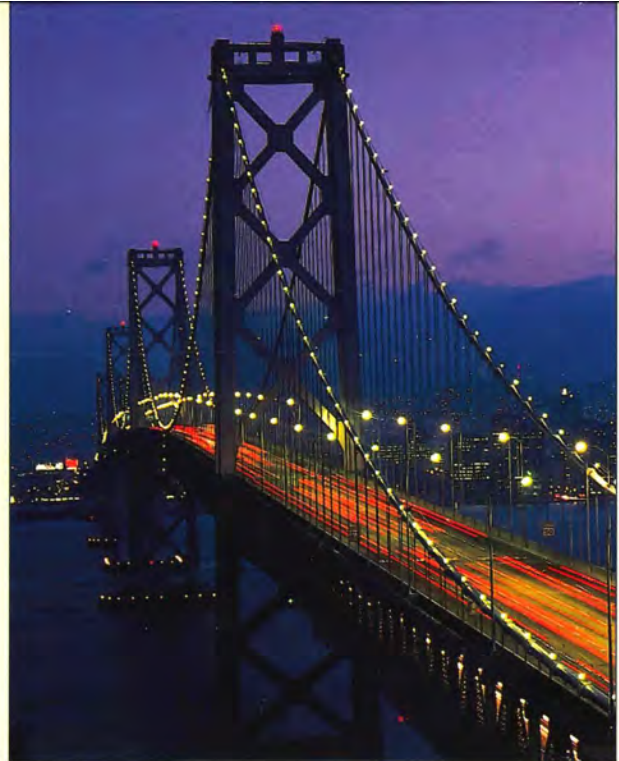


**M  
TP**  
MACMILLAN  
TECHNICAL  
PUBLISHING  
U.S.A.

MACMILLAN NETWORK  
ARCHITECTURE AND  
DEVELOPMENT SERIES



# SWITCHED, FAST, AND GIGABIT ETHERNET

THIRD EDITION

---

*Understanding, Building, and Managing  
High-Performance Ethernet Networks*

---

*Robert Breyer & Sean Riley*

MACMILLAN NETWORK ARCHITECTURE  
AND DEVELOPMENT SERIES

*Switched, Fast, and*  

---

*Gigabit Ethernet,*  

---

*Third Edition*  

---

Robert Breyer  
Sean Riley

**M**  
**TP**  
MACMILLAN  
TECHNICAL  
PUBLISHING  
U.S.A.

**Copyright © 1999 by Macmillan Technical Publishing**  
**THIRD EDITION**

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.

International Standard Book Number: 1-57870-073-6

Library of Congress Catalog Card Number: 98-85496

2002 01 00 99 4 3 2 1

Interpretation of the printing code: The rightmost double-digit number is the year of the book's printing; the rightmost single-digit, the number of the book's printing. For example, the printing code 99-1 shows that the first printing of the book occurred in 1999.

*Composed in Palatino and MCPdigital by Macmillan Computer Publishing*

*Printed in the United States of America*

**Trademark Acknowledgments**

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Macmillan Technical Publishing 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.

**Warning and Disclaimer**

This book is designed to provide information about Switched, Fast, and Gigabit Ethernet. 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 authors and Macmillan Technical Publishing 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.

**Feedback Information**

At Macmillan Technical Publishing, 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 at [ciscopress@mcp.com](mailto:ciscopress@mcp.com). Please make sure to include the book title and ISBN in your message.

We greatly appreciate your assistance.

**Publisher**

*Jim LeValley*

**Executive Editor**

*John Kane*

**Managing Editor**

*Caroline Roop*

**Acquisitions Editor**

*Brett Bartow*

**Development Editor**

*Kitty Wilson Jarrett*

**Project Editor**

*Theresa Mathias*

**Copy Editors**

*Hugh Vandivier*

*Keith Cline*

**Indexer**

*Tina Trettin*

**Acquisitions Coordinator**

*Amy Lewis*

**Manufacturing Coordinator**

*Brook Farling*

**Book Designer**

*Anne Jones*

**Cover Designer**

*Karen Ruggles*

**Proofreader**

*Megan Wade*

**Production**

*Steve Balle-Gifford*

*Darin L. Crone*



## About the Authors

**Robert Breyer** joined the networking industry in 1990, when he joined Intel in Folsom, California, as a product marketing engineer. His work there involved the sales and marketing of Ethernet 10BASE-T transceiver and controller chips. One of Robert's projects was the LAN on Motherboard program, which integrated Ethernet directly into PCs.

In 1993, Robert moved to Intel's network adapter group in Portland, Oregon, to develop the concept of value-added software for Intel's new line of network adapters. Subsequently, Robert was part of the team that launched the first 10/100 PCI adapter, as well as the first 10/100 intelligent server adapter.

Robert moved back to Johannesburg, South Africa, in 1995, where he became Intel's South African representative. During this time, Robert received the Intel Achievement Award, Intel's highest award. Robert was part of a team that set up an International Telecommunications Union (ITU) video-conferencing telemedicine demonstration featuring Dr. Andrew Grove, president and CEO of Intel, and Nelson Mandela, president of South Africa. From 1996 to 1998, Robert established and subsequently managed the official Intel sales subsidiary in South Africa. Robert is currently on a one-year leave-of-absence from Intel.

Robert has a bachelor's degree in electrical engineering from the University of the Witwatersrand in Johannesburg, South Africa. In addition, Robert holds a master's of science in electrical engineering and a master's of business administration from the University of California at Davis. He has written a number of white papers and application notes on local area networking technologies.

**Sean Riley**, currently the director of marketing for Intel's networking products in Europe, the Middle East, and Africa, began work in the networking industry in 1992. Beginning in Intel's Folsom, California, factory as a technical marketing engineer, Sean worked on several early versions of Ethernet silicon products. In 1993, Sean moved to Portland, Oregon, to develop and champion a new technology called Fast Ethernet.

During the next four years, Sean was involved early on in the High-Speed Working Group of the Institute of Electrical and Electronic Engineers (IEEE). This resulted in Sean becoming a founding member of the Fast Ethernet Alliance and a signing member of the IEEE 802.3u subcommittee, which standardized 100BASE-T Fast Ethernet.

Concurrently, Sean was a key member of Intel's Fast Ethernet product development team. For work in developing early Fast Ethernet products and technology, Sean won the Intel Achievement Award in 1995. In 1996, Sean led Intel into the Fast Ethernet hub and switch business.

Sean Riley has a bachelor's and master's of science in electrical engineering from Arizona State University. He has written a multitude of white papers and application notes on various networking technologies, including a chapter in the *Gigabit Ethernet Handbook*.

### About the Technical Reviewer

This book's technical reviewer, Matt Birkner, contributed his considerable practical, hands-on expertise to the development process for *Switched, Fast, and Gigabit Ethernet*. As the book was being written, he reviewed all the material for technical content, organization, and flow. His feedback was critical to ensuring that *Switched, Fast, and Gigabit Ethernet* fits our readers' need for the highest quality technical information.

**Matthew H. Birkner** has been working in the networking industry for seven years. He has been a network design engineer, network operations center engineer, and technical support engineer. Formerly senior networking engineer in MCI's outsourcing unit, he was responsible for the resolution of complex customer networking issues. He is currently employed as a network consulting engineer at Cisco Systems, where he works on enterprise network designs and performance/analysis. Matt holds a bachelor's of science from Tufts University, where he majored in electrical engineering, and he is a Cisco Certified Internetwork Expert (CCIE).

## Acknowledgments

When we got together to co-author the first edition of this book, back in 1994, we were breaking new ground. At that time, Ethernet switching and Fast Ethernet were both new technologies. There were a few books around that covered switching in general; and one or two that discussed the emerging 100Mbps version of Ethernet. Yet no single book really approached both subjects together. Many of the network professionals we knew were struggling with how to put both technologies into their networks simultaneously; hence the inspiration for the first edition of this book, which was published in 1995.

In 1996, ATM started getting a lot more press coverage, so we agreed to a second edition that included a chapter on ATM integration. It had been just over a year since the publication of the first edition, yet we were surprised by how many new Ethernet developments there had been in such a relatively short time period. The phenomenal success of Switched and Fast Ethernet meant that both technologies became mainstream virtually overnight, finding their way into large and small networks everywhere. Soon after the publication of the second edition, we noticed that the "all roads lead to ATM" hype was starting to subside. Instead, the debate shifted to the merits of Gigabit Ethernet and Layer 3 switching. We saw the same promise in them as we originally saw in Switched and Fast Ethernet. It also became clear to us that Gigabit Ethernet and Layer 3 switching went hand-in-hand.

This brings us to the third edition of this book. In June 1998, the IEEE ratified 802.3z, the Gigabit Ethernet standard, as well as the new 802.1p priority switching standard. The networking industry was off on yet another high-bandwidth race. When we first spoke to Macmillan Technical Publishing about the third edition, we vastly underestimated the number of new Ethernet developments. Gigabit Ethernet and Layer 3 switching are, in fact, just two out of about a dozen recent Ethernet innovations.

Our original goal was to have this book in print at around the same time as the Gigabit Ethernet standard would be approved. That didn't happen because it took us a lot longer than anticipated to research all the new topics and to separate the facts from the hype. Finally, you have the finished product in your hands: A practical guide to deploying a Switched, Fast, and Gigabit Ethernet LAN, complete with an in-depth discussion of Layer 3 switching. Along the way, we were able to incorporate all the latest IEEE Ethernet, VLAN, trunking, priority switching, and 1000BASE-T standards. We were also able to include the latest hot-off-the-press EIA/TIA cabling and associated tester developments.



Numerous individuals have helped make this new and improved third edition a reality. We would like to thank John Chambers, Doug Wills, and Maureen Kasper from Cisco Systems for contributing the Foreword to this book. We are also grateful to Kevin Tolly and Charlie Bruno from the Tolly Group for their interesting perspectives on Ethernet's future, covered in the Epilogue. Many thanks go to our technical reviewer, Matt Birkner, who taught us a thing or two about the IETF, bridging, and routing. We also want to thank Tony Beam from AMP, who provided some valuable input for Chapter 6, "Cabling and More on Physical Layers," and Sven Iversen who lent us his considerable knowledge of network management.

