems." *Communications Technology*, April 1997.

Interactive and On-Demand Services

Chapter 8 Interactive Services

Chapter 9 Interactive Cable System Case Studies

Chapter 10 On-Demand Services

Chapter 11 On-Demand Cable System Case Studies

Interactive Services

Interactive services provide extensions to the cable system to provide a new class of services. There are almost an infinite number of possible interactive services. Some examples are

- Home shopping
- Home banking
- E-mail
- Web access
- Electronic games
- Stock tickers

Interactive service uses local processing in the set-top combined with data communications services provided by the cable network to deliver new services via the television receiver. Moreover, interactive services can be integrated with broadcast television to provide *enhanced* television services, such as interactive advertising. With this promise, why are interactive services not already widely deployed on cable systems?

There are a number of reasons:

- Although cable operators have experimented with interactive services in trials and limited deployments for many years (see "The Trials and Travails of Interactive TV," in *IEEE Spectrum*, by Tekla Perry), cable operators have yet to deploy, in volume, an application platform that supports interactive services. (See the section Limitations in Chapter 3, "The Analog Set-Top Converter," for a discussion of the limitations of analog set-tops.)
- Interactive services have yet to prove their capability of generating real revenues in large-scale deployment. There is no single *killer* application, so a suite of interactive services is required.
- Television viewing is fundamentally a passive mode of interaction for most television viewers. In addition, many advertisers that support television channels might be reluctant to change the current relationship they have with their audience.

Nevertheless, this situation is ripe for change with the introduction of digital services. A digital cable system already has many of the necessary resources to support interactive applications, and the emergence and acceptance of the Internet are educating customers to expect more from their television sets.

Internet Convergence

Web browsing has increased phenomenally in popularity over recent years and most companies now host a Web site to promote their services or products. The ubiquitous connectivity provided by the Internet enables the Web, but the Internet was not built to support any particular application. The Internet has its origins in ARPANET, which was funded by the Advanced Research Projects Agency (ARPA) (see Chapter 1 of *Internetworking with TCP/IP, Third Edition, Volume 1*, by Douglas E. Comer). In fact, it was 25 years after ARPANET first went online (in 1969) before Web browsing was developed.

The Web originated in universities to facilitate the exchange of information, and these sites still provide the opportunity for serious research. However, the Web has become a tangle of pornographic, commercial, special interest, and vanity publishing sites. The noise of the Web often drowns out the valuable information that is buried in there, somewhere, like a needle in a haystack.

The emergence of the Web has provided a major impetus to interactive services on cable by providing a software toolbox for the applications developer. The challenge is to transfer the Internet-centric technology to a television environment. At first sight, it seems easy to emulate the success of the Internet by offering a Web-browsing service, but this approach is not as simple as it seems:

- Basic digital set-tops simply do not have the memory or performance to run a standard Web browser application written for the PC.
- Most Web sites are designed for presentation on a high-resolution computer monitor display. NTSC was not designed to handle these kinds of images and, in general, Web sites transfer poorly to a television environment.
- The Internet is a totally uncontrolled environment; applications and applets are downloaded at the customers' risk. In a cable system, the cable operator is responsible for the availability of the service and for repairing, or replacing, failed set-tops. Problems due to uncontrolled, untested applications and computer viruses could seriously damage a cable operator, financially and operationally.
- Internet applications are generally provided free of charge. An Internet service provider (ISP) bills the customer for the cost of Internet access. This has been called the "all you can eat" billing model. In contrast, cable operators resell content from a programmer on a service-by-service basis.

Goals of Interactive Services

There are a number of common goals for interactive services, regardless of the specific services offered. These goals must be satisfied for a service to be successful.

Interactive services must add value in some way for the programmer, cable operator, and customer. This added value can come from a number of sources:

- Customer-generated revenue—If an interactive service is useful or compelling, the customer will be prepared to pay for it. For example, advanced analog experience shows that customers are prepared to pay several dollars per month for an interactive program guide.
- Advertiser-generated revenue—If an interactive service can increase the effectiveness of existing commercials or allow the customer to access a companies product and service information, the advertiser will be prepared to pay the cable operator for that service.
- Enhanced customer perception and retention—Increased competition from DBS, MMDS, and others make it even more important to cable operators that their customers are happy with their cable service. Informational services have proven to be useful in terms of enhancing the customer's perception of the overall cable service.

Interactive services must also provide a satisfying experience for the customer. A poorly developed interactive service is worse than no interactive service at all. There are a number of areas where care must be taken:

- Interactive services must be friendly and intuitive to use. The service needs to be easily accessible from the remote control with little or no customer training.
- Interactive services must present an attractive, graphically rich interface. The customer is accustomed to the high production values of most television programming and will not tolerate an uninteresting or static user interface.
- Interactive services must be reliable and highly available. If an interactive service becomes popular with customers, they will be frustrated if it is unreliable or unavailable.

Interactive Versus On-Demand Services

Until recently, interactive television services were considered to be almost synonymous with on-demand services, specifically video-on-demand (see *Interactive Television* by Winston William Hodge). However, video-on-demand is a technology that supports a special class of interactive services and presents its own set of functions and limitations. To distinguish video-on-demand (and all related on-demand services) from interactive services, I will refer to them as on-demand services. On-demand services are described in Chapter 10, "On-Demand Services."

The following definitions serve to differentiate interactive services from on-demand services:

- Interactive services provide a wide-range of services to the customer, including e-mail, home shopping, home banking, network games, and many other services. Many of these services use the cable system as a two-way data communications medium.
- On-demand services also provide an interactive service to the customer but are differentiated by their control of a streaming media service. An on-demand service provides a dedicated video and audio stream to the customer via the cable system *in addition to* using the cable system as a two-way data communications medium.

Interactive Services

Most interactive services that have been developed to date can be grouped into one of the following categories:

- Navigation—Navigation services are actually a set of related services developed originally for the advanced analog set-top.
- Information—Informational services provide some kind of news or information service to the customer—for example, news, weather, sports, stocks, or calendar-of-events information.
- Communication—Communications services support some form of communication between customers—for example, e-mail, chat, telephony, or video conferencing.
- Electronic commerce services—Electronic commerce (e-commerce) provides the customer with a secure monetary transaction of some kind—for example, home shopping, home banking, or the ability to withdraw electronic cash.
- Video games—There are a wide variety of video games that fit into two major categories: single-player games where the user plays against the computer, and networked games where the user plays against another user over the network.
- Enhanced television services—Enhanced television is a term used for any service that takes an existing streaming media service and enhances it with a related interactive service. For example, enhanced commercials, play-along game shows, opinion polls, and surveys.

The next sections consider the interactive services in each of these categories in more detail.

Navigation

Traditionally, navigation services in a cable system include a number of services that are bundled together. These services include

- Analog and digital tuning—This service allows the customer to select the desired channel by using the up-and-down channel button or by direct entry of the channel number.
- Interactive program guide—The electronic program guide (EPG) provides the majority of the navigation functions for a typical customer. The EPG provides rapid access to program schedules and descriptions.
- Impulse pay-per-view (IPPV) interface—The IPPV interface allows the customer to select and purchase IPPV events using the remote control.
- VCR remote control—The VCR remote control functions allow one-touch recording of a selected event and are linked into the IPG and IPPV services.
- Parental control—Parental control allows the customer to block programs based on rating or channel.
- Subscriber preferences and configuration—This part of the navigation service allows the customer to customize the configuration of the set-top.
- Digital music service—The navigation service allows the customer to select music channels and to view a description of each track as it is playing.
- Emergency alert system (EAS) support—EAS messages warn of extreme weather conditions and other hazards. (Although this is more of an informational service than a navigation service, EAS support is usually included in the navigation application.)

As new services are introduced to a cable system, each needs support from the navigation service to allow the customer to select it. New services can be advertised using the EPG, by allocating them to channel numbers, activating them with dedicated keys on the remote, or by means of on-screen menus.

Information Services

Information services naturally fit into the interactive mold when customers know what information they want and they want to access it immediately. The information can be news, weather, sports, stock reports, or any other information. In any case, an interactive service on a cable system can offer significant benefits to the customer:

- An interactive service can display the latest information using text and graphics. The information can be instantly available at any time the customer is watching television.
- The service can be tailored to the area served by the cable system. This kind of local service can provide information that is not readily available elsewhere.
- The service can be tailored to the individual customer according to preferences that he or she has expressed.

Communications Services

Communications services can be provided as interactive services. Most of these services require some kind of additional hardware—for example, a wireless keyboard. There are a number of candidate services from the Internet:

- E-mail—An e-mail service is attractive to customers who don't own a PC, or who don't have access to the Internet.
- Chat—Chat is a very popular application and has already been linked to television programming. Chat allows an online, multiperson conversation using short typed messages that are delivered in real-time.
- IP telephony—IP telephony uses packet-mode transmission and reception of voice conversation. This service holds a great deal of potential, but it is more likely to be successful if it is linked into television programming in some way.
- Video-conferencing—Video-conferencing from the television could become a new mode of social interaction. It is probably a bit further away because of the bandwidth requirements to do effective conferencing.

Electronic Commerce Services

Interactive services that provide for electronic commerce (e-commerce) are a very attractive target for the cable operator. Services could include home shopping, home banking, and even the ability to withdraw electronic cash from your bank account using the set-top. E-commerce has the following attractive properties:

- The amount of data that is required to support a transaction is relatively small.
- The set-top box contains an extremely sophisticated digital security system. This can be used to ensure that electronic transactions cannot be forged or altered.
- The billing model is simple; as in an ATM machine, a per-transaction fee could be charged either to the customer or to the service provider.

Video Games

Video games come in many different shapes and sizes. The dedicated video game console is a highly competitive business, with each supplier leapfrogging the others to give superior graphics, better realism, and more action. The digital set-top cannot match a dedicated game console with its highly optimized architecture for dedicated game playing. Nevertheless, a digital set-top can provide sufficient graphics and CPU performance to run some of the older, *classic*, video games. In addition, most games in the educational game category can be supported by a digital set-top.

There are two main categories of video games:

- **Single-player**—Single-player games match the player against the computer. A single-player game requires no communication with the outside world after it has been successfully downloaded into the set-top.
- **Networked games**—A networked game allows the player to play against other players in a shared *game-space*. Two-way, real-time communications are required to support networked games.

Both kinds of video games can be supported as an interactive service on a digital set-top, with some limitations:

- It is not economically feasible to keep pace with the performance of the latest generation of video game consoles. However, most classic and educational video games can be supported by a digital set-top.
- Some games require more memory than is available in a digital set-top. A solution is to break them into *levels* that are downloaded as the player progresses through the game.

An alternative approach is to use the set-top to provide a video game console with a data communications link into a network game space. In this model, the set-top provides communications resources to the game console via a standard network link—for example, Ethernet. The communications resource is used to download game software and to link game players over the network. Because the video game console is connected to a television, it is physically close to the set-top.

Enhanced Television Services

One of the most exciting interactive services is enhanced television. Enhanced television provides a bridge between the television model, where the viewer is completely passive, and the PC model, where the user is always active. This is achieved by giving the television viewer the option of jumping into an interactive session at certain predefined points in the television program. These predefined entry points are authored into the television content so that they become an intuitive and natural part of the television program.

Enhanced television programming is still in its infancy, mainly because there are few customers with the necessary platform to support it. Nevertheless, enhanced television services show a great deal of promise, and there is a range of potential services:

- **Enhanced commercials**—Enhanced commercials allow the customer to express interest in a broadcast commercial as it is playing. An icon appears during the commercial that tells the customer that this is an interactive commercial. If the customer expresses interest (by pressing a button on the remote control), he is directed to more information about the product or service being advertised and can arrange for a follow-up contact from the advertiser.

- Play-along game shows—In a play-along game show, the customer can match his wits against the contestants and the set-top automatically keeps track of the score. The highest-scoring customers could be invited to participate in future shows (in person!).
- Opinion polls and surveys—This service would allow the programmer to get nearly instantaneous feedback on various issues.

Applications Model

All interactive services are delivered by means of some application software or *application*. The application can provide a single interactive service (for example, an electronic game) or a general-purpose mechanism for delivering services (for example, a Web browser).

The application can run in the set-top or in a server (located at the headend or distribution hub), or it can be distributed between the set-top and the server. An application that runs in the set-top is called the *client* application, whereas an application that runs in the server is called the *server* application. A distributed application has both server and client components.

Client Applications

A client application is entirely contained in the set-top. Client applications are limited by a number of factors:

- Set-top physical resources—These include memory, processor bandwidth, graphics processing, and so on.
- Locality of information—A client application can provide a useful interactive service only if the information is local to the set-top. Information can be broadcast, such as program guide information, but the set-top has a limited capacity to store information.
- Communications—Although it is technically possible for two client applications to communicate, this approach raises so many synchronization, addressing, and security problems that it is essentially useless.

Nevertheless, many useful interactive services can be implemented by a client application. Examples are

- Navigation—Navigation services can be provided for any programming that can be described by means of broadcast guide information. Schedules for television programming, premium, and pay-per-view programming change relatively slowly and can be described by a finite amount of data.
- Games—Single-player video games are implemented as client applications.
- Information services—If the information data-set can be contained on the set-top or broadcast to it, informational services can be implemented as client applications.

- Enhanced television—If the client application code and data is delivered with the television programming, enhanced television services can be implemented as client applications.

Any interactive service that requires two-way communications with the network cannot be implemented as a client-only application. Although client applications are limited, they do have some significant advantages:

- Ease of implementation—Client applications are easy to code and easy to debug.
- No communication resources—A client application consumes no network bandwidth.
- No transactional resources—A client application requires no transactional support from a server. Therefore, client applications are inherently *scalable*—that is, the number of client instances can be increased without limit.

For these reasons, a client application is generally the best solution if it can provide the desired interactive service.

Server Applications

A server-based or *server* application runs on a server (typically located in the cable headend). Server applications are sometime called screen-server applications because the server generates all the information that is displayed on the television screen.

Server applications have been used for many years in traditional computing. The first computer terminals (visual display units or VDUs) provide only display and keyboard input functions. When a set-top is decoding an MPEG-2 stream, it provides a (sophisticated) display function.

A server application does all application processing and graphics rendering in the server and then ships the result to the set-top for display. (Typically, the screen image is compressed as an MPEG I-frame).

Server applications can be used for almost any application except those that require a very low delay between the user input and a screen update—for example, video games. In addition, server applications are easy to implement and can easily be updated with no impact on the set-top. Despite these qualities, they are very limited for mass deployment:

- Communication resources—A server application consumes significant network bandwidth. Regardless of how small the change, each update completely refreshes the screen contents. (Some applications use MPEG P-frames to reduce bandwidth, but the information density is still low.)
- Transactional resources—A server application requires a thread of execution for every set-top that is concurrently accessing the service. Therefore, server applications are inherently difficult to scale.

Server-based applications are generally inferior to client or distributed applications. However, server-based applications can be useful in some circumstances:

- If a set-top cannot support a client application, it is sometimes possible to use a server application instead. This approach can be used to extend the lifetime of digital set-tops that have limited memory space, graphics, or CPU performance. For example, suppose that an interior decoration service becomes popular and that it employs three-dimensional modeling techniques that require 32 MB of DRAM and a 300 mips CPU. Older set-tops can still support the service if it is implemented as a server application and the results transmitted as MPEG-2 frames.
- If the interactive service requires access to a large database, placing the application on the same server as the database can simplify implementation and optimize performance.

Most server-based applications can be better implemented as distributed applications, with both client and server parts.

Distributed Applications

A distributed application model provides the most flexible way to implement interactive services. This model allows the application processing to be shared between the client (set-top) and the server. Typically, the client application provides functions related to input and output (for example, graphics rendering, audio processing, keyboard input, and so on), and the server application provides computational functions. The distributed applications model provides the combined advantages of client and server applications:

- The client part of the distributed application is inherently scalable—Careful design of the client allows the distributed application to be scaled up without placing excessive transactional loads on the server.
- The server part of the distributed application can provide extended resources to the client—For example, the server part can be used to manage a shared database, to provide persistent storage, or to provide fast computation.

If an interactive service cannot be implemented as a client application, a distributed application is usually the best choice. However, distributed applications bring certain challenges:

- Implementation—Distributed applications can be difficult to code and difficult to debug. The programmer has to consider not only the client code and the server code but also the interactions between them.

- Communication resources—A distributed application consumes network bandwidth, which must be carefully managed.
- Fault tolerance—A distributed application operates correctly only when the server, and the communications link to it, are in service. Therefore, redundancy in the network and the server are necessary to make the service highly available.
- Transactional resources—A distributed application consumes server bandwidth. The server usually maintains a thread of execution for every client instance. Therefore, the server part of the application must be carefully designed so that it can support the required number of users.

Application Requirements

There are an infinite variety of interactive applications, each with its own unique presentation, features, and functions. It is helpful to divide interactive applications into a number of categories based on the resources that they require from the cable system:

- Software download—Interactive services are often provided by a client application program that executes in the set-top. Downloading the client application program when it is needed allows many applications to share the set-top memory. For example, a particular electronic game might be played by only 10 percent of customers and can be downloaded from a data carousel only when a customer selects it.
- Activation—The activation of an interactive service is achieved by executing a client application on the set-top. There are a number of different *triggers* for application execution: The user can explicitly request the service, it might be activated in response to in-band data, or it might be activated in response to an out-of-band message. For example, the EAS application is triggered when the set-top receives an emergency alert message.
- Communications—Many interactive services require a communications resource from the network. Different applications require connections with varying bit rates and quality-of-service parameters. For example, an e-mail application might establish a TCP/IP connection to check the customer's mailbox on the mail server.
- Streaming media—Enhanced television services often manipulate broadcast streaming media. For example, an interactive advertising application might generate overlay graphics and check for a remote control button-press.

Software Download

The application client software must be loaded on the set-top before the interactive service can be launched. There are many ways of loading the application, and the loading method is often used to categorize the application:

- Resident application—A *resident application* is normally resident all the time on the set-top. A resident application is usually fundamental to the operation of the set-top, so it makes sense to always have it loaded in memory. For example, the navigation application is usually resident.
- Broadcast application—A *broadcast application* is broadcast in-band as part of a streaming media service. This allows the application to be received simultaneously by all set-tops tuned to the streaming media service and has the advantage of placing no transactional load on the cable system.
- Solicited application—A *download request* is made to the cable system to load a solicited application. A resident or broadcast application usually requests a solicited application as a result of some action by the customer. For example, the customer selects channel 97 and causes the navigation application to download the home-shopping application.

Activation

Interactive services can be initiated in a number of different ways, and the set-top application can be executed in response to a number of different events:

- Synchronous applications—*Synchronous applications* are activated in response to in-band data that is carried as part of a streaming media service. For example, a synchronous application is invoked during a commercial to present some interactive options.
- Asynchronous applications—*Asynchronous applications* are activated in response to an out-of-band message. An example is the delivery of a new mail notification. A notification message is received on the out-of-band channel that activates the e-mail application. The e-mail application fields the message and illuminates the message-waiting light on the set-top.
- Menu applications—*Menu applications* are activated in response to a customer input. The input is usually received from the remote control or keypad, but other input devices, such as a keyboard or joystick, can also trigger application activation.
- Timer applications—*Timer applications* are activated in response to a timer expiration. For example, the unattended recording function supported by the navigator application sets a timer event. When the timer expires, it activates the unattended recording application.

Communications

Applications have a wide range of communications requirements, depending on the offered service and the application design. A well-designed application will always attempt to minimize its use of communications resources; such an application is sometimes called a *green* application:

- One-way applications—A *one-way application* uses only one-way communication resources. Nevertheless, such an application can provide *local* interactivity. An example is the interactive program guide application. The guide data is broadcast to the set-top and saved in memory. The customer interacts with the guide application locally, to find and display the program listing, without using further communications resources

All other interactive applications use two-way communications, but there are many different ways of providing two-way communications services in cable systems (see Chapter 5, "Adding Digital Television Services to Cable Systems"). Two-way applications can be subdivided into several categories based on the quality of service of the return channel (see "Residential Broadband Internet Services and Applications Requirements," by Timothy Kwok, in *IEEE Communications*):

- Store-and-forward—A *store-and-forward* communications resource is provided by advanced analog set-tops and by some first-generation digital set-tops. These set-tops are able to transmit data only when polled by the headend controller. A good example of a store-and-forward application is impulse pay-per-view (IPPV). The customer uses the IPPV application to order a movie and, as a result, access is granted to an encrypted channel for a period of time. A record of the purchase is stored in the secure microprocessor until it can be retrieved by a poll from the headend controller. Store-and-forward applications place a minimal transactional load on the headend controller and require minimal data transmission capacity in the network.
- Telephone-return—*Telephone-return applications* are very similar in their limitations to store-and-forward applications. They are typically used only to support IPPV services, but recently some providers have experimented with using them for applications that generate a real-time transaction. For example, an enhanced commercial that is designed to stimulate an impulse purchase. During the commercial, the customer is given the option of requesting more information by pressing a button on the remote control. If the customer decides to buy the advertised item (concert tickets, a compact disk, and so on), the application dials a toll-free number using the telephone modem in the set-top.
- Best-effort QoS—A two-way communications resource with a *best-effort* quality of service (QoS) can be used to support a wide range of services. These include home-shopping, home-banking, e-mail, Web browsing, and so on. These applications can provide an acceptable service even if the two-way communications resource supports only a best-effort QoS—that is, the data rate and the delivery of the data packets are not guaranteed. For example, a home-banking application that establishes a secure

TCP-IP session with the service provider via the cable system can tolerate best-effort QoS. If data packets are lost, the TCP-IP protocol will retransmit the packets. If the data rate is low, the customer will experience some delay in the service. Nevertheless, the customer will be satisfied as long as the network is engineered to provide sufficient capacity even during busy periods.

- **Guaranteed QoS**—Some services require QoS *guarantees* from the cable network. QoS guarantees specify a number of parameters, including packet delay, jitter, error rate, and loss rate. For example, a network game application might be designed to tolerate delays of up to 500 milliseconds, whereas a video-conferencing application might tolerate a delay of only 100 milliseconds.