We would also like to thank Kitty Jarrett and Amy Lewis from Macmillan Technical Publishing, who literally made the third edition happen. Finally, we would like to thank Brett Bartow at Macmillan for backing this project 100% (even though his authors resembled slugs at times).

We wish you luck with your new Switched, Fast, and Gigabit Ethernet networks. We hope you learn as much from reading this book as we did from writing it. Feel free to email us with comments, especially fourth edition suggestions!

—Sean Riley (Sean.Riley@intel.com)

Robert Breyer (Robert\_A\_Breyer@yahoo.com)

October 1998



## Contents at a Glance

Foreword	xv
Introduction	xix
<b>Part I: The History of Ethernet</b>	<b>1</b>
1 The History of Ethernet	3
<b>Part II: Repeating, Bridging, Switching, and Layer 3 Switching</b>	<b>37</b>
2 An Overview of High-Speed LAN Technologies	39
3 Ethernet, Fast Ethernet, and Gigabit Ethernet Standards	83
4 Layer 2 Ethernet Switching	135
5 VLANs and Layer 3 Switching	161
<b>Part III: Determining Your Ethernet Needs</b>	<b>207</b>
6 Cabling and More on Physical Layers	209
7 Bandwidth: How Much Is Enough?	277
8 Network Components	317
<b>Part IV: Deploying, Managing, and Troubleshooting</b>	<b>381</b>
9 Deployment	383
10 Deployment Examples	445
11 Managing Switched, Fast, and Gigabit Ethernet Networks	469
12 Troubleshooting	511
Epilogue	545
<b>Part V: Appendixes</b>	<b>551</b>
A Gigabit and Fast Ethernet Vendors	553
B Further Reading	557
C Useful Web Links	567
Index	577

## Table of Contents

xv	Introduction	xix
xix		
1	<b>Part I: The History of Ethernet</b>	<b>1</b>
3	<b>1 The History of Ethernet</b>	<b>3</b>
37	The Origins of Ethernet: The ALOHA Radio System (1968–1972) .....	3
39	Building the First Ethernet at Xerox PARC (1972–1977) .....	5
83	DEC, Intel, and Xerox Standardize Ethernet (1979–1983) .....	7
135	3Com Productizes Ethernet (1980–1982) .....	8
161	StarLAN: A Great Idea, Except for Its Speed (1984–1987) .....	12
207	The Demise of StarLAN .....	13
209	The History of 10BASE-T and Structured Wiring (1986–1990) .....	14
277	Fiber Ethernet and UTP Ethernet .....	14
317	Structured Wiring: StarLAN and Token Ring .....	14
381	SynOptics Communications Is Founded .....	15
383	10BASE-T Is Approved as an IEEE Standard .....	15
445	Novell NetWare: The Networked “Killer Application” .....	16
469	The Demise of Token Ring .....	17
511	Ethernet Switching and Full-Duplex Emerge (1990–1994) .....	18
545	Full-Duplex Ethernet .....	19
551	Fast Ethernet Emerges (1992–1995) .....	19
553	The IEEE 802.3 100Mbps Standards Wars .....	20
557	Taking Ethernet into the Next Century: VLANs and Layer 3 Switching	
567	(1996–1998) .....	21
577	VLANs .....	23
	Layer 3 Switching .....	23
	Industry Merger and Acquisition Trends (1993–1998) .....	26
	3Com .....	27
	Bay Networks .....	27
	Cisco Systems .....	27
	Cabletron .....	28
	The Giants of the Networking Industry .....	28
	Gigabit Ethernet (1995–1998) .....	30
	Summary .....	36
	<b>Part II: Repeating, Bridging, Switching, and Layer 3 Switching</b>	<b>37</b>
	<b>2 An Overview of High-Speed LAN Technologies</b>	<b>39</b>
	A Historical and Somewhat-Biased Perspective .....	40
	Migration Issues .....	41
	Connection Costs .....	42
	Performance Considerations .....	42
	Quality of Service .....	44
	Ease of Migration .....	45

Understanding the Technology .....	46
Scalability .....	47
Multivendor Support .....	47
Standards .....	48
An Ethernet Primer .....	48
Revolutionary Versus Evolutionary Technologies .....	51
The Most Common High-Speed Technologies Compared .....	52
FDDI .....	52
TCNS and 100VG-AnyLAN .....	55
Iso-Ethernet .....	57
Fibre Channel and HIPPI .....	58
ATM .....	59
High-Speed Token Ring Options .....	74
Summary .....	77
<b>3 Ethernet, Fast Ethernet, and Gigabit Ethernet Standards</b> .....	<b>83</b>
Ethernet and the OSI Reference Model .....	84
Layer 1: The Physical Layer .....	86
Layer 2: The Data Link Layer .....	86
Layer 3: The Network Layer .....	87
10Mbps Ethernet MAC and PHY Standards .....	88
The Ethernet CSMA/CD MAC .....	88
Ethernet PHYs .....	93
Half- and Full-Duplex Ethernet .....	100
100BASE-T/Fast Ethernet .....	103
The Fast Ethernet CSMA/CD MAC .....	104
Fast Ethernet PHYs .....	105
100BASE-FX: Fast Ethernet for Fiber-Optic Cabling .....	105
100BASE-TX: Fast Ethernet for Category 5 UTP .....	106
100BASE-T4: Fast Ethernet for Four-Pair Category 3 UTP .....	106
100BASE-T2: Fast Ethernet for Two-Pair Category 3 UTP .....	107
Fast Ethernet MII .....	108
Gigabit Ethernet .....	110
The Gigabit Ethernet MAC .....	111
Gigabit Ethernet PHYs .....	113
The IEEE Auto-Negotiation Standard .....	117
Ethernet Repeaters .....	120
How Repeaters Work .....	121
The Repeater Collision Domain .....	122
10Mbps Ethernet Repeater Rules .....	124
100BASE-T Repeater Rules .....	126
Gigabit Ethernet Repeater Rules .....	131
Gigabit Ethernet Full-Duplex Repeaters .....	132
Summary .....	134



.....46	Ethernet Repeaters .....	120
.....47	How Repeaters Work .....	121
.....47	The Repeater Collision Domain .....	122
.....48	10Mbps Ethernet Repeater Rules .....	124
.....48	100BASE-T Repeater Rules .....	126
.....51	Gigabit Ethernet Repeater Rules .....	131
.....52	Gigabit Ethernet Full-Duplex Repeaters .....	132
.....52	Summary .....	134
.....55	<b>4 Layer 2 Ethernet Switching .....</b>	<b>135</b>
.....57	Ethernet Bridging and Switching .....	136
.....58	Repeaters .....	136
.....59	Bridges .....	137
.....74	Routers .....	141
.....77	The 802.1D Spanning-Tree Algorithm .....	143
83	Differences Between Bridges and Switches .....	144
.....84	Full-Duplex Ethernet .....	145
.....86	The New IEEE 802.3x Full-Duplex/Flow-Control Standards .....	146
.....86	Switch Architectures and Performance .....	151
.....87	Link Aggregation .....	154
.....88	Summary .....	157
.....88	<b>5 VLANs and Layer 3 Switching .....</b>	<b>161</b>
.....93	Unicast, Multicast, and Broadcast Frames .....	162
.....100	Unicast Frames .....	163
.....103	Multicast Frames .....	163
.....104	Broadcast Frames .....	163
.....105	Broadcast Storms .....	164
.....105	Broadcast Domains .....	165
.....106	VLANs .....	168
.....106	Benefits of VLANs .....	169
.....107	VLAN Membership .....	171
.....108	The Evolution of VLANs and VLAN Membership .....	178
.....110	Inter-VLAN Communication .....	181
.....111	Centralized VLAN Routing .....	181
.....113	802.1p Priority Switching .....	184
.....117	So Why Haven't VLANs Taken Over the World? .....	186
.....120	Layer 3 Switching .....	187
.....121	Layer 3 Background .....	188
.....122	How Routers Work .....	191
.....124	What Exactly Is a Layer 3 Switch? .....	193
.....126	Market Trends .....	196
.....131	A History of Layer 3 Switches .....	198
.....132		
.....134		



Layer 4 Switching .....	202
Port-Based Prioritization .....	203
Layer 4 Applications .....	204
Layer 4 Switching Issues .....	205
Summary .....	205
<b>Part III: Determining Your Ethernet Needs</b> .....	<b>207</b>
<b>6 Cabling and More on Physical Layers</b> .....	<b>209</b>
Encoding and Compression Techniques .....	210
Number of Pairs .....	210
Dual-Duplex Transmission .....	211
Bandwidth-Efficient Encoding .....	211
Cabling Standards .....	220
Some Cabling Standards History .....	220
TIA/EIA, ISO/IEC, and DIN .....	221
The TIA/EIA 568 Specifications .....	222
Ethernet MDI/MDI-X and Crossover Cables .....	230
Fiber Optics .....	231
Cable Lengths .....	237
IEEE802.3/Ethernet Cabling Standards .....	240
Fast Ethernet Physical Layers .....	240
Gigabit Ethernet Physical Layers .....	245
Fiber to the Desktop: Is It a Reality? .....	254
Bandwidth .....	255
Length .....	255
Cabling Cost .....	255
Hardware Availability and Pricing .....	256
EMI Immunity .....	257
Wiring Closet Considerations and the Connector Choice .....	257
Future-Proofing .....	258
FTTD Conclusions .....	258
Running Fast or Gigabit Ethernet on Your Existing UTP Cable Plant .....	259
Category 3 or 4 Cabling .....	259
IBM Type 1 Cabling .....	260
Category 5 Cabling .....	260
Cross-Connect Devices .....	261
Possible Category 5/1000BASE-T Problem Areas .....	263
Definitions of Some Important Cabling Test Variables .....	263
dB (Decibel) .....	263
Attenuation .....	264
NEXT, FEXT, Power Sum Measurements, and ELFEXT .....	264
Return Loss .....	266
Propagation Delay .....	266