Streaming Media

Interactive services are often linked to streaming media services. Interactive services that are ported from the Internet and provide a static visual display (text and graphics) to the television environment often look uninteresting. However, generating a streaming media session and allocating network resources for each customer is in the domain of on-demand services (see Chapter 10). Therefore, considerable effort has been put into linking applications to a broadcast stream to link interactive services into the viewing of regular broadcast television services (see "Interactive Multimedia Services to Residential Users" by Thodore Zahariadis and others in *IEEE Communications*):

- **Media-linked applications**—*Media-linked applications* are linked to a streaming media service and are *authored* to work in conjunction with the content of the streaming media. Most enhanced television services fall into this category—for example, a commercial that is enhanced with a media-linked application. The application uses data in the streaming media to allow the customer to interact with the commercial to get more information or to request follow-up from a retailer.
- **Media-related applications**—*Media-related applications* are not linked with the content of a streaming media service but are somehow related to streaming media services. An example of a media-related application is the interactive program guide—it is not linked to the content of any of the streaming media services, but it does allow the customer to select the desired service.
- **Media-independent applications**—*Media-independent applications* are completely independent of streaming media services. Native Internet applications, such as e-mail, chat, and Web browsing, are media-independent. However, it is easy to conceive of a media-linked or media-related version of a Web browser. A media-linked browser might allow the customer to surf directly to the home page of an advertiser by pressing a remote control button during a commercial. A media-related Web browser might enhance a Web page with streaming media services; for example, the HBO Web page could include a small panel to show what is currently playing on HBO.

Application Resources

Each interactive application requires its own set of resources in the set-top, the network, and the server. An interactive cable system provides a set of mechanisms to provide resources to the application. These mechanisms must be supported by the hardware components and by networking protocols. Fortunately, many of the mechanisms have already been developed by the computer industry in the field of distributed computing (see *Internetworking with TCP/IP*, Third Edition, Volume 1, by Douglas E. Comer).

The Time Warner Full Service Network (see Chapter 9, "Interactive Cable System Case Studies") was developed using distributed computing techniques and pushed the applications envelope for a cable system. However, many issues in large distributed computer networks do not yet have good solutions. Therefore, using a broadcast model to support application resources is the best approach where it is feasible. For example, software download is generally accomplished using a point-to-point file transfer protocol in a distributed computing environment. In a broadcast cable system, software download uses a data carousel channel to continuously broadcast the file.

The following sections describe

- Set-top resources—These include CPU performance, graphics acceleration, memory, applications environment, and communication resources.
- Software download mechanisms—Data carousels, MPEG-2 Private Data, and the out-of-band channel can be used to support software download.
- Activation and synchronization mechanisms—These include resources required in a multimedia operating system to support applications.
- Communication mechanisms—These include one-way and two-way communications resources.

Set-Top Resources

A digital set-top that is designed to support interactive services must provide more resources than a basic broadcast digital set-top in the following areas:

- CPU performance
- Graphics acceleration
- Memory
- Applications environment
- Communications

CPU Performance

Digital set-tops designed to support interactivity incorporate a faster CPU than a broadcast digital set-top—offering an instruction rate of between 50 and 200 mips. The faster CPU offers a number of advantages to the cable operator:

- There is adequate performance to support a wide range of applications, including fast-action video games.
- Application efficiency is not so critical, and this enables portable applications environments, such as HTML, JavaScript, and Java.
- An interpreted language environment (such as Java) allows real-time bounds checking of all instructions. This provides an effective way of limiting the damage inflicted by a poorly tested application.

Graphics Acceleration

Many interactive applications need graphics acceleration to produce fast, rich, smooth graphics. Graphics acceleration is often required to support video games and can be used to enhance navigation services with animation effects. Graphics acceleration is particularly important in a television environment, because television is a highly visual medium with excellent production values—poor graphics look especially bad in this environment.

Memory

Most interactive services are implemented as client or distributed applications. In either case, the client code requires memory space to load and execute. A broadcast digital set-top might have as little as 1 MB of application memory, whereas an interactive set-top requires at least 4 MB.

Applications Environment

A developer-friendly applications environment is needed to allow third-party applications developers to add their applications to a digital set-top. The applications environment must support resource management so that multiple applications can share the set-top resources, including window management, memory management, and input-device management.

Communications

An interactive set-top must support communications resources to allow software download of applications. In addition, real-time, two-way communications resources are essential to support many interactive services.

Software Download Mechanisms

Placing an application in a set-top using software download over the cable network has significant advantages for the cable operator:

- New applications can be added at the headend and propagated to any set-top over the cable network.
- Applications can be stored at the headend and downloaded to the set-top only when required. This saves memory in the set-top.
- Applications can be updated and any bugs fixed. The new version can be downloaded to the set-top when required.

However, downloading applications each time they are needed has some serious disadvantages:

- It takes time to download an application. If the application is small, it can be downloaded over a high-speed channel in a fraction of a second. However, downloading a large application introduces delays and consumes large amounts of network capacity.
- If the download server or download channel is unavailable, the customer will see a loss of service. Making certain applications resident—for example, the navigator—increases the availability of the services provided by those applications.

Several download mechanisms are commonly used in interactive cable systems:

- Data carousel—The *data carousel* provides a mechanism to download applications software or data by continuously transmitting a set of files over a QAM channel. A standard mechanism is provided by the DSM-CC data carousel (see the standard DSM-CC, Digital Stored Media—Command and Control).
- MPEG-2 private data—An MPEG-2 transport stream can be used to download applications software or data by placing the data in a separate program elementary stream (PES).
- Out-of-band channel—The out-of-band channel provides a point-to-point datagram service between the set-top and a headend server.

Data Carousel

DSM-CC specifies a data carousel mechanism that allows information to be broadcast to a population of digital set-tops (see DSM-CC, Digital Stored Media—Command and Control). All, or part, of a QAM channel can be used to transmit a set of files in a continuous cycle. When a set-top requires a file, it tunes to the carousel and listens to the carousel until it retrieves the file it needs. Typically, several data carousels are used, each for a specific application—for example, program guide data, application download, and so on.

MPEG-2 Private Data

MPEG-2 transport supports the carriage of arbitrary binary data with MPEG-2 private data sections. MPEG-2 private data is used to carry application code and data as part of the MPEG-2 program to support enhanced television applications. When a set-top selects the MPEG-2 program, it activates a loader application, which listens to the MPEG-2 private data. As the enhanced television applications are received, the loader application places them in memory and activates them.

Out-of-Band Channel

The out-of-band channel provides a two-way communications path between the set-top and the headend. This path can be used to support standard download protocols, such as Boot Protocol (BOOTP) or Trivial File Transfer Protocol (TFTP). However, the out-of-band channel could easily become saturated with download traffic if it were used to load all interactive applications. For this reason, use of the out-of-band channel for download should be discouraged as a policy decision by the cable operator.

Activation and Synchronization Mechanisms

All applications need to be activated after they have been successfully downloaded. The activation might be immediate or synchronized to some event, such as a message or a timer expiration. In any case, the operating system is responsible for activation.

Synchronous applications need to synchronize their execution to streaming media. The application is authored with the streaming media and synchronized to it using the MPEG-2 display timestamps. Even so, the timing of the application download must be carefully planned by the application developer—the application must be loaded into memory before it can be activated. However, it must not be downloaded too early; that way, a previously loaded application has time to release memory and exit.

A multimedia operating system provides facilities for the actual synchronization of the application to the MPEG-2 display timestamp. (An application cannot be granted access to the MPEG-2 display hardware.) The operating system dispatches an event to the application at the appropriate display time. Examples of operating systems with synchronization facilities are PowerTV and OpenTV (see the section Internet Resources, later in this chapter, and see Chapter 9).

Communication Mechanisms

Many interactive applications require a communications resource. There are two main types of communications resource in an interactive cable system:

- One-way communications—In this case, the information required by the application must flow over a known channel so that the application can locate it.
- Two-way communications—In this case, some kind of session must be established by the application with a server.

Each type of communications has its own set of problems and solutions.

One-Way

In one-way communications, there are several modes of communication: broadcast, multicast, and unicast. Each has its uses according to the type of application:

- Broadcast—In this case, information that is useful to all applications is simply broadcast over the entire network. An example is program guide data or embedded MPEG-2 private data associated with a broadcast program.
- Multicast—Multicast communications are useful when the information is useful only to a group of set-tops. In a one-way network, the cable hub provides a convenient multicast group, which is used by SI and EAS messages.
- Unicast—Unicast communications are used for CAS messages in a one-way environment, but although the message is destined for only one set-top, it must be broadcast to all parts of the network where that set-top might be located.

Two-Way

In two-way communications, the same modes of communication exist in the forward direction (from the network to the set-top) as in one-way communications. However, the communications in the reverse direction are always unicast to a particular address in the network:

- Broadcast—In this case, the information is useful to all applications and can be simply broadcast over the network. An example is the broadcast carousel or embedded MPEG-2 private data in a broadcast program. The broadcast mode is used in two-way networks but must be designed never to elicit a response from the set-top, because this could trigger a *broadcast storm* of responses.
- Multicast—Multicast communications are useful when the information is sent to a group of set-tops. Communications protocols have been developed to support the general definition of multicast groups, and these can be applied to two-way networks. Thus, an application can signal on the reverse channel to add itself to a multicast group while it is active. Pointcast is an example of such an application.

Communication Mechanisms

Many interactive applications require a communications resource. There are two main types of communications resource in an interactive cable system:

- One-way communications—In this case, the information required by the application must flow over a known channel so that the application can locate it.
- Two-way communications—In this case, some kind of session must be established by the application with a server.

Each type of communications has its own set of problems and solutions.

One-Way

In one-way communications, there are several modes of communication: broadcast, multicast, and unicast. Each has its uses according to the type of application:

- Broadcast—In this case, information that is useful to all applications is simply broadcast over the entire network. An example is program guide data or embedded MPEG-2 private data associated with a broadcast program.
- Multicast—Multicast communications are useful when the information is useful only to a group of set-tops. In a one-way network, the cable hub provides a convenient multicast group, which is used by SI and EAS messages.
- Unicast—Unicast communications are used for CAS messages in a one-way environment, but although the message is destined for only one set-top, it must be broadcast to all parts of the network where that set-top might be located.

Two-Way

In two-way communications, the same modes of communication exist in the forward direction (from the network to the set-top) as in one-way communications. However, the communications in the reverse direction are always unicast to a particular address in the network:

- Broadcast—In this case, the information is useful to all applications and can be simply broadcast over the network. An example is the broadcast carousel or embedded MPEG-2 private data in a broadcast program. The broadcast mode is used in two-way networks but must be designed never to elicit a response from the set-top, because this could trigger a *broadcast storm* of responses.
- Multicast—Multicast communications are useful when the information is sent to a group of set-tops. Communications protocols have been developed to support the general definition of multicast groups, and these can be applied to two-way networks. Thus, an application can signal on the reverse channel to add itself to a multicast group while it is active. Pointcast is an example of such an application.

- Unicast—Most two-way communications are point-to-point—that is, they use a unicast mode in the forward and return direction. This mode is used to establish a session between the client and server parts of a distributed application and requires some mechanism for them to discover each other.

Chapter 9 provides some practical examples of the implementation of these various types of communications resources.

Summary

This chapter considered the development of interactive services for cable systems. To be useful, interactive services must satisfy some basic goals and usability criteria that are significantly different than those for PC applications. They must be designed to fit into the television services so that they support, enhance, and extend them in a natural, intuitive, and friendly way.

An almost infinite variety of interactive services can be conceived, but they can be grouped into navigation, information, communication, electronic commerce, video games, and enhanced television applications. Some applications span categories—for example, an enhanced television application can also provide a communications service.

Interactive services are created for applications that can run in the set-top (client applications), the server (server applications), or both (distributed applications). Each type of application has its own advantages and disadvantages, but a rule of thumb is that client applications are easiest to implement. Distributed applications provide the most general model and are most suitable for use in a two-way environment.

Each application requires a different set of resources from the set-top, network, and server components. Application resources are divided into set-top, software download, activation and synchronization, and communications resources. An interactive cable system must provide resources to the application in all these areas to support interactive services.

References

Books

Comer, Douglas E. *Internetworking with TCP/IP, Third Edition, Volume 1*. Upper Saddle River, NJ: Prentice Hall, 1995.

Hodge, Winston William. *Interactive Television*. New York, NY: McGraw-Hill, 1995.

Periodicals

Kwok, Timothy, "Residential Broadband Internet Services and Applications Requirements." *IEEE Communications*, June 1997.

Milenkovic, Milan. "Delivering Interactive Services via a Digital TV Infrastructure." *IEEE Multimedia*, October–December 1998.

Perry, Tekla. "The trials and travails of Interactive TV." *IEEE Spectrum*, April 1996.

Zahariadis, Thodore, Corrado Rosa, Marco Pellegrinato, Anita Byork Lund, and George Stassinopoulos. "Interactive Multimedia Services to Residential Users." *IEEE Communications*, June 1997.

Standards

ISO/IEC 13818-6. "DSM-CC—Digital Stored Media—Command and Control."

Internet Resources

OpenTV

<http://www.opentv.com>

PowerTV

<http://www.powertv.com>

Interactive Cable System Case Studies

This chapter discusses two interactive cable systems—the Time Warner Full Service Network and the Time Warner Pegasus program. These case studies are closely related; the same core team of engineers is responsible for both programs, and the Pegasus program has many of the same goals as the Full Service Network.

Time Warner Full Service Network

The Full Service Network (FSN) was conceived primarily as a vehicle for video-on-demand (VOD) services. In fact, in the early 1990s, interactive and on-demand services were considered to be much the same thing. VOD was the major goal of the FSN for a number of reasons. First, VOD has been the Holy Grail of engineers for many years. The idea that video can be captured, stored, and manipulated using digital techniques is fascinating, if only because it is such a wonderful engineering challenge. Second, the limitations of pay-per-view programming can be frustrating, and evidence suggests that customers really want a VOD service and are prepared to pay for it. This evidence is backed up by the success of the video rental business. In a short time, the (once-laughable) idea of renting videotapes of movies to customers for a fee has become a multibillion-dollar-a-year business.

Although the FSN was conceived as a VOD service, many other services were added to the rapidly evolving specifications in early 1993. These included other on-demand services, such as *sports-on-demand*, *news-on-demand*, and even *HBO-on-demand* (internally, these became known affectionately as SOD, NOD, and HOD). In addition, many interactive services were included in the FSN umbrella, including video games, home shopping, and later, even Web browsing. The framework for interactive services is an important FSN contribution and is discussed in detail in this chapter, but the development of on-demand services is examined separately in Chapter 10, "On-Demand Services."

FSN Network Architecture Goals

Time Warner Cable issued "a request for assistance in building a full service telecommunications network," in February 1993. At that stage in the FSN project, the service definition was broadly stated as "traditional CATV services, full video-on-demand

with instant access, interactive television, interactive gaming, long distance access, voice, video telephone and personal communication services."

In contrast, the network architecture definition was precise and described the infrastructure to support the services in considerable detail:

"Our current vision for this project involves:

- a server at the headend,
- a switching system capable of performing all the necessary routing functions from the server to any of the full service customers,
- assembling of the data on a node by node basis using add drop multiplexing or equivalent technologies,
- digital modulation devices using QAM modulation or equivalent techniques,
- optical links from the headend to the node locations fed by single or multiple laser configurations,
- coaxial plant to the customers homes, and
- appropriate customer terminal equipment to translate the bit streams into usable services."

From its inception, the FSN was envisioned to provide a broad range of services, most of which were not fully defined at the beginning of the project. The FSN designers understood that to achieve this kind of flexibility, a *service-independent* approach was required.

Traditionally, telephone, computer, and cable networks are all designed for a particular service, and, as a result, they can carry only a single traffic type. Carrying video, voice, and data over a single network became the fundamental goal of the FSN, because achieving that goal enabled the broad service flexibility that was required to meet the FSN's (deliberately) open-ended vision.

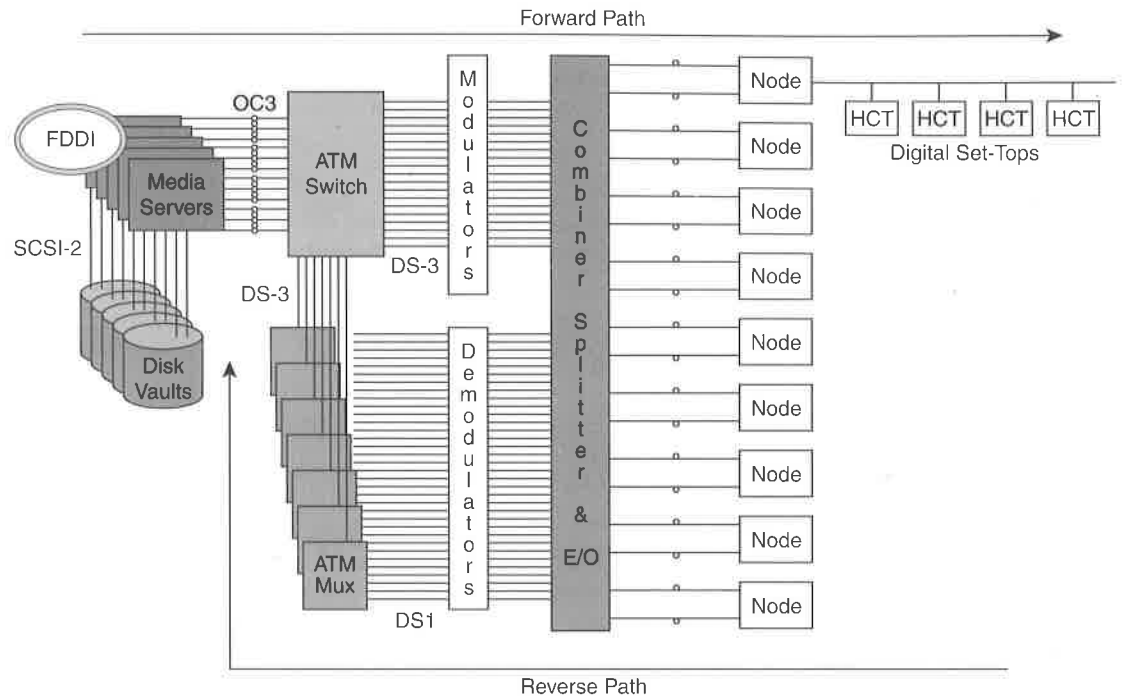
The FSN was also intended to prove that interactive services were feasible in a real cable system with a significant number of subscribers. Many trials of interactive services with only 50 or 100 subscribers have failed to discover any of the problems of system scaling. As a complex system increases in size, second- and third-order effects come into play—the problems of scaling. The FSN subscriber population was chosen to be 4,000 subscribers because this number is sufficiently large that scaling problems could be discovered (and hopefully resolved) during the design process.

To support interactive services, the FSN also specified a high-speed reverse data path from the set-top to the headend. The reverse data path transformed the cable system into a two-way system—it could be used to provide low-latency signaling for interactive games, for telephone conversations, or for video conferencing.

Network Overview

Figure 9-1 provides an overview of the FSN network. The description is divided into forward and reverse paths for clarity.

Figure 9-1 Full Service Network



Forward Path

Eight media servers (SGI Challenge XL) are connected to disk vaults using fast and wide SCSI-2 interfaces. The disk vaults can be configured to provide a total of 3,000 gigabytes of media storage capacity, enough for about 500 movies.

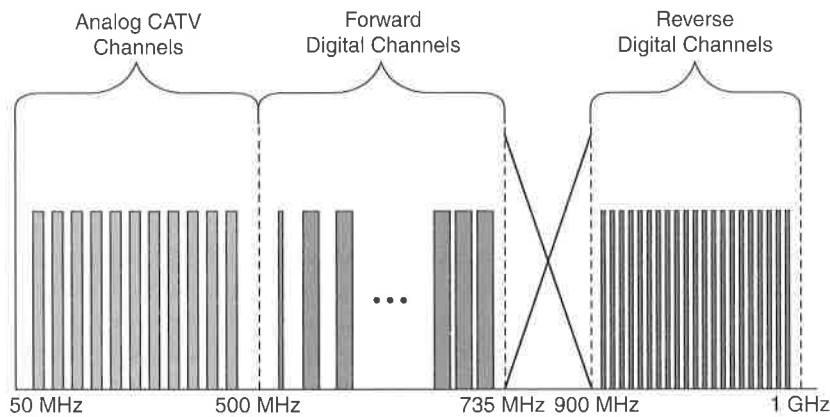
The media servers are connected to an AT&T GCNS-2000 ATM switch. Each server is connected to the switch with six SONET OC3 connections. A total of 48 OC3s provide a total of 5,184 Mbps of usable payload bandwidth (after accounting for SONET and ATM overhead).

The media servers are also interconnected with an FDDI ring, which is used to transfer media content to the disk vaults and to collect billing records from the servers at 100 Mbps. (A separate FDDI ring was used to expedite development with the understanding that in the future ATM switching can support all communications.)

The ATM switch is connected to a bank of 64-QAM modulators supplied by Scientific Atlanta. 152 unidirectional DS3 links provide a total of 5,600 Mbps of payload capacity from the ATM switch to the neighborhoods (after accounting for SONET and ATM overhead).

The QAM modulator outputs are tuned in the frequency range 50 to 735 MHz, and spaced at 12 MHz. This allows the outputs to each neighborhood to be combined into a broadband RF signal. Conventional analog television channels (spaced at 6 MHz) are also combined. The spectral frequency diagram is shown in Figure 9-2.

Figure 9-2 Full Service Network



The composite RF signal from 50 to 735 MHz is used to amplitude-modulate a laser. The laser is coupled to a single-mode fiber, which takes the signal out to the neighborhood about 10 miles away. At the neighborhood, the optical signal is converted back into the RF domain by the fiber node and used to feed a coaxial feeder network, which passes about 500 subscribers.

The RF signal enters the subscriber residence and feeds the home communications terminal (HCT) or digital set-top converter. The HCT is actually a powerful RISC-based multimedia computing engine with video and audio decompression and extensive graphics capabilities. The HCT also incorporates an analog set-top converter (the Scientific Atlanta 8600X) to tune analog channels.

An ATM addressing scheme is designed to allow any server to send data to any HCT.