.....202	Delay Skew .....	267
.....203	Attenuation to Crosstalk Ratio .....	267
.....204	Test Equipment .....	269
.....205	UTP Testers .....	270
.....205	1000BASE-T Cable Testers .....	270
	Fiber Optic Cable Testers .....	273
<b>207</b>	Some General Tips for Dealing with Cabling Suppliers .....	274
<b>209</b>	Summary .....	276
.....210	<b>7 Bandwidth: How Much Is Enough?</b> .....	277
.....210	Are All of Today's Ethernets Overcrowded? .....	278
.....211	Has the Network Infrastructure Become the Weakest Link? .....	279
.....211	How to Tell if Your Network Is Overloaded .....	279
.....220	The Customer Way: Unhappy Users .....	280
.....220	The Technical Way: Measuring Utilization and Throughput .....	280
.....221	The Bandwidth Limits of Ethernet .....	281
.....222	Definitions and Terminology .....	281
.....230	What Exactly Is the Capacity of Ethernet? .....	285
.....231	Improving the Ethernet MAC .....	289
.....237	Ethernet Capacity Recap .....	291
.....240	Tools to Measure Loading .....	293
.....240	Traffic Monitoring Tools and Protocol Analyzers .....	294
.....245	Hardware-Assisted Protocol Analyzers .....	295
.....254	Network Management Software .....	295
.....255	Handheld Diagnostic Tools .....	296
.....255	Upgrading Your Network .....	298
.....255	Which Nodes Are Generating the Most Traffic? .....	298
.....256	What to Do When Your Network Is Overloaded .....	300
.....257	How to Upgrade Your Network .....	302
.....257	Why Today's Networks Are Running Out of Bandwidth .....	303
.....258	Moore's Law .....	303
.....258	Practical Reasons Why You Are Running Out of Bandwidth .....	304
ant.....259	Gigabit Ethernet to the Desktop .....	314
.....259	Summary .....	315
.....260	<b>8 Network Components</b> .....	317
.....260	NICs .....	319
.....261	Desktop NICs .....	319
.....263	Mobile NICs .....	329
.....263	Server NICs .....	331
.....263	NIC Wrap-Up .....	342
.....264	Workgroup Components .....	342
.....264	Standalone Repeaters .....	343
.....266	Stackable Hubs .....	347
.....266	Gigabit Buffered Repeaters .....	351

Workgroup Switches .....	351
Desktop Switches .....	364
Backbone Components .....	366
Bridges .....	366
Routers .....	367
Backbone Switches .....	372
Summary .....	379
<b>Part IV: Deploying, Managing, and Troubleshooting</b>	<b>381</b>
<b>9 Deployment</b>	<b>383</b>
General Rules of 10BASE-T and 100BASE-T .....	385
The 100m and 5m Rules .....	385
10BASE-T Shared Network Rules .....	386
10BASE-T Switched Network Rules .....	386
100BASE-T Shared Network Rules .....	387
100BASE-T Switched Network Rules .....	388
100BASE-FX Fiber Network Rules .....	388
Gigabit Ethernet Network Rules .....	389
The Golden Rules of Switched, Fast, and Gigabit Ethernet .....	390
Deployment Step 1: Adding 10Mbps Switches .....	391
Standard Workgroup Deployment .....	392
Using a Switch of Hubs .....	394
Managing Switches .....	397
Knowing When to Upgrade .....	397
Deployment Step 2: Adding Fast Ethernet NICs .....	398
10/100Mbps Future Proofing .....	400
Using 100BASE-TX Versus 100BASE-T4 .....	401
Upgrading Servers to Fast Ethernet .....	402
The Patch Panel Approach .....	403
Deployment Step 3: Converting Workgroups to Fast Ethernet .....	404
Existing Workgroups with No Local Servers .....	404
Existing Workgroups with Local Servers .....	406
New Workgroups with No Local Servers .....	408
New Power Workgroups with Local Servers .....	410
Future Workgroup Considerations .....	412
Workgroup Options .....	413
Deployment Step 4: Switched and Fast Ethernet in the Backbone .....	415
The Location of Servers in a Backbone .....	415
Distributed Backbones .....	417
Collapsed Backbones .....	419
Server Farms .....	423
Managing Backbone Switches .....	425
Backbone Options .....	425



.....351	Deployment Step 5: Routers, Layer 3 Switching, and VLANs	
.....364	in the Backbone.....	427
.....366	Fast Ethernet Routers.....	427
.....366	Layer 3 Switches in the Backbone.....	429
.....367	VLANs.....	435
.....372	Deployment Step 6: Gigabit Backbone Switches.....	438
.....379	Connecting Multiple Workgroups.....	439
	Scaling Gigabit Ethernet Switches.....	440
	Connecting Gigabit Ethernet Switches to	
	Gigabit Ethernet Servers.....	440
<b>381</b>	Summary.....	442
383	<b>10 Deployment Examples</b>	<b>445</b>
.....385	Example 1: Deploying Switched and Fast Ethernet in a New	
.....386	Network.....	446
.....387	The Backbone Solution: A 100BASE-FX Switched Backbone.....	447
.....388	The Workgroup Solution: A Combination of Switched	
.....388	and Fast Ethernet.....	449
.....389	Cost Analysis.....	454
.....390	Example 2: Adding Layer 3 Switching Capability to the Backbone.....	456
.....391	Example 3: Deploying Switched and Fast Ethernet in a LAN	
.....392	with an Existing FDDI Backbone.....	458
.....394	Adding 100BASE-T Workgroups to an FDDI Ring.....	459
.....397	Extending an FDDI Ring with 100BASE-FX.....	460
.....397	Example 4: Deploying a Gigabit Ethernet Switched Backbone.....	461
.....398	Example 5: Deploying a 10-Mbps Switch as a Switch of Hubs.....	463
.....400	Example 6: Deploying Fast Ethernet in a Branch Office.....	465
.....401	In the Real World, Every Network Is Different.....	467
.....402	<b>11 Managing Switched, Fast, and Gigabit Ethernet Networks</b>	<b>469</b>
.....403	Network Management.....	470
.....404	SNMP.....	471
.....404	RMON.....	475
.....406	Other Aspects of Network Management.....	487
.....408	Desktop Management.....	490
.....410	Asset and Applications Management.....	491
.....412	Server Monitor Services Management.....	491
.....413	Help Desk Services Management.....	492
.....415	Business Management.....	492
.....415	Chargeback.....	492
.....417	Cost Management.....	493
.....419	Key Issues for Managing Switched Networks.....	493
.....423	VLANs.....	494
.....425	Managing Service Levels.....	500
.....425	Managing Switches at Gigabit Speeds.....	503
	Planning for Future Needs.....	505
	The Future of Managing Switched Networks.....	508



<b>12 Troubleshooting</b>	<b>511</b>
Dedicated Troubleshooting Products .....	512
Handheld Cable Testers .....	512
Handheld Frame Analyzers .....	512
LAN Analyzers .....	515
Traffic Generators .....	515
Dedicated RMON Probes .....	516
PC-Based Troubleshooting Products .....	517
Software-Based Sniffers .....	517
LAN Management Suites .....	518
Summary of Troubleshooting Product Categories .....	518
Troubleshooting Techniques in a Shared Environment .....	520
Intermittent Connections .....	520
Traffic-Level Analysis .....	520
Individual Frame Analysis .....	523
Troubleshooting Switched Networks .....	528
Traffic-Level Analysis in a Switched Network .....	528
Individual Frame Analysis .....	531
Troubleshooting Switch Configurations .....	533
Troubleshooting Fast Ethernet Networks .....	536
Changing Diagnostic Hardware .....	536
Changing Diagnostic Software .....	537
Potential Problems with Fast Ethernet .....	538
Troubleshooting Gigabit Ethernet Networks .....	539
Gigabit Ethernet Analyzers .....	539
Gigabit Ethernet Internal Switch Management .....	542
Summary .....	544
<b>Epilogue Fast-Track Future</b>	<b>545</b>
Can Applications Respond to the Challenge? .....	546
How Far Can Ethernet Evolve? .....	547
Will Ethernet Fall into the ATM Trap? .....	549
Evolution in Support of Revolution .....	549
<b>Part V: Appendixes</b>	<b>551</b>
<b>A Gigabit and Fast Ethernet Vendors</b>	<b>553</b>
<b>B Further Reading</b>	<b>557</b>
<b>C Useful Web Links</b>	<b>567</b>
Vendor and Organization Web Sites by Chapter .....	568
Interesting Publications .....	576
Newsgroups .....	576
<b>Index</b>	<b>577</b>

## 10Mbps Ethernet MAC and PHY Standards

Ethernet is strictly based on a layered OSI model. As a result, you can easily combine the Ethernet MAC with different PHYs. This section discusses the 10Mbps Ethernet MAC standards, and the four major baseband PHY specifications.

### The Ethernet CSMA/CD MAC

The Ethernet MAC technology is called *carrier-sense multiple access with collision detection* (CSMA/CD). Figure 3.2 illustrates the CSMA/CD flow.