Reverse Path

The HCT transmits a QPSK-modulated signal in the 900 to 1000 MHz band. Reverse carrier frequencies are defined at a spacing of 2.3 MHz. The QPSK channel supports a data rate of 1.152 Mbps (after accounting for ATM overhead). Each reverse channel is slotted using a time division multiple access (TDMA) scheme. This allows a single reverse channel to be shared among a number of HCTs. The time-slot assignments are made at the headend and sent to the HCT over a forward channel so that only one HCT is enabled to transmit in any given time slot. By default, each HCT has access to a constant bit rate ATM connection with a bit-rate of 46 Kbps. More importantly, the access latency of a typical packet is 25 ms, worst-case.

The reverse channels from a neighborhood are transported by the coax plant back to the fiber node. At this point, the reverse spectrum is used to modulate a laser, which is coupled to a single-mode fiber. Separate fibers (in the same cable sheath) are used for the forward and reverse directions.

At the headend, the optical signal is first converted back into the RF domain, and then it is fed to a bank of QPSK demodulators. These convert the cell stream into an ATM-format T1 link.

The outputs of the demodulators are combined by seven Hitachi AMS5000 ATM multiplexers. A standard, bidirectional ATM-format DS3 is used to connect each multiplexer to the ATM switch.

The ATM switch passes the reverse data to the media servers. An ATM addressing scheme is designed to allow any HCT to send data to any server.

ATM Addressing

The ATM switch and multiplexers are configured with a mesh of permanent virtual connections (PVC).

In the forward path, ATM virtual paths are configured from each OC3 port to each DS3 port. This allows the server to address any forward applications channel (FAC) by selecting the appropriate VPI. The HCTs tuned to a FAC ignore the VPI and reassemble connections based on VCI. This creates a two-level switching hierarchy—VP switching in the ATM switch and VC switching in the neighborhood—and allows any media server to send data to any HCT.

In the reverse path, the ATM multiplexers perform a traffic aggregation function from DS1 to DS3 rates. A mesh of virtual paths is provisioned from the multiplexer DS1 ports to the server OC3 ports. This allows any HCT to send cells to any server by using the appropriate VPI.

Connection Management

The connection manager is composed of a distributed set of processes that run on the media servers. In response to an application request for a connection with a given quality of service, the connection manager determines a route, allocates connection identifiers, and reserves link bandwidth. The connection identifiers are passed to the server and client applications at the media server and HCT, respectively.

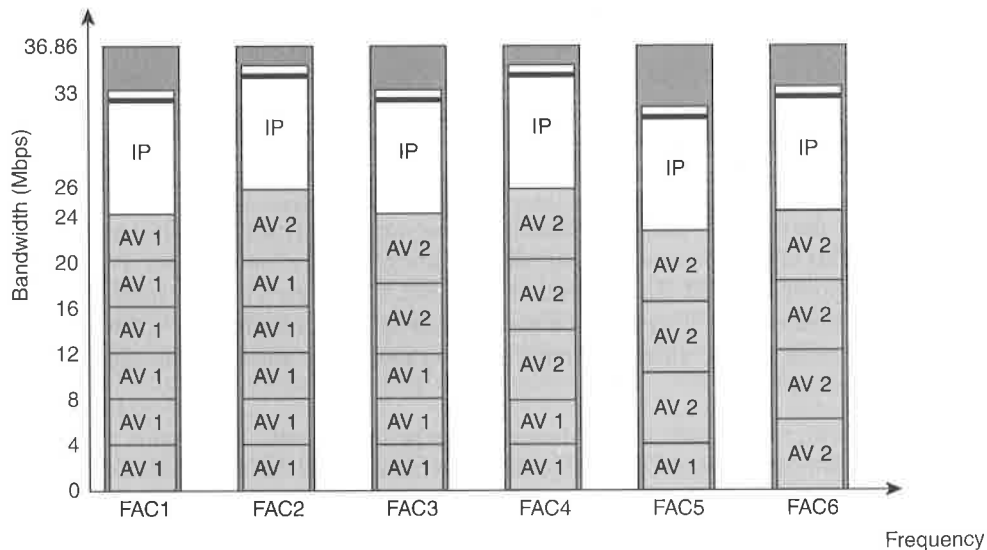
On-demand services use the connection manager to establish a constant bit rate ATM connection for each media stream from a server to the HCT. This approach is used to guarantee the quality of service of the connection so that the cell loss rate is less than 10^{-10} , sufficient for high-quality delivery of MPEG compressed video streams.

Connectionless Signaling

The allocation of a connection for each *application* request would create far too much overhead in the distributed application environment of the FSN. In practice, only on-demand services generate requests to the connection manager, and all other communications sessions are created using a connectionless network.

Figure 9-3 illustrates how the forward application bandwidth is shared between on-demand connections for audio and video (labeled AV) and static connections for IP networking (labeled IP).

Figure 9-3 Forward Applications Channel Partitioning



Each HCT is allocated *three* IP addresses when it is booted—one for each of the three IP networks:

- Fast IP—This network is created using a fixed, 8 Mbps connection in each forward applications channel.
- Slow IP—This network is created using a fixed, .714 Mbps connection in each forward applications channel.
- Control IP—This network is created using the forward QPSK channel and has an approximately 1 Mbps capacity.

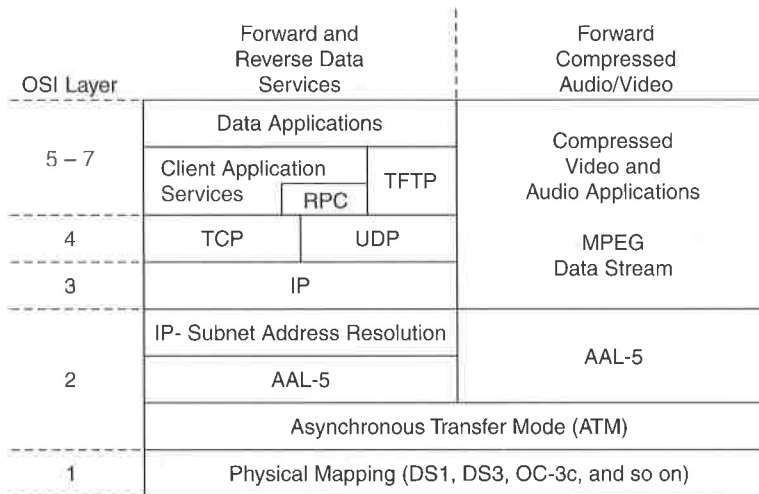
Three IP networks are used for different applications and to work around some of the deficiencies of the HCT:

- A fast IP network is mapped into an 8 Mbps ATM connection over each Forward Applications Channel. Fast IP is used for application download, which is initiated by a file transfer request from the HCT. The HCT is capable of receiving data at bursts of up to 8 Mbps when it is dedicated to downloading a file. Applications are compressed to reduce download time even further, and the largest application can be reliably loaded in less than one second.
- The slow IP network is mapped into a 0.714 Mbps ATM connection over each Forward Applications Channel. Slow IP is used for general communications between the client and server parts of a distributed application. The HCT is capable of receiving and transmitting data at this lower rate while the client application is executing.
- The control IP network is mapped into a 1 Mbps ATM connection over each Forward Control Channel. Control IP is used for general control signaling to all HCTs in a neighborhood.

Although the requirement for connection management of on-demand sessions was obvious at the start of the development of the FSN, the need for connectionless signaling was less apparent. Distributed applications typically have less stringent quality of service requirements than on-demand applications, but distributed applications might use multiple sessions for short bursts of communication. For this reason, distributed applications require a connectionless signaling mechanism, such as that provided by IP networking.

Figure 9-4 summarizes the communications protocol stack for the FSN. The connectionless signaling protocols that support the distributed applications environment are on the left of the diagram. The connection-oriented protocols that support the streaming services are on the right of the diagram.

Figure 9-4 FSN Protocol Stack



Services

Interactive services provided by the FSN include navigation, games, and home shopping. (VOD services are described in Chapter 11, "On-Demand Cable System Case Studies.")

Navigation

Navigation services provided by the FSN include

- Analog tuning
- Interactive program guide—The Preview guide look-and-feel is implemented on the HCT by an application that allows the customer to scroll through a grid of program descriptions over time and channel. (The Preview guide is now known as *TV Guide*.)
- Parental control—Parental control allows the customer to block programs based on rating or channel.
- Subscriber preferences and configuration—This part of the navigation service allows the customer to customize the configuration of the HCT.

In addition, the navigator maps each major category of interactive service (or *venue*) to a channel number between 90 and 99, as shown in Table 9-1.

Table 9-1 *Interactive Services and Channel Assignments*

Channel	Service
90	Music
91	Sports
92	Education
93	News
94	Services
95	Controls
96	Games
97	Shopping
98	Movies
99	TV

Mapping interactive services to channel numbers provides an alternative to the three-dimensional *Carousel* navigator. When the HCT is turned on, the FSN Carousel provides an effective showcase for the range of services. As customers become more familiar with the services, however, they want a shortcut to select the desired service without going through a series of menus.

Video Games

The FSN provides two ways for the customer to play video games: *native* video games, which are integrated into the HCT, and *console* games, which are played on an external game console.

Native games include card games, Klondike and Gin, and two 3D action games (PODS and TVBots). The native games are networked so that the player enters into a contest with another FSN customer. The FSN provides no mechanism to select a specific opponent, using the first-come, first-served model. PODS is a 3D flying game in which the successful player wins by dragging and dropping worms onto squares to capture the most territory on a board. TVBots puts the player into a 3D game in competition with up to 15 other players. The game is everyone for him- or herself, and players are vanquished from the maze when they take a given number of hits from other players. (An interesting phenomenon occurred when a group of younger customers started playing as a team, effectively wiping out individual players by playing cooperatively.)

Console games were provided on the Atari Jaguar console. Each game was downloaded on request over the Fast IP channel into the memory of the Jaguar console. The Jaguar console was connected to the HCT by means of a high-speed parallel interface.

Home Shopping

The home shopping venue allows the customer to select from a number of different stores: Crate and Barrel, Chrysler, LL Bean, The Nature Company, Sharper Image, Spiegel, The U.S. Post Office, Viewers Edge, Warner Brothers Studio Store, Williams Sonoma, and others. Each store provides its own unique user interface, personalized by the applications developer. Most of the stores support catalog-style ordering of goods that are delivered to the customer. Payment is by credit card.

Online shopping is now a multibillion-dollar-a-year business over the World Wide Web. There is no reason why interactive television should not take a share of this market if these same services can be delivered via a set-top box.

Enhanced Television

The Full Service Network was originally conceived as a highly enhanced television service; all the interactive services used supporting video. This approach creates an entirely new kind of television service with its own requirements for production. (The FSN even had its own Digital Production Center, which was responsible for authoring the interactive *content*, or programming.)

To support a mass rollout of interactive services, interactive programming must be adopted by existing programmers. But this will not happen until there is a significant population of set-top boxes that can support this type of service. Therefore, there has been considerable interest in interactivity that uses existing programming so that the cost of production is not incurred twice.

The FSN demonstrated this approach of enhancing existing broadcast television by adding an interactive application that displayed sports scores as an overlay. This application enables the customer to watch NFL Sunday football and browse through the other sports scores without missing any part of the game.

Applications Model

The FSN applications model was based heavily on a distributed computing model. Each application is distributed between the HCT and the server complex and includes a client and set-top part that communicates via the slow IP network. A *Remote Procedure Call* (RPC) mechanism provides a convenient abstraction of the communications layer to the applications layer. The RPC mechanism allows a procedure in the client application to invoke a remote procedure in the server application in the same way as a local procedure, greatly simplifying the development of distributed applications.

FSN client applications rely on the considerable resources in the HCT to perform the lion's share of processing, and are considered to be *thick* clients. A thick client is responsible for

all the presentation layer functions and performs these functions with no assistance from the server. Several advantages are gained from the thick client model:

- The abstraction of the data is high. As an example, if a simple string of text is retrieved from the server and presented as an animated overlay by the client, the communications requirements are tiny compared to the server sending an animated overlay for display by the client. In other words, less communications resources are required, resulting in less demand on the network compared to a thin client model.
- Less processing resources are required by the server because it is primarily retrieving data objects. Moreover, a server can be designed to support many client instances without having to maintain a separate context for each client.

The combination of an RPC mechanism and a thick client allows scaling of the FSN applications. Nevertheless, the FSN applications model has its limitations because it is dependent on the availability of the network and the server components. For example, the FSN navigation application is implemented as a distributed application and is dependent on the availability of the slow IP network and the server resources. Thus, navigation of broadcast channels is tied to interactive resources, and a server failure affects some customers.

Application Requirements

The FSN applications described (navigation, games, and shopping) require significant network resources in each area identified in Chapter 8, "Interactive Services":

- Software download—All FSN application client software is downloaded in response to a user action—for example, selecting a channel or responding to an on-screen dialog. Thus, all FSN applications can be classified as solicited applications.
- Activation—All FSN applications are activated in response to a user action and can be classified as menu applications.
- Communications—All FSN applications require two-way communications with varying quality of service requirements. The navigation and home shopping services used best-effort quality of service in most cases, but when streaming media was used to support the application, guaranteed quality of service was required to deliver it. The game services required a bounded delay, especially for the action games, and this quality of service was achieved through careful management of the connectionless delivery network.
- Streaming media—All FSN services are independent of broadcast programming and are correctly classified as media-independent interactive services. However, this is misleading because many of the interactive services use on-demand streaming media to provide multimedia support (see Chapter 11).

Applications Resources

The FSN offers considerable resources to its distributed applications, both at the HCT and at the server. The network resources are engineered to support on-demand services to 25% of customers at any given time, and this dictates most of the 5 Gbps of forward network capacity. In comparison to this figure, the network resources allocated to interactive services are relatively small (about 250 Mbps or 5% of the total network capacity). This demonstrates the potential cost savings of an interactive service over an on-demand service.

Digital Set-Top

The central processing unit (CPU) performance of the first prototype FSN HCT was approximately 100 mips. This seemed like an unbelievable and inexhaustible resource in 1993 but soon proved not to be the case. Real-time compositing of live video and graphics in software puts a tremendous load on the CPU. At 60 fields per second, the CPU has just 16 milliseconds to render graphics before the field is displayed, and the field cannot be late. Even though video and audio decompression was done in hardware, a faster, 140 mips, version of the R4000 CPU was required to support FSN applications.

Software Download

Software download in the FSN allows remote upgrades to the operating system and applications. This is essential in an interactive system so that new applications can be downloaded to provide new services. However, the FSN downloads each application as a solicited application—that is, as a result of a user action. This model has some serious drawbacks when the size of the network is increased:

- There is a possibility that many users could request a new service at the same time; statistically, this is unlikely in an uncorrelated system. However, when a popular broadcast ends (or goes to commercial break), a significant fraction of customers might decide to switch to interactive services. Engineering the system for a maximum download time can be very expensive if peaks of activity are taken into account.
- Each time an application is loaded, it generates a transactional load on a server. Once again, server capacity must be engineered for peak activity.
- Any failure in any part of the network or server complex might cause an application download to fail. The navigator application is also used to select broadcast channels, and their availability can be compromised if the supporting interactive services are not completely reliable.

Although the FSN developed a broadcast carousel technique to reduce transactional load due to HCT boot, availability of all services was dependent on the two-way operation of the network and correct operation of the servers.

Communications

FSN applications required some communications resources that were not anticipated at the start of development:

- Lightweight remote procedure call (RPC)—See the section Applications Model, earlier in this chapter.
- Distributed naming service—After the client application is loaded, it needs to bind to a specific server resource in the headend. Use of a physical address does not allow server resources to be distributed across servers, so a logical address, or name, is used. At the server, a naming service is used to resolve the name, and the naming service must be distributed to provide high availability.

Lessons Learned

The Full Service Network provided some surprising and valuable lessons that can be applied to interactive service development.

Interactive services require a different communications model from on-demand services. Interactive services tend to use multiple sessions for short bursts of communication, so they require a connectionless signaling mechanism, such as that provided by IP networking. In contrast, on-demand services require a continuous stream of data that is best provided by a connection-oriented network, such as that provided by ATM. In particular, the TDMA reverse scheme was found to be very wasteful of return spectrum because it used a fixed bit rate allocation to each HCT, and the allocation is wasted most of the time. As a result, the DAVIC out-of-band protocol was developed to include a reservation protocol that allows many more set-tops to share a given return channel.

Application development cost is significant and must be carefully managed by the choice of development tools and execution environment. Application development tools must allow rapid prototyping and testing of the application in an environment that accurately simulates the cable network. The execution environment must provide some level of hardware independence so that applications can be ported from one set-top to another without the need for a complete rewrite. A robust execution environment is also very important so that applications cannot crash or permanently disable the set-top.

In addition, certification of interactive applications becomes critical to ensure that a new application does not disrupt other applications on the set-top or other functions on the network. In the FSN, a separate certification network was required to perform quality assurance of new applications before they were placed on the production network. Quality assurance of applications is key to ensuring reliable services in the cable environment.

Interactive television must have similar production values to the television programming that it accompanies. This means that applications must provide a rich graphical interface that is designed for a television display (not a computer display).

Pegasus

The Pegasus program is a direct descendant of the FSN. The program started in early 1995, even as the FSN was being rolled out to customers and the Orlando staff was grappling with stability and maintenance of the FSN. Originally, Pegasus was known internally as the Deployable FSN but this name was soon changed to Pegasus. The Pegasus name is derived from the Trojan horse strategy that it uses.

In 1995, the major stumbling block was the cost of an interactive set-top. Cost was not really an issue for the 4000 Orlando set-tops, but for general deployment of interactive services the cost of the set-top is critical. The Trojan horse strategy defines an advanced digital set-top that includes interactive features at a small cost premium over a basic digital set-top (see Chapter 6, "The Digital Set-Top Converter"). This set-top is universally deployed—initially to provide broadcast services but with future interactive services following soon after. This not only makes the set-top future-proof but also allows applications software and interactive services to be provided on the same platform. To make the interactive features cost-effective, they are included in every digital set-top; but, like the Greeks in the Trojan horse story, they appear only when interactive services are developed and delivered by the cable operator.

In much the same way, the Pegasus network architecture is designed to support interactive services as they are deployed. To minimize costs, the network communications paths are engineered to support interactive services at a relatively low service penetration, but are designed to be easily scaled up to handle additional signaling traffic as service penetration increases.

Pegasus Phase 1.1 Goals

The Pegasus Phase 1.1 goals were articulated in a Request for Quotation (RFQ) issued in 1997, one year after the initial Pegasus RFQ. Phase 1.1 adds to the interactive services provided by the resident navigation application. To enable new interactive services the RFQ identified two major goals:

"A true applications platform that will allow developers of new programs and services to use ubiquitous tools and, to the extent possible, ignore the details of the underlying broadband network architecture. This leaves the deployment of these programs and services as purely a business decision (much as deploying a new analog broadcast TV channel is today).

"Deploying a network and hardware platform architecture that will incent more vendors to enter the market in addition to the initial three mentioned above. In particular, it is hoped that the basic Pegasus hardware and network standards (with the exception of Conditional Access) will allow vendors of those products to build modulators, set-top terminals, etc., with ever better cost/performance benefits.

"The addition of a ubiquitously-supported, software runtime (execution) environment on top of, or as part of, the set-top native operating system is seen as the way to achieve both of the goals above. The execution environment and associated tools give software developers the ability to create new services and applications independent of the underlying hardware and, to some degree, the network architecture. At the same time, this abstraction allows set-top hardware vendors the flexibility to pick the best CPU's, memory

architectures, graphics capabilities, operating systems, etc., to improve the cost/performance equation while supporting all applications developed for the runtime environment."

The complete RFQ is available on the Web (see the section Internet Resources at the end of this chapter).

Network Overview

The Pegasus network is described in Chapter 7, "Digital Broadcast Case Studies." The important aspects of the network for the support of interactive services are the real-time, two-way network that links the Pegasus set-top to the headend equipment.

The two-way network is based on standard networking protocols and equipment and is engineered to support interactive services (albeit at relatively low service penetration). These standards are described in Chapter 5, "Adding Digital Television Services to Cable Systems."

Lessons learned from the FSN allow the Pegasus network to support all the same interactive services as the FSN at considerably reduced network cost. This is possible because

- DSM-CC data carousels are used wherever possible to reduce the transactional and network traffic required for downloading interactive applications.
- The DAVIC OOB protocol definition is designed to support many bursty traffic sources in the shared-media cable return environment. This reduces the number of QPSK demodulators per distribution hub by a factor of 15 compared to the FSN.
- The operating system software and navigation software are always resident in the Pegasus set-top. This greatly reduces the network resources required for software download.

In addition, network resources can be further reduced by careful application design. The design of these so-called *green* applications is discussed in the next sections.

Services

Pegasus is designed to support all the interactive services described in Chapter 8. These can be summarized as

- Navigation—Navigation services include analog and digital tuning, interactive program guide, IPPV interface, VCR remote control, parental control, subscriber preferences and configuration, digital music service, messaging, and EAS support.
- Information—Informational services provide news, weather, sports, stocks, and calendar-of-events information.
- Communications—Communications services include e-mail and chat.

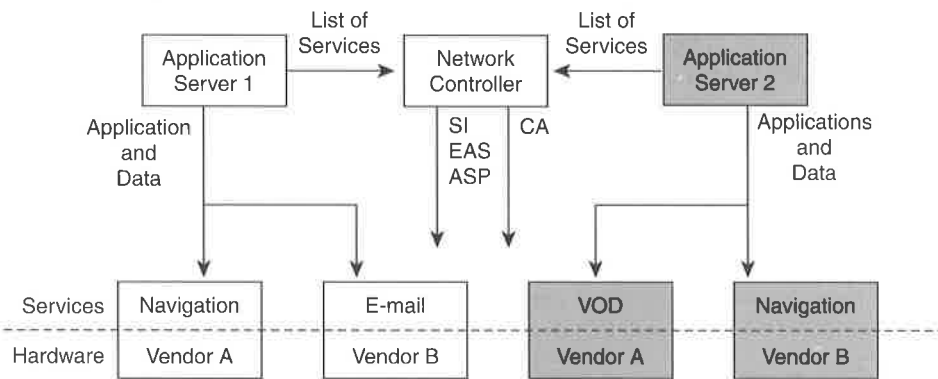
- Electronic commerce services—E-commerce services include home shopping and home banking (including the ability to withdraw electronic cash).
- Video games—Video games include single-player games, where the user plays against the computer, and networked games, where the user plays against other users over the network.
- Enhanced television—Enhanced television services include interactive commercials, play-along game shows, opinion polls, and surveys.