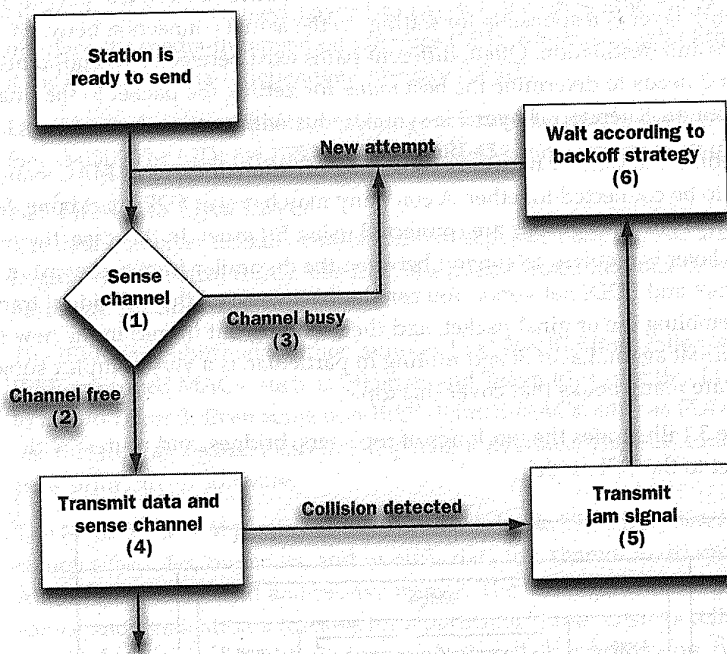


FIGURE 3.2 A flow diagram illustrating the CSMA/CD MAC.

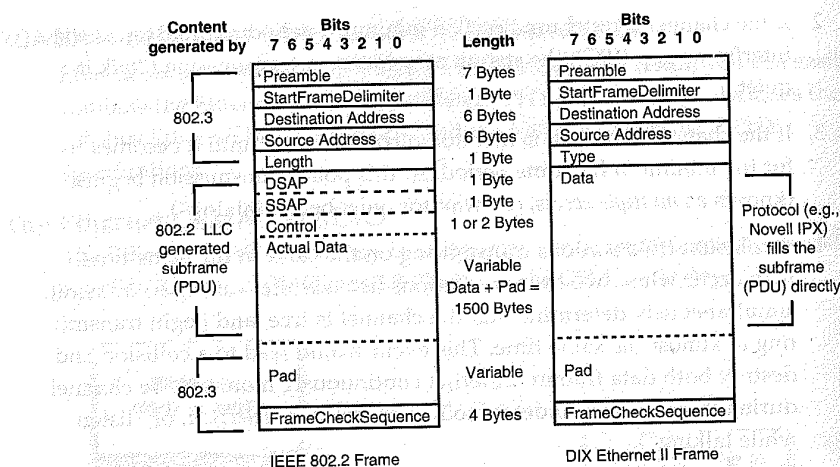
CSMA/CD is the shorthand version for about seven different steps that make up an Ethernet transmission. Notice the analogy to human speech among multiple individuals:

1. A station wanting to transmit a frame of information has to ensure that no other nodes or stations are currently using the shared media, so the station listens to the channel first. (This is the *carrier sense* part, also known as "listen before talking.")

2. If the channel is quiet for a certain minimum period of time, called the interframe gap (IFG), the station may initiate a transmission ("talk if quiet").
3. If the channel is busy, it is monitored continuously until it becomes free for the minimum IFG time period. At this point, transmission begins (known as *multiple access*, or "wait for quiet before talking").
4. A collision (two stations transmitting on the cable at the same time) may occur when two or more stations listen while waiting to transmit, simultaneously determine that the channel is free, and begin transmitting at almost the same time. This event would lead to a collision and destroy both data frames. Ethernet continuously monitors the channel during transmission to detect collisions (*collision detection*, or "listen while talking").
5. If a station detects a collision during transmission, that transmission immediately stops. A jam signal is sent to the channel to guarantee that all other stations detect the collision and reject any corrupted data frame they may have been receiving (also part of the collision detection, or "one talker at a time").
6. After a waiting period (called a *backoff*) the stations that wish to transmit attempt to make a new transmission. A special random backoff algorithm (called binary exponential backoff, or BEB) determines a delay time that the different stations will have to wait before attempting to send their data again. Of course, another collision could occur after the first one, especially when many nodes are trying to obtain access at the same time. After 16 consecutive collisions for a given transmission attempt, the packet will be dropped. This can and does happen if the Ethernet channel is overutilized. This is also part of the multiple access method.
7. The sequence returns to step 1.

Ethernet uses frames of data to transmit the actual information, also known as *payload*, from source to destination. Like most other LANs in existence today, Ethernet transmits a frame of variable length. The length of the frame changes because the payload or data field can vary.

The original Ethernet frame specified by Digital, Intel, and Xerox is known as the DEC-Intel-Xerox Ethernet V2.0 frame, or just DIX or Ethernet II frame. The official IEEE Ethernet frame subsequently replaced it. The only difference is in the 2-byte frame type/length field. Figure 3.3 shows both frame types.



**FIGURE 3.3** The official IEEE and the older DIX Ethernet II frame differ in some respects.

Let's look at the individual fields that comprise the Ethernet frame:

- The preamble is sent to allow the receiver to synchronize with the incoming transmission and locate the start of the frame. The preamble is a sequence of 01010101..., 7 bytes long.
- The start of frame delimiter (SFD) indicates that the MAC frame is about to commence. The SFD octet or byte is specified as 10101011.
- The source address denotes the sender. Each node has a unique address. The first three bytes of the address are called the block ID or organizationally unique identifier (OUI) and identify the manufacturer of the equipment. The IEEE assigns them. Intel, for example, is identified by the 00-AA-00 (hex) address, 3Com uses the 00-20-AF address, and Cisco uses 00-00-0C. The other three bytes are called the device ID and are assigned by each manufacturer. These are always unique. Three bytes of device ID allows for 16 million different and unique addresses:  $2^{24}-1$  or 16,777,215 to be exact. Some of the major Ethernet vendors have shipped more than 16 million Ethernet MAC devices and have started using new OUIs. Some of the references given in Appendix C, "Useful Web Links," show the complete list of vendor OUIs.
- The destination address specifies where the frame will be sent.



- The protocol type is a field in the original DIX V2.0 frame type. It specifies which kind of Layer 3 protocol the data contained. A hexadecimal value of 08 00 indicates a TCP/IP packet, for example, and 81 37 indicates a Novell NetWare packet.
- The newer IEEE frame differs from the Ethernet II or DIX frame type in one key area. The IEEE frame is much more popular these days and no longer uses the protocol type field, where it has been replaced with the length field. This specifies the total length of the data that will be transmitted, which can vary from 0 to 1500 bytes. If the contents of this field are 0 to 1500 bytes, you can be sure that it is the length we are discussing. If the contents of the type length/type field are greater than 1500 bytes (for example, 8,137), we are talking about the old DIX frame type.
- The data field can vary from 0 to 1500 bytes in length. The data field is also known as the protocol data unit (PDU). We discuss the contents of the data field in more detail later.
- If the actual data is less than a minimum length required, the MAC will add a variable pad to maintain a minimum total frame size of 64 bytes. If the data is longer than the maximum frame allows, Layer 3 will typically split the packet into more than one frame.
- Finally, a frame check sequence (FCS) ensures accurate transmission. The cyclical redundancy check method (CRC) checks for invalid frames. This value is calculated from the rest of the packet's data and sent along in the FCS frame. The receiving station performs the same calculations and compares its results with the FCS transmitted with the packet. If the results are different, the packet is rejected.
- The maximum total frame size is 1518 bytes. (The frame officially starts with the source address.)

There are a total of four different frame types. The DIX Ethernet II frame discussed previously, plus three different versions of the IEEE frame type: the Ethernet 802.3 frame (also known as Novell Raw LLC), the 802.2 frame, and the 802.2 with SNAP frame. All look slightly different. Figure 3.4 shows the contents of the 1500-byte data field for the four different frame types. The following list details the three different versions of the IEEE frame type:

- The Novell Raw frame format is used only by Novell's NetWare operating system, and uses the 802.3 frame shell type without adding an IEEE 802.2 LLC header within the data field. In this case, NetWare adds its

own IPX information. This type of frame is called "Novell Raw" because it encapsulates the IPX data in raw form without any 802.2 LLC information. Because the frame doesn't use the IEEE 802.2 LLC information, this frame type is actually Novell-proprietary and not 802.3 compliant.

- For the 802.2 frame type, the data field contains the 802.2 LLC-embedded information. The IEEE frame contains the protocol type information within the LLC subframe. Three fields are located at the beginning of the data field. The DSAP field, a 1-byte SSAP field, and a 1-byte Control field. The IEEE frame assigns Service Access Point numbers; among those currently defined are E0 for Novell, F0 for NetBIOS, 06 for TCP/IP, and AA for the Subnetwork Access Protocol (SNAP).
- The 802.2 with SNAP frame is almost identical to the 802.2 frame, except it supports more than 256 protocol types by adding a 5-byte Protocol Identification field. On any SNAP packet, both the DSAP and SSAP fields are set to AA, and the Control field is set to 03. The 5-byte protocol field then follows.

Table 3.2 lists all the relevant Ethernet MAC frame parameters. Most 10Mbps Ethernet/802.3 MAC parameters are listed in bit times. The Ethernet MAC is inherently scalable. All the parameters can be measured in terms of the time taken to transmit 1 bit of data, referred to as *bit times*. Note that the actual speed of Ethernet (10Mbps) is not mentioned in the MAC specification at all. This makes it very easy to run Ethernet at different speeds.