Applications Model

The Pegasus applications model supports both client and distributed applications. Client applications provide local interactivity, in which the customer interacts with the client application. This model is used for all navigation functions to reduce transactional load on the headend servers.

In addition, Pegasus separates the interactive applications functions from other system functions, allowing multivendor applications to be supported. This is illustrated by the example in Figure 9-5.

Figure 9-5 Pegasus Applications Model



In Figure 9-5, a single network controller is responsible for broadcast functions common to all customers. These services include the delivery of service information (SI), emergency alert system (EAS) messages, application service protocol (ASP), and conditional access (CA) system messages, in which

- Service information defines the MPEG source identifier mapping for broadcast services on the cable system according to SCTE (proposed) standard DVS-234.
- Emergency alert system messages are used to signal emergencies according to SCTE standard DVS-208 (EIA 814).

- All services are accessed by the Pegasus navigator using the application service protocol (SCTE proposed standard DVS-181). There is a conscious effort to abstract all services into a single channel space at the user interface level. The application service protocol (ASP) *advertises* services available on the cable system, effectively providing a naming service to allow the set-top to locate broadcast and interactive services. Interactive services are provided by the application servers, and each application server supplies a list of services to the network controller for insertion into the ASP carousel.
- Conditional access (CA) system messages deliver entitlement and de-scrambling keys according to the selected conditional access system (in this case, PowerKEY).

Application Portability

Application portability is designed to allow applications to run on multiple hardware platforms—in this case, digital set-tops. All digital set-tops are built to conform to the common network protocol specifications described in the section Applications Model, earlier in this chapter. Figure 9-5 shows two digital set-top implementations, from Vendor A and Vendor B, and two application servers, Application Server 1 and Application Server 2.

Each application server provides interactive services to either type of set-top by providing applications code and data. Applications code is continuously broadcast over a DSM-CC data carousel. Applications data is distributed over a DSM-CC data carousel or is accessed using the two-way IP network that connects the set-tops to the servers.

In the example shown in Figure 9-5, Vendor A set-tops have loaded a navigation application from Application Server 1 and a VOD application from Application Server 2. The VOD application is a special case of interactive application because it requires a media server and a connection management service to operate. (VOD is discussed in more detail in Chapter 10, "On-Demand Services.") Vendor B set-tops have loaded an email application from Application Server 2 and a navigation application from Application Server 2.

For this scenario to work, there must some mechanism to ensure that applications can execute on the two different vendor set-tops. There are three ways to achieve this goal, which defines the level of *application portability*:

- Each application is individually authored to each type of set-top—in other words, there is no application portability. This approach quickly becomes unmanageable as the number of set-top types and applications increases.
- Each application is authored to a common operating system. In this case, the operating system provides the portability layer, but the operating system has to be ported onto every set-top. In the PC world, the dominant operating system is Windows, and it provides a portability layer between different hardware vendors.

- Each application is authored to a portable *middleware* layer of software. To support different operating systems, only the middleware has to be ported to each operating system, which is much less effort than porting all applications. In the computer industry, the Web browser provides the middleware for application portability and enables Web applications to be independent of both the hardware platform and the operating system.

The Pegasus system uses the second and third options to promote portability. *Native* applications are authored to the PowerTV operating system. *Portable* applications are authored to a middleware software, which includes Hypertext Markup Language (HTML), JavaScript, and, optionally, Java components.

In Pegasus Phase 1.0, the only applications required are those that support navigation. These are implemented as native applications and can be ported to a new operating system if required. A wide range of interactive services requires a portable applications environment, and the Pegasus 1.1 RFQ was issued to select a middleware software to support such an environment.

Applications Resources

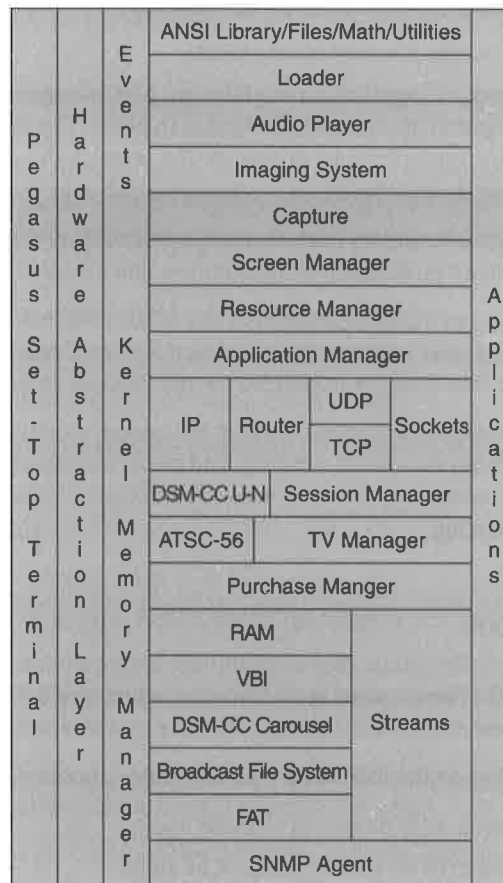
Pegasus applications require the same set of resources as the FSN applications; these resources are the set-top, network, and server resources.

Set-Top

The FSN experience taught us that many of the video, graphics, and communications functions should be implemented in hardware in the Pegasus set-top; because hardware state machines operate in parallel, they provide considerable performance improvement over a software implementation, which operates serially.

The Pegasus set-top relies on a sophisticated multimedia operating system to manage the basic set-top resources: CPU, graphics, memory, and communications. The operating system is responsible for delivering these resources to applications so that they can operate properly. It provides support for both broadcast and interactive applications, and through the PowerTV operating system application programming interface (API), provides access to the set-top terminal functions for other applications. The operating system components are shown in Figure 9-6.

Figure 9-6 PowerTV Operating System Components



The PowerTV operating system components are described in the following sections.

Operating System Kernel

At the heart of the operating system is a high-performance, multitasking kernel, which serves as the system administrator, and the set-top terminal Memory Resource Manager. For optimal real-time performance, the operating system kernel provides

- Priority-based, preemptive task scheduling, supporting multithreaded applications
- Unified events system
- Dynamic linking and dispatching of functions and modules

- Interprocess communication facilities
- Interrupt and exception handling services
- Efficient dynamic memory management

Because all system and application modules are position-independent, complete modules or single functions can be dynamically loaded in place. The operating system is fully reentrant, allowing multiple tasks to execute the same module, while each task uses separate stack and data spaces.

The kernel manages all memory resources on a dynamic basis with a flat memory model, using the hardware memory protection and management unit available in the Pegasus platform.

The operating system includes a subset of the ANSI C libraries that provide string manipulation, character testing, memory allocation and deallocation, and date and time functions. These functions are optimized for the Pegasus environment.

A loader module supports dynamic linking of resident modules, operating system extensions, applications, shared libraries, and resources into the system at runtime. The loader is stream-based, supporting dynamic relocation of system and user modules as well as forward referencing.

Audio Functions

Abstract devices called audio players provide multichannel pulse code modulation (PCM) audio playback for one or more applications, masking the differences of underlying hardware implementations and assuring the best possible audio playback.

Audio players support the following types of audio, stored in AIFF data format:

- Uncompressed 8-bit mono/stereo PCM audio
- Uncompressed 16-bit mono/stereo PCM audio
- International Multimedia Association (IMA) ADPCM compressed audio
- 8-bit μ law compressed audio

Audio players enable applications to

- Add and remove audio players.
- Play sounds, allowing the audio mixer to perform any necessary sample rate or bit depth conversions.
- Play any portion of a sound.
- Loop any portion of a sound any number of times.
- Pause and resume playback.
- Adjust the audio playback volume and stereo balance.
- Adjust playback pitch.

Five basic system sounds (ActionAccept, ActionReject, Attention, KeyAcknowledge, and KeyClick) are provided with the system.

Imaging Functions

The imaging system provides graphics rendering support by including a set of graphics primitives: lines, circles, ellipses, rectangles, round rectangles, arcs, and polygons. A set of graphics functions and effects enables rendering of complex images with extensive pixel map manipulations, including scaling, transparency, and translucency support. In addition to its support for pixel maps and graphics primitives, the imaging system supports

- Filling shapes with solid colors and patterns
- Drawing with brushes of application-defined widths and patterns
- Creating drawing patterns using brush masks

The imaging system supports drawing in the following color formats:

- 8-bit red green blue (RGB)
- 16-bit RGB
- 8-bit color lookup table (CLUT)
- 8-bit CLUT with an 8-bit alpha plane value

The imaging system supports opaque, transparent, and translucent graphics overlays on the video plane. In addition, the digital video can be captured in real-time with arbitrary vertical and horizontal scale factors and then manipulated as a standard pixel map.

Antiflutter and horizontal smoothing filters are controlled by the imaging system to provide high-quality graphics on interlaced television displays.

Supported fonts include monospaced fonts, proportional bitmap fonts, and their anti-aliased equivalents in a variety of styles. The system utilizes the NFNT font format definition, enabling applications to take advantage of a wide range of pre-existing font families as well as existing font creation tools on desktop platforms.

Screen Manager

The Screen Manager oversees the screen requirements of applications and controls which screen of the active application is currently visible. The Screen Manager provides support for applications using shared, private, and private-purgeable screens (that is, private screens created from purgeable memory). The use of private-purgeable screens allows the operating system to reclaim these memory resources when there is a critical memory requirement and the application that owns the purgeable screen is in the background.

Application Manager

Multiple applications must be able to execute concurrently on networked consumer devices. Although some subscribers are familiar with multiple windows-based applications within a PC environment, consumer devices using televisions as their display require a different viewer metaphor. The operating system is designed to allow the viewer to interact directly with one application at a time while all other applications continue to execute in the background.

In the operating system, applications use a messaging protocol that supports multiple applications sharing the same system. The Application Manager, which provides this protocol, is responsible for launching, releasing, switching, and monitoring applications within the system, as well as identifying and removing errant applications. The Application Manager makes resource switching optional at the application level. If an application enables resource switching, it can request that the Application Manager inform it when it regains focus and it can control which features—for example, channel or screen—of the application context are subject to resource switching.

Memory Resource Manager

The Memory Resource Manager keeps track of resources in memory and allows applications to acquire (that is, read) resources. A resource is data of any kind stored in the Apple-defined resource format in a module's resource partition. For example, the operating system has a built-in resource partition that includes system fonts and sounds.

Applications and system software interpret the data for a resource based on its resource type. Developers typically use resources to store descriptions of user interface elements, such as menus, fonts, sounds, and text. Because resources are separate from application code (which is usually stored in a module's data partition), developers can create and manage resources without recompiling their applications.

A developer can create resources using a resource compiler, resource editor, or other resource tools, such as ResEdit. Those resources are then put into a resource partition and the partition added to a module using a supplied packaging program. After the module is loaded into the set-top terminal, an application can ask the loader for the location of the resource partition and pass the handle returned to the Memory Resource Manager, thereby gaining access to the resource data.