**TABLE 3.2 KEY 10MBPS ETHERNET MAC AND FRAME PARAMETERS, SPECIFIED IN BIT TIMES**

Parameter	Value (Bit Times)
Time	512 bit times
MinInterFrameGap	96 bit times (or 9.6 $\mu$ s)
AttemptLimit	16 (tries)
BackoffLimit	10 (exponent)
JamSize	32 bits
MaxFrameSize	12144 bits (1518 bytes)
MinFrameSize	512 bits (64 bytes)
AddressSize	48 bits

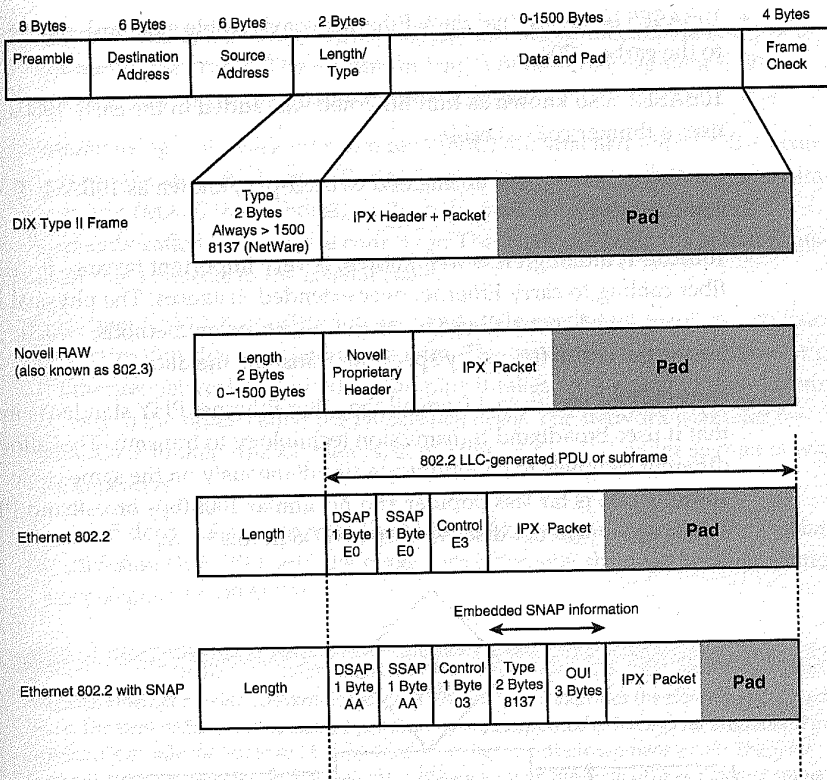


FIGURE 3.4 The data field of the four different frame Ethernet types in use.

Calculating the time to transmit 1 bit for 10Mbps Ethernet transmission becomes very easy:

$$1 \text{ bit-time} = \frac{1 \text{ bit}}{10\text{Mhz}} = 0.1\mu\text{s} \text{ or } 100\text{ns}$$

For 1Mbps Ethernet/StarLAN, the frame looks exactly the same. The only thing that changes is the bit time. For StarLAN, the bit time is  $1/1\text{MHz} = 1\mu\text{s}$ , or 1000ns. Fast Ethernet works exactly the same way: The frame is identical again, but the bit time is reduced to  $1/10$ , or 10ns.

### Ethernet PHYs

This section looks at the different PHY implementations for 10Mbps Ethernet (see Figure 3.5). There are officially five ways to transmit 10Mbps Ethernet:

- 10BASE5 is the original thick Ethernet coaxial cable standard, dating back to the early 1970s.
- 10BASE2, also known as thin Ethernet, was added in the early 1980s and uses a thinner coaxial cable.
- In 1990, Ethernet over unshielded twisted-pair, known as 10BASE-T, was standardized.
- 10BASE-F, although less well known, is very important because it utilizes fiber cabling to carry Ethernet over extended distances. The physical layers mentioned so far all use baseband transmission methods, which means the entire frequency spectrum transmits the data.
- 10BROAD36 is different from all the other Ethernet PHY standards in that it uses broadband transmission technology to transmit. This allows different channels to communicate simultaneously on the same cable. 10BROAD36 is far less popular and no similar 100Mbps broadband PHY exists, so we will not discuss 10BROAD36 in this book.

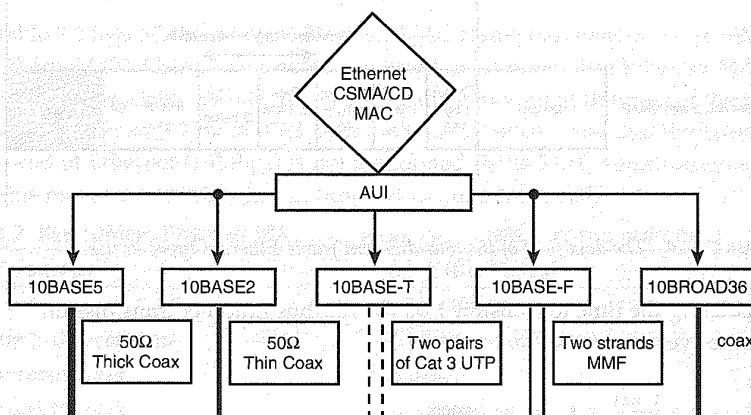


FIGURE 3.5 The different Ethernet/802.3 PHYs.

#### 10BASE5: Thicknet

10BASE5 is the original Ethernet 802.3 standard. 10BASE5 utilizes a thick, often yellow, coaxial cable that has a physical diameter of 10mm. The cable has to be terminated with a 50Ohm/1W resistor. Up to 100 stations per segment are allowed.



**Note**

*A two-user switched connection is a prerequisite to Full-Duplex Ethernet because full-duplex requires a point-to-point connection with only two stations present. Note that Switched Ethernet does not automatically imply full-duplex operation.*

Full-duplex has played an increasingly important role with the introduction of Fast and Gigabit Ethernet. In 1997, the 802.3X full-duplex/flow control standard was approved. (Chapter 4 covers this in more detail.) Some people in the IEEE even advocated making Gigabit Ethernet a full-duplex-only technology, which means dropping the all-too-familiar CSMA/CD MAC concept altogether. In the end, the half-duplex followers prevailed, and CSMA/CD operation was included in the standard. It is doubtful whether we will ever see any half-duplex shared Gigabit Ethernet equipment in the marketplace, however.

A trend is currently in place to move from repeated half-duplex to Switched full-duplex Ethernet. With the advent of Gigabit Ethernet in particular, it looks like CSMA/CD and collisions are heading for extinction. What will live on is the language of Ethernet, namely the 802.3 frame format. Our well-known Ethernet frames will be zooming along at 1, 10, 100, 1000, and someday 10000Mbps.

**100BASE-T/Fast Ethernet**

100BASE-T is the 100Mbps version of the classic Ethernet standard. The IEEE officially adopted the new IEEE 802.3u Fast Ethernet/100BASE-T specification in May 1995. The salient features of 100BASE-T are as follows:

- The 100BASE-T MAC uses the original Ethernet MAC operating at 10 times the speed.
- The 100BASE-T standard is designed to include multiple physical layers. Three different 100BASE-T physical layers are part of the 802.3u standard: two for UTP and one for multimode fiber. An additional UTP PHY was added later as the 802.3y specification.
- Like 10BASE-T and 10BASE-F, 100BASE-T requires a star-wired configuration with a central hub.
- 100BASE-T also includes a specification for an MII, a 100Mbps version of today's AUI. The MII layer is a digital interface connecting MAC and PHY and allows for external transceivers.

The differences between 10BASE-T and 100BASE-T are in the PHY standards and network design areas. That's because the new IEEE 802.3u 100BASE-T specification contains many new rules for repeaters and network topology. Figure 3.8 provides an overview of the new IEEE 802.3u standard.

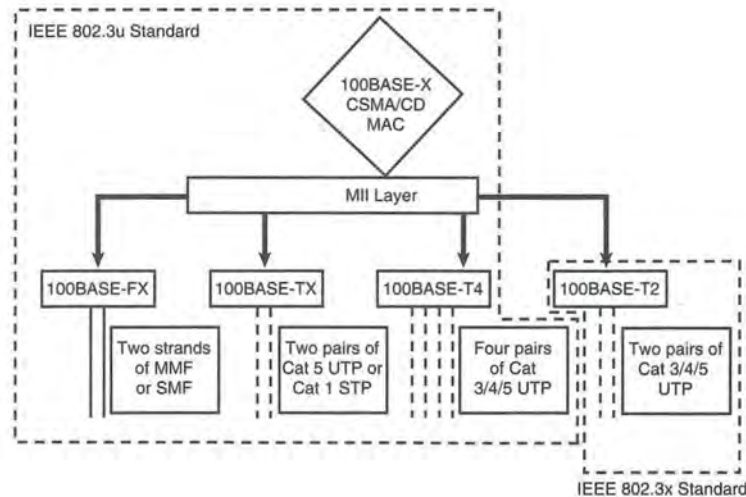


FIGURE 3.8 An overview of the 100BASE-T 802.3u standard showing MAC, MII, and the three original PHYs. (100BASE-T2 was added as 802.3y a year later.)

### The Fast Ethernet CSMA/CD MAC

The 100BASE-T MAC is almost identical to the 10Mbps “classic” Ethernet MAC. As mentioned earlier, the 802.3 CSMA/CD MAC is inherently scalable, which means that it can run at different speeds and be interfaced to different physical layers. StarLAN/1BASE5 took advantage of this scalability to run Ethernet at 1Mbps. Table 3.6 compares 10Mbps and the new 100Mbps Ethernet MAC standards. Note that the 100Mbps Ethernet MAC retains all the 10Mbps Ethernet MAC parameters except for *InterFrameGap*, which has been decreased to one-tenth its original value, from 9.6 $\mu$ s to 0.96 $\mu$ s. A Fast Ethernet frame has the same framing format as a 10Mbps Ethernet frame, except that it is transmitted across the wire at 10 times the speed.

TABLE 3.6 10Mbps COMPARED TO 100Mbps ETHERNET MAC PARAMETERS (MOST VALUES ARE IDENTICAL WHEN MEASURED IN BIT TIMES, BUT ONE-TENTH THE VALUE WHEN MEASURED IN REAL TIME)

Parameter	Ethernet/802.3	Fast Ethernet/802.3u
slotTime	512 bit times	Same <sup>1</sup>
Minimum InterFrameGap	96 bit times (9.6 $\mu$ s)	Same (=0.96 $\mu$ s)
attemptLimit	16 (tries)	Same
backoffLimit	10 (exponential number)	Same
jamSize	32 bits	Same

Parameter	Ethernet/802.3	Fast Ethernet/802.3u
maxFrameSize	1518 bytes	Same
MinFrameSize	64 bytes (512 bits)	Same
AddressSize	48 bits	Same

<sup>1</sup> Same means same value in bit times as 10 Ethernet (for example, the value in the column to the left).

You can calculate the 100BASE-T bit time as follows:

$$1 \text{ bit-time} = \frac{1 \text{ bit}}{100\text{Mhz}} = 10\text{ns}$$

### Fast Ethernet PHYs

As with 10BASE-T, 100BASE-T combines the CSMA/CD MAC with different physical layer specifications. The IEEE 802.3u specification contains three new physical layers for 100Mbps Ethernet:

- 100BASE-TX, which requires two pairs of Category 5 UTP or Type 1 STP cabling
- 100BASE-FX, which uses two strands of multimode fiber
- 100BASE-T4, which requires four pairs of Category 3 or better cabling

Then there is the IEEE 802.3y 100BASE-T2 PHY standard, which was completed about a year after the original Fast Ethernet standard. 100BASE-T2 utilizes only two pairs of Category 3 cable, but is virtually unheard of. 100BASE-TX and 100BASE-FX are the most popular 100BASE-T physical layers. They were based on the FDDI/CDDI physical layer, utilizing the 4B/5B encoding method. (Chapter 6, "Cabling and More Physical Layer Details," gives more information on the 4B/5B encoding method.) 100BASE-TX is the de facto standard for high-speed horizontal connections; 100BASE-FX is becoming increasingly popular for vertical/backbone connections. 100BASE-T4 is rarely used, whereas we know of no 100BASE-T2 products that are shipping.

Like 10BASE-T and 10BASE-F, all 100BASE-T PHY specifications require a star configuration with a central hub. All 100BASE-T UTP PHYs use the RJ-45 connector and specify a maximum 100-meter cable distance.

This section provides an overview of the different 100BASE-T physical layers.

### 100BASE-FX: Fast Ethernet for Fiber-Optic Cabling

100BASE-FX (like 100BASE-TX) borrows its physical layer from the ANSI X3T9.5 FDDI physical layer media-dependent (fiber PMD) standard. 100BASE-FX utilizes two strands of the multimode (62.5/125μm) fiber cabling made popular by FDDI. The maximum segment length for fiber-optic connections varies.



TABLE 3.7 A SUMMARY OF THE DIFFERENT 100BASE-T PHYSICAL LAYERS

PHY Parameter	10BASE-T	100BASE-TX	100BASE-FX
Origin of Technology		CDDI TP PMD (ANSI X3T9.5)	New; developed to 1996
IEEE standard	802.3i-1990	802.3u-1995	802.3u-1995
Encoding	Manchester	4B/5B	8B/6T
Cabling required	UTP Category 3/4/5	UTP Category 5 or STP Type	UTP Category 5 or STP Type
Signal frequency	20MHz	125MHz	25MHz
Number of pairs required	2	2	4
Number of transmit pairs	1	1	3
Distance	100m	100m	100m
Full-duplex capable?	Yes	Yes	No

<sup>1</sup> 150m for repeater-DTE, 412m for DTE-DTE, 2000m for full-duplex DTE-DTE, and 10km for single-mode full-duplex DTE-DTE.

Table 3.7 compares 10BASE-T with the four 100BASE-T physical layer specifications. We discuss the different physical layers and their cabling requirements in more detail in Chapter 6.

### Fast Ethernet MII

MI I is similar to AUI for 10Mbps Ethernet. The MII layer defines a standard electrical and mechanical interface between the 100BASE-T MAC and the various PHY layers. This standard interface works like AUI in the classic Ethernet world in that it allows manufacturers to build media- or wiring-independent products, with external MAUs being used to connect to the actual physical cabling.

The electrical signals differ between MII and AUI. AUI has a stronger signal, capable of driving 50-meter cable lengths; the MII signals are digital logic-type signals, capable of driving 0.5 meters of cable. MII uses a 40-pin connector, similar to the SCSI connector, although it is smaller. Figure 3.9 shows an external MII 100BASE-FX transceiver. We don't think you will see too many of these because most Fast Ethernet hubs and NICs already include the transceiver.

MI I never became as popular as AUI. AUI became a de facto standard because Thick Ethernet hardware wasn't available with an onboard transceiver. That only came later, with 10BASE2 or 10BASE-T. Yet with 10BASE2 and 10BASE-T, the AUI connector provided the capability for backward compatibility through an external transceiver. With 100BASE-T, most hardware manufacturers decided to skip MII and offer different products that integrate the physical layer transceiver already. NIC manufacturers, for example, often sell the same basic design in a 100BASE-TX and in a 100BASE-T4 version. You have to choose up front which media type you need. Think of it as a choice between Thinnet-only or 10BASE-T-only products, as opposed to combo or AUI-only cards using

external  
Pro/100  
flexibility

FIGURE 3.9  
Transceivers  
Transceiver.



TX	100BASE-T4	100BASE-FX	100BASE-T2
PMD (ANSI)	New; developed from 1995 to 1996	FDDI PMD (ANSI X3T9.5)	New; developed from 1994 to 1995
5	802.3u-1995	802.3u-1995	802.3y-1996
ory 5 or STP	8B/6T	4B/5B	PAM5x5
	UTP Category 3/4/5	Multimode or single-mode fiber	UTP Category 3/4/5
	25MHz	125MHz	25MHz
	4	2	2
	3	1	2
	100m	150/412/2000m <sup>1</sup>	100m
	No	Yes	Yes

single-mode

er specifica-  
irements in

standard elec-  
various  
rnet world  
t products,  
ing.

signal,  
ogic-type  
ector, simi-  
xternal MII  
se because

l because  
r. That  
BASE-T,  
through  
rs decided  
er trans-  
sic design  
front  
nly or  
ig

external transceivers. An example of a device with an MII connector is the Intel Pro/100 Server adapter, which uses an MII connector to provide you with the flexibility of either attaching an external fiber or a UTP PHY transceiver.



FIGURE 3.9 Fast Ethernet MII transceivers are similar to the old Ethernet AUI transceivers, but are very rare because most Fast Ethernet hardware includes an onboard transceiver. MII uses a 40-pin connector very similar to SCSI-2.

## Gigabit Ethernet

In 1995, Fast Ethernet seemed like such an ingenious, yet simple, idea. 100BASE-T became an overnight success, and it was only a matter of time before the frequency would be increased another order of magnitude. Work started on Gigabit shortly after the Fast Ethernet standard had been ratified. Three years later, on June 25, 1998, the IEEE 802.3z standard was officially adopted.

If you have read this chapter's information on Fast Ethernet, we should just ask you to go back and read everything with an extra zero added everywhere. Gigabit Ethernet is that simple. Some differences arise in the physical layers, network design, and minimum frame size that we cover here and in Chapter 6. For the most part, however, Gigabit Ethernet is just supercharged Fast Ethernet.

The IEEE 802.3z standard includes the Gigabit Ethernet MAC, as well as three physical layers that use the 8B/10B encoding originally developed as part of the ANSI Fibre Channel technology.

Essentially, the IEEE engineers bolted the existing Fibre Channel PHY to the Ethernet MAC running at 10 times the speed of the Fast Ethernet MAC. The 802.3z standard encompasses two fiber physical layer standards, 1000BASE-LX and 1000BASE-SX as well as one copper PHY, 1000BASE-CX. The copper-based PHY was also inherited from the Fibre Channel standard and was included to enable cost-effective and quick cross-connects. A fourth PHY, called 1000BASE-T, is still under development.

### Note

*The ANSI X3T11 committee is responsible for the Fibre Channel technology. ANSI decided to spell Fibre with an re rather than an er to differentiate it from the fiber-optic cable on which it runs.*

The IEEE has also defined a Gigabit MII (GMII), which is similar to the Fast Ethernet MII and connects Gigabit MAC and PHY. The GMII is only an electrical interface specification, and unlike the Fast Ethernet MII, the specification does not include a connector. The biggest benefit of the GMII is that it allows circuit designers to use existing Fibre Channel PHY chips and will allow for easy redesign of existing 1000BASE-LX equipment to accommodate future PHY transceivers.

Some people in the IEEE wanted to make Gigabit Ethernet a full-duplex-only technology; others wanted to preserve the classic CSMA/CD algorithm and the half-duplex shared operation. The reason for preserving the CSMA/CD part of

Ethernet was twofold. First, many vendors didn't want to redesign their Ethernet MAC chips completely. (Running it at 10 times the clock frequency didn't take a complete redesign.) Second, some IEEE members just wanted to preserve the 25-year heritage of CSMA/CD itself.

Figure 3.10 shows all the different Gigabit Ethernet components.

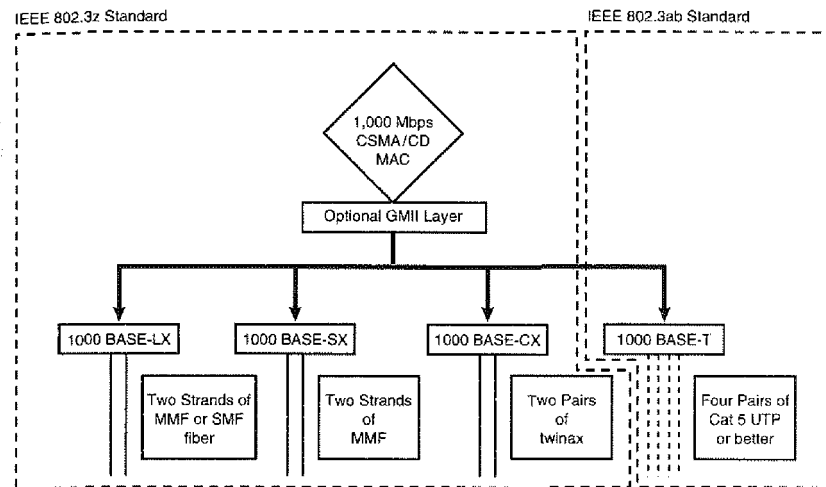


FIGURE 3.10 A Gigabit Ethernet block diagram showing the MAC, GMII, and four different physical layers.