The operating system provides an RFC-compliant Transmission Control Protocol/User Datagram Protocol-Internet Protocol (TCP/UDP-IP) network stack with a Berkeley Sockets application programming interface. The TCP/UDP-IP network stack provides transparent access to the RF broadband and Ethernet interfaces provided on the Pegasus set-top terminal.

Session Manager

The Session Manager is responsible for establishing, maintaining, monitoring, and tearing down network sessions between set-top terminal client applications and content-provider servers. It supports multiple, simultaneous sessions, administering the set-top terminal resources associated with each session.

The Session Manager supports the DAVIC 1.2 standard lower- and mid-layer protocol and physical interfaces, and implements the DSM-CC user-network signaling protocol.

In addition, the Session Manager supports client/server application download services, enabling applications and associated data to be downloaded on demand into the set-top terminal.

TV Manager

The TV Manager provides a logical channel-tuning interface to guarantee that an application is able to tune to a particular service, regardless of whether the channel is analog or digital. It provides a consistent programming interface to analog and digital services.

The TV Manager also interfaces with the PowerKEY conditional access system to guarantee that subscribers have access to only those channels for which they are authorized.

The TV Manager provides a logical interface for applications to the system information (SI) delivered across the network. The TV Manager controls the set-top terminal's master volume, stereo sound, and muting.

Purchase Manager

The Purchase Manager provides an interface for performing secure transactions for services such as pay-per-view and impulse pay-per-view. It simultaneously supports multiple purchase devices, internal as well as plug-in devices. Applications may retrieve a list of available events and request access to such events.

Streams Management

The operating system provides a streams component that provides an abstracted interface to data delivery sources in the set-top terminal. The streams interface provides access to streams whose origin is

- RAM, stored in the set-top terminal (can also be ROM or FLASH)
- VBI, delivered in the vertical blanking interval
- Data carousel, delivered in the digital domain using the DSM-CC data carousel standard both in-band on the carousel FAT channel and out-of-band on the forward data channel

For development purposes, the streams component also supports a SCSI interface, stored on a locally attached device such as a hard drive. PowerTV supports the FAT filing system standard for PCs.

In addition, a broadcast file system is supported, enabling applications to traverse a hierarchical broadcast delivery system with indications of file updates relayed automatically to applications.

By modifying one line of source code, applications can retrieve data from different sources. Thus, local storage media, such as hard drives, can be used for development; the application can then be moved to the real RF network with a change to only one line of code on the client.

The operating system also incorporates an SNMP agent that provides an interface to system diagnostic information that can be interrogated over the RF network.

The development environment for applications is Windows 95–based with C and C++ compiler support in addition to source-level debugging capabilities. Development set-top terminals are available that include Ethernet and SCSI interfaces with additional operating system support for FTP and FAT file systems to aid development.

Software Download

Pegasus supports all the download mechanisms described in Chapter 8:

- Data carousel—The data carousel is the preferred means of software download. The broadcast file system (BFS) layers a file system abstraction onto the DSM-CC data carousel to provide a hierarchical naming space for software files. (SCTE proposed standard DVS-111 describes the BFS.)
- MPEG-2 private data—A separate MPEG-2 packetized elementary stream (PES) is used to download synchronous applications.
- Out-of-band channel—Software download over the out-of-band channel is discouraged to preserve the out-of-band channel capacity for two-way communications.

The size of each of these download files becomes very critical, and general traffic analysis, which is not really currently done, becomes a new requirement in launching interactive applications, especially as more applications are launched on a single system.

Synchronization

The PowerTV operating system provides facilities for application synchronization to the MPEG-2 display timestamp by dispatching events to the application at the appropriate display time.

Communications

Pegasus provides one-way communications and two-way communications facilities (see Chapter 8, "Interactive Services"). One-way communications channels can support a limited subset of local interactive services in a one-way cable plant. (One-way operation might be dictated for business reasons in some areas or might be the result of a temporary operational problem; in either case, it is important to support as many services as is technically possible over a one-way system.) The DAVIC out-of-band channels provide two-way communications channels.

Lessons Learned

The Pegasus design, implementation and deployment experience has provided a number of valuable lessons:

- Broadcast services generate the primary source of revenue for cable systems and will continue to do so for the foreseeable future. This fact dictates that interactive services are layered onto a cable system with great care so as not to detract from the total value of services provided to the customer.
- Application development cost is significant and must be carefully managed by paying attention to application development tools and application portability. Application development tools must be able to simulate a cable system environment for ease of development and testing. Application portability enables multivendor set-top hardware and is best achieved through a middleware strategy.
- To be successful, interactive television services must incorporate the same production values as television programming. This represents a tremendous challenge to the application developer but can be mitigated by a rich suite of multimedia operating system facilities.
- Communications facilities must be carefully engineered to consider the most efficient way of delivering resources to the application. Total network traffic, transactional server capacity, availability, and reliability must all be considered when selecting the communications channel and data transfer mode.

Summary

This chapter summarized two interactive cable systems: the Time Warner Full Service Network and the Pegasus system. In each case, the interactive services, applications model, and applications resources are described.

The FSN provides a model for a fully interactive system and describes the ATM-based connection-oriented approach to network design. In addition, the connectionless signaling support required to support a distributed applications environment is described. FSN services and the applications model are summarized; this leads to a discussion of

applications resources and how these are delivered in the FSN. The lessons learned during the FSN development include the communications infrastructure required by interactive services and the application development issues surrounding a television-centric environment.

Pegasus builds on the lessons learned from the FSN development experience and focuses on cost-effective solutions for the delivery of interactive services; the connectionless network infrastructure is optimized for interactive applications, and applications are designed to make efficient use of network resources. The Pegasus applications model is layered onto broadcast services and supports client and distributed applications to support a wide range of services. The issue of application portability is addressed through the selection of a middleware strategy that supports maximum application portability.

References

Standards

SCTE DVS-111. "Revision 1, Digital Headend and Distribution CFI Response System Description (Digital Broadcast Delivery System)." November 24, 1997. Society of Cable Telecommunications Engineers.

SCTE DVS-181. "Revision 1, Service Protocol." February 4, 1999. Society of Cable Telecommunications Engineers.

SCTE DVS-208. "Revision 3, Proposed Draft Standard: Emergency Alert Message for Cable." April 19, 1999. Society of Cable Telecommunications Engineers.

SCTE DVS-234. "Service Information for Out-of-Band Digital Cable Television." June 23, 1999. Society of Cable Telecommunications Engineers.

Internet Resources

Pegasus requirements

<http://timewarner.com/rfp>

PowerTV

<http://www.powertv.com>

On-Demand Services

What are on-demand services? Generally, an on-demand service provides a dedicated multimedia stream to the customer and provides controls to manipulate the multimedia stream. The most common example provides a video and audio stream to the customer—video-on-demand (VOD). However, on-demand services are not limited to video-on-demand; they include any service that manipulates a dedicated multimedia stream.

Some examples of on-demand services are

- Movies-on-demand
- Music-on-demand
- Post-broadcast on-demand
- Special interest (or niche) programming
- Distance learning
- Library access
- Video mail

From a network developer's point of view, these services have the following in common:

- They require a session that provides some quality of service (QoS) guarantees to an individual client. This implies either a connection-oriented session establishment or a connectionless network that can guarantee QoS.
- They require some form of stream control. The client application controls the stream by signaling to the server application. For example, in a video-on-demand application, the client application sends play, pause, and fast-forward commands to the media server.

On-demand services represent the Holy Grail for the developer and the *killer* application for the marketer. On-demand services can be added to the real-time, two-way network infrastructure by adding new server applications and hardware.

On-demand services can support the dominant set-top application, which is watching television, and transform it by putting control into the customers' hands. Although interactive services based on computer technology might fail in the living room due to static or uninteresting appearance, on-demand services are a sure thing because the user-interface can use full-screen video and high-quality audio.

Video-on-demand has been in development for many years and has experienced various trials and tribulations (see Chapter 11, "On-Demand Cable System Case Studies"). Why has it taken video-on-demand so long to reach the market place?

There are a number of reasons:

- A video-on-demand service requires a tremendous increase in bandwidth from the cable system. For example, in a system serving 100,000 digital set-tops at 10% peak utilization, 10,000 discrete video streams must be simultaneously delivered over the cable system.
- Allowing the control of each video-on-demand session to be placed in the hands of the customer requires an efficient and reliable signaling mechanism for all 10,000 simultaneous video-on-demand sessions.
- The cost of the video servers, storage, modulation, and transport has to be reduced to the point where a business case can be made for video-on-demand services.

In a remarkably short period of time, all the technical and economic problems have been solved and video-on-demand is ready for prime time. In fact, multiple solutions exist for most problems, and this chapter describes and compares the alternatives for on-demand services.

Interactive Versus On-Demand Services

The distinction between interactive services (see Chapter 8, "Interactive Services") and on-demand services is the need for a session with some well-known QoS. Interactive services can be designed to reduce the network bandwidth requirements to a minimum (green applications) but there is no equivalent in on-demand services. In general, on-demand services have very different QoS requirements than interactive services. See Table 10-1.

Table 10-1 *Comparison of QoS Requirements for On-Demand and Interactive Services*

QoS	On-Demand	Interactive
Bit rate	2–6 Mbps (standard definition)	20–200 Kbps
Delay	Constant	Variable
Loss	10^{-10}	10^{-6}
Error	10^{-10}	10^{-6}
Retransmission	Not allowed	Allowed

Internet Comparisons

On-demand services are appearing on the World Wide Web and are called *streaming media* applications. Streaming media applications allow the user to listen to radio stations, preview CDs, and view movie clips. In contrast to the strict QoS requirements for on-demand services based on MPEG-2, streaming media applications are designed to work in an Internet environment where QoS guarantees are often nonexistent. This leads to some very different approaches for the delivery of video and audio (see Tables 10-2 and 10-3).

Table 10-2 *Comparison of Streaming Video and On-Demand Video Services*

Parameter	MPEG-2 Compression	Streaming Video
Bit rate	2–6 Mbps (standard definition)	20–200 Kbps
Resolution	480 x 544; 480 x 720	120 x 160
Frame rate	30 fps	5–15 fps

Table 10-2 compares streaming video and on-demand video services. MPEG-2 video requires a minimum bit rate of about 2 Mbps and is designed to guarantee a certain constant resolution and frame rate; for these reasons, its bandwidth is relatively constant. In contrast, streaming video services use compression schemes that adapt to network bandwidth by adjusting their frame rate and resolution dynamically. At higher bit rates (about 500 Kbps), streaming video compression can support 240 x 360 resolution. Moreover, they are designed to tolerate much higher loss and error rates than MPEG-2 compression.

Table 10-3 *Comparison of Streaming and On-Demand Audio Services*

Parameter	AC-3 Compression	Streaming Audio
Bit rate	192–384 Kbps	8–100 Kbps
Frequency response	20 Hz–20KHz	20 Hz–4, 8, or 20 KHz
Channels	2–6	1–2

Table 10-3 compares streaming audio and on-demand audio services. AC-3 video requires a bit rate from about 192 Kbps (for stereo) up to 384 Kbps (for 5.1 channel audio) and is designed to guarantee a certain level of audio fidelity at a fixed bit rate. Streaming audio services use compression schemes that adapt to network bandwidth by adjusting audio fidelity while tolerating much higher loss and error rates than AC-3 compression.