100BASE-X refers to the original Fast Ethernet MAC and FX/TX PHYs. Similarly, 1000BASE-X refers to the 1000Mbps MAC and the LX/SX/CX transceiver technology.

### The Gigabit Ethernet MAC

The main points of the Gigabit Ethernet MAC are the following:

- Gigabit Ethernet uses the official 802.3 frame format, identical to that of 10Mbps and 100Mbps Ethernet (refer to Figure 3.3).
- Like 10Mbps and 100Mbps Ethernet, Gigabit Ethernet can operate in both half- and full-duplex mode.
- The 1000Mbps MAC uses the original Ethernet MAC operating at 10 times the speed. Running the Ethernet MAC at Gigabit speeds has created some challenges in terms of the implementation of CSMA/CD. To make CSMA/CD work at 1GHz, a minor modification was required. The slot time has been increased to 512 bytes, as opposed to 64 bytes for 10



**TABLE 3.10 A SUMMARY OF THE DIFFERENT GIGABIT ETHERNET PHYSICAL LAYER IMPLEMENTATIONS**

PHY Parameter	1000BASE-T	1000BASE-SX	1000BASE-LX	1000BASE-CX
IEEE Standard	802.3ab <sup>1</sup>		802.3z (July 1998)	
Number of pairs required	Four pairs	Two strands	Two strands	Two pairs
Cable Category required	Category 5 or better	50 or 62.5 $\mu$ m MMF	50 or 62.5 $\mu$ m MMF or 8–10 $\mu$ m SMF	150- $\Omega$ Twinax
Cable length	100m	220–550m	5000m (SMF) 550m (MMF) <sup>2</sup>	25m
Encoding	4D-PAM5	8B/10B	8B/10B	8B/10B
Connector specified	RJ-45	SC	SC	HSSC or DB-9
Proven	Not yet	3	3	3
Full-duplex capable	Yes	3	3	3
Broad product support	None shipping yet <sup>1</sup>	33	3	Probably not <sup>3</sup>

<sup>1</sup> 802.3ab approval is expected in mid-1999. Products are expected around the same timeframe.

<sup>2</sup> Requires special DMD patch cord. See Chapter 6 for more details.

<sup>3</sup> By September 1998, not a single Gigabit Ethernet vendor was shipping CX-based product.

## The IEEE Auto-Negotiation Standard

When 100BASE-T was invented, it was no longer safe to assume that a typical Ethernet RJ-45 connector would be carrying 10BASE-T. Instead, a UTP cable could carry one of seven different Ethernet signals: 10BASE-T, 10BASE-T full-duplex, 100BASE-TX, 100BASE-TX full-duplex, 100BASE-T4, 100BASE-T2, or 100BASE-T2 full-duplex. The IEEE incorporated a very elegant scheme into the 802.3u standard that has simplified your life as a LAN manager tremendously. The IEEE's auto-negotiation technology, originally developed by National Semiconductor and known as NWay, can tell what speed the other end of the wire can handle. The repeater, switch, or NIC will then automatically adjust its speed to the highest common denominator—the fastest speed that both can handle. This is how auto-negotiation works:

- Both devices need to contain the auto-negotiation logic. (All Fast Ethernet equipment shipping after mid-1996 includes this feature.)
- Auto-negotiation is an enhancement of the 10BASE-T link integrity signaling method and is backward compatible with link integrity.

provide a recap of the different IEEE Fast and Gigabit Ethernet physical layers that have been developed to run on the TIA/EIATIA/EIA 568-A system. (We have already briefly discussed these in Chapter 3, "Ethernet, Fast Ethernet, and Gigabit Ethernet Standards.") We examine the cabling requirements for the new 1000BASE-T/802.3ab standard in detail, and we provide some practical tips on how to certify your cabling infrastructure and how to deal with cabling contractors. Finally, we provide some tips on what to pay special attention to when running Fast and Gigabit Ethernet over Category 5 UTP.

We also cover two related subjects: the pros and cons of running fiber to the desktop and encoding/compression techniques.

### Encoding and Compression Techniques

Cabling equipment is rated in terms of megahertz (MHz). For example, Category 3 cabling is rated up to a maximum frequency of 16MHz and is synonymous with 10BASE-T, which transmits 10Mbps. Yet the same cable can transmit 100Mbps using 100BASE-T4 technology. How can that be possible?

Sometimes people confuse the terms *MHz* and *Mbps*. In the old days, one could say that it took a hertz (which equals a clock cycle) to transmit a bit (or a million hertz to transmit a million bits). Cabling is rated in MHz, which is the maximum frequency that can be transmitted reliably. To attain higher transmission rates, engineers today typically use several tricks to reduce the frequency of the signal to meet the bandwidth requirements of cable. The important thing to remember is that the data rate and the bandwidth required can be very different. Let's look at encoding in a bit more detail.

### Number of Pairs

The number of pairs is the most obvious way of increasing the transmission capability of cable. 10BASE-T uses only one dedicated transmission pair as well as one dedicated reception pair. Some faster versions of Ethernet use multiple transmission pairs. The data is split up over these multiple pairs, thereby reducing the bandwidth requirements for each individual pair.

For example, 100BASE-T4 utilizes a total of four pairs. Two pairs are dedicated as either transmit or receive, but the other two can act as either transmit or receive pairs. This means that a total of three transmission pairs are available at any one time. The downside of this *floating pair* scheme is that full-duplex transmission is not possible. 1000BASE-T goes one step further and uses four pairs for transmission.

### Switched Ethernet Cabling Requirements

The switched versions of 10, 100, or 1000Mbps Ethernet can all operate on the same cable as their shared-media versions because their wire speed is the same. The only difference is the number of end stations on that segment, which is two for a fully switched connection. If the switched connection operates in full-duplex mode, the cabling distance can often be extended, as no collisions will be

present in a full-duplex environment. Note that not all Ethernet physical layers are full-duplex capable. 10BASE-T, 10BASE-FL, 100BASE-TX, 100BASE-FX, 100BASE-T2, and all Gigabit Ethernet physical layers are full-duplex capable, but 10BASE5, 10BASE2, 10BASE-FP, 10BASE-FB, and 100BASE-T4 are not. Please refer to Chapter 3 for more details on full-duplex Ethernet.

### Dual-Duplex Transmission

Recent advances in DSP technology allow for *simultaneous* transmission and reception on the *same* pair of cabling, also known as dual-duplex. The new 1000BASE-T physical layer uses both simultaneous transmission and reception on all four pairs. This means full-duplex transmission.

### Bandwidth-Efficient Encoding

The earlier 10Mbps versions of Ethernet all used Manchester encoding, which is a simple, but not very efficient, encoding scheme. With the higher data rates of Fast and Gigabit Ethernet, engineers have been forced to use more sophisticated coding methods for transmitting data over limited-capacity UTP cabling. Let's look at the encoding technology in a bit more detail.

Ethernet transmission uses two different kinds of encoding schemes: block and line encoding. Block encoding takes several bits (four or eight, for example) and sends it out as one new code word. Line coding takes the parallel data stream and encodes it as a serial bit stream, ready to be transmitted across the wire in a serial fashion.

When calculating the overall bandwidth requirements (or efficiency) of a transmission scheme, you need to take into consideration the overall data rate, the block encoding efficiency, and the line encoding implications.

#### Block Encoding

Block encoding takes Ethernet raw parallel data (a series of bits) and encodes it according to a predetermined lookup table. 100BASE-TX, for example, uses a block-encoding scheme called 4B/5B. The 4B/5B encoding takes 4 bits of data (4B) and expands it to 5 coded bits (5B). Four of the coded bits contain the



actual data in accordance with some table. The fifth, "spare," bit is added to transmit additional information. There are three reasons for adding this extra bit as part of the coding scheme:

- The receiving station needs to receive the clock signal in order to decode the data accurately. Sending the clock separately would not be a very efficient use of the cabling infrastructure. The 5-bit code word can be chosen so that a clock transition is present often enough for the receiving station to regenerate the clock.
- The extra bit also allows the coded words to be chosen carefully so that an equal number of logical 0s and 1s are transmitted, which means the digital signal has no DC component, allowing for an easier design of the hardware.
- The 4B/5B encoding allows some error correction capability to be incorporated into the data stream. The 4B/5B coding scheme was originally developed for use with FDDI.

All coding schemes are typically abbreviated  $x\text{B}/y\text{B}$ . This always means  $x$  data bits are encoded into  $y$  coded bits. Sometimes three-level encoding schemes are used, in which case the shorthand becomes  $x\text{B}/y\text{T}$ .

The most common block encoding schemes are, of course, 4B/5B used for 100BASE-TX/FX, and the newer 8B/10B used for 1000BASE-X, which was originally developed by IBM for Fibre Channel. The lesser known 8B/6T coding scheme is used for 100BASE-T4 and is a new development. The most complex coding scheme is the PAM 5x5, which uses five different amplitudes (that is, voltage levels) and five different phases to generate 25 different codes.

The downside of the  $x\text{B}/y\text{B}$  encoding schemes is that they create a higher transmission rate. For example, the extra bit transmitted with the 4B/5B encoding of 100BASE-TX increases the transmission rate to 125Mbaud. (You can calculate this by taking the five coded bits and dividing them by the four raw bits, multiplied by the original data rate.)



### Symbols and Bauds

The speed or bit rate of a digital transmission is measured in bits per second (bps). Some transmission schemes, for example PAM5, don't just use 0's and 1's to transmit data, they use more complicated signals. PAM5 uses five different codes, not just two, to transmit the data. In this case, five different states or *symbols* are present. A particular voltage level actually contains two bits of information, or two *bits/symbol*.

The number of symbols transmitted per second is called the *baud rate*. The baud rate is only of interest to someone looking "under the hood" of a particular data transmission technology. That's because, ultimately, a certain baud rate will always be decoded into a bit rate again.

With simpler encoding schemes, such as Manchester, NRZ, and NRZI, the baud rate is typically twice the bit rate because a simple 0 to 1 transition represents an individual data bit. All the newer and more complex encoding schemes, such as PAM5, 4B/5B, 8B/10B, and 8B/6T, use more than two symbols, and therefore the bit and baud rate are quite different. For example, 100BASE-TX uses 4B/5B encoding that creates 5-bit words out of 4 bits of data via a lookup table. This kind of encoding contains 0.8 bits/symbol. That's because an extra bit is carried along for every four bits of data transmitted. This extra bit fulfills other purposes but does not actually constitute any useful data transfer. In this case, the bit rate is 100Mbps, but the baud rate is  $5/4 \times 100$ , or 125 Mbaud.

### Line Encoding

Line encoding converts a digital word (or series of bits) into a serial bitstream that can be transmitted over the wire. The common types of line encoding in use today with Ethernet are Non-Return-to-Zero (used for all fiber optic based Ethernet versions), Manchester (10BASE-T), and Multilevel-Threshold-3 (100BASE-TX). Let's look at the different line or serial bit encoding schemes in a bit more detail:

- Non-Return-to-Zero (NRZ) is a very simple encoding scheme. Some would say it's not an encoding scheme at all because it is similar to the raw data. NRZ signal is high for a logical 1 and low for a logical 0. NRZ encoding has a bandwidth efficiency index of 1: the bandwidth requirement equals half the data transmission rate.
- Return-to-Zero (RZ) code stays low for 0s. For logical 1s, RZ goes high for half a bit period and low for the remaining half bit.
- Non-Return-to-Zero Invert on Ones (NRZI) is a scheme in which a logical 0 is represented by a change in level. A 1 means change no level, which means that the level will go from low to high or high to low if a 0 is

transmitted (that is, the “invert on ones” part) but will stay the same for logical 1s. NRZI is used by 100BASE-FX. NRZI, like NRZ, has a bandwidth efficiency index of 1: the bandwidth requirement equals half the data transmission rate.

#### Note

*The problem with NRZ, RZ, and NRZI encoding is that they have no logical transitions in the bitstream if a series of 0s or 1s are transmitted: the receiver cannot make out the clock. In all three cases, some block encoding guarantees bit transitions in order for the receiver to decode the clock.*

- Manchester code has automatic transitions. It has a level transition in the middle of each bit. For a binary 1, the first half of the bit is high, and the second half is low. For a logical 0, the first half is low, and the second half is high. This encoding scheme guarantees easy clock decoding on the receiving end, but the drawback of this scheme is that it requires a lot of bandwidth. Manchester encoding has a bandwidth efficiency index of 2: the bandwidth requirement equals the data transmission rate. Because bandwidth was not a problem with the early versions of Ethernet (coaxial cable has a much higher bandwidth than UTP), all 10Mbps versions of Ethernet used this encoding scheme (including 10BASE-T).
- Multilevel-Threshold-3 (MLT-3) is unofficially named after its three inventors at Crescendo Communications: Mario Mazzola, Luca Cafiero, and Tazio De Nicolo. (Crescendo pioneered running FDDI over copper, which then became known as CDDI. Cisco acquired Crescendo in 1993.) MLT-3 is an encoding scheme that uses three different voltage levels (all previous ones discussed use two). The three levels are -1V, 0V, and +1V. For this reason, it can only be used with copper cabling because fiber uses either an on or off state. The level stays the same for consecutive logical 0s or 1s. A change in a bit means a change in the voltage level. The change occurs in a circular pattern of 0V, +1V, 0V, -1V, 0V, +1V. This circular, sinusoidal pattern change means that MLT-3 encoded data always resembles a smooth sine wave of a much lower frequency than the original bitstream, making it ideally suited for high-speed data transmission over UTP cabling. MLT-3 has a bandwidth efficiency of two-thirds, meaning the transmitted signal requires a bandwidth of only a quarter of the original transmission rate.
- PAM5 stands for pulse amplitude modulation—5. It is similar to MLT in that it uses more than two levels: five different voltage levels in this case.



Figure 6.1 shows the 100BASE-T physical layer, its 4b/5B block encoding/decoding, and the MLT-3 line encoding/decoding.

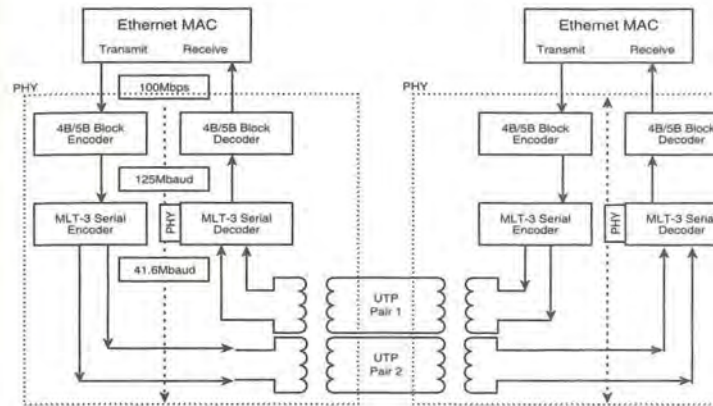


FIGURE 6.1 The Ethernet physical layer architecture consists of block encoding/decoding and serial or line encoding/decoding. The figure shows 100BASE-TX as an example.

Table 6.1 shows the most popular Ethernet PHYs in use (column 1). The table provides an overview of the different encoding and compression schemes that we just discussed. Columns 1 through 6 deal with the digital part of a physical layer. Columns 7 through 11 refer to the cabling and analog part of the physical layer. Let's discuss this table in a bit more detail:

- On the left, read the number of transmit pairs used for a particular PHY (column 2). The data rate per pair in column 3 is equal to the nominal data rate or wire speed divided by the number of transmit pairs.
- The PHY design engineers choose the block encoding scheme listed in column 4. Next, the coding efficiency is a function of the particular block encoding used (column 5).
- Next, go to column 6. This is the symbol rate in Mbaud for each cabling pair, which can be calculated by taking the data rate per pair (column 3) and dividing it by the coding efficiency (column 5).

same for  
band-  
half the

the bit-  
all three  
the clock.

on in the  
and the  
cond half  
n the  
s a lot of  
dex of 2:  
ecause  
et (coaxial  
ions of

ree  
Cafiero,  
copper,  
n 1993.)  
vels (all  
rd +1V.  
fiber  
utive log-  
el. The  
his circu-  
ways  
he origi-  
nission  
ls, mean-  
of the

MLT in  
his case.

TABLE 6.1 DIFFERENT ETHERNET STANDARDS AND THEIR ENCODING METHODS PRODUCE DIFFERENT DATA RATES AND BANDWIDTH USAGE

1. Technology	2. # of Transmit (TMT) Pairs	3. Data Rate/ Pair (=Data Rate / # of TMT Pairs)	4. Block Encoding Used <sup>3</sup>	5. Coding Efficiency <sup>2</sup> (Bits/ Symbol on Each Pair)	6. TMT1 Symbol Rate/Pair (=Data Rate / Coding Efficiency)
<i>UTP PHYs</i>					
10BASE-T	1 <sup>6</sup>	10Mbps	Manchester <sup>7</sup>	0.5 <sup>8</sup>	20 Mbaud
100BASE-TX	1 <sup>6</sup>	100Mbps	4B/5B <sup>9</sup>	0.8 <sup>10</sup>	125 Mbaud
100BASE-T4	3 <sup>11</sup>	33Mbps	8B/6T <sup>12</sup>	1.33 <sup>13</sup>	25Mbaud
100BASE-T2	2 <sup>15</sup>	50Mbps	PAM5 <sup>16</sup>	2 <sup>17</sup>	25Mbaud
1000BASE-T	4 <sup>15</sup>	250Mbps	PAM5 <sup>18</sup>	2 <sup>17</sup>	125Mbaud
<i>Fiber PHYs</i>					
10BASE-F	1 strand	10Mbps	NRZI <sup>7</sup>	0.5	20Mbaud
100BASE-FX	1 strand	100Mbps	4B/5B <sup>9</sup>	0.8 <sup>10</sup>	125Mbaud
1000BASE-X	1 strand	1000Mbps	8B/10B or 8B/1Q <sup>4</sup> <sup>14</sup>	0.8 <sup>19</sup>	1250Mbaud

<sup>1</sup>The required symbol rate can be calculated by dividing the data rate by the coding efficiency.<sup>2</sup>The encoding efficiency is an indication of how much data is transferred over the total cable within a given unit of time.<sup>3</sup>Block encoding is a tried and tested digital transmission technology. Many different encoding schemes are in use today.<sup>4</sup>This point refers to the frequency point at which there is no more signal (or power spectral density) present.<sup>5</sup>Nyquist, a famous mathematician, came up with a theory that the minimum theoretical channel bandwidth needs to be half the symbol rate in order to recover the original data stream. Another famous mathematician, Shannon, came up with a theory to calculate exactly how much bandwidth was required, but we won't bore you with these mathematical details.<sup>6</sup>Both 10BASE-T and 100BASE-TX use a dedicated transmit and receive pair.<sup>7</sup>In the old days, cable bandwidth was plentiful. (Remember that coaxial cable had lots more bandwidth than today's UTP versions.)<sup>8</sup>Manchester is terribly inefficient, only being able to transfer one bit in two clock cycles or 0.5 bits per symbol.<sup>9</sup>The 4B/5B takes four bits and generates five code bits. 4B/5B was derived from the ANSI FDDI standard.<sup>10</sup>100BASE-TX uses the available bandwidth more efficiently than Manchester. It can transfer 0.8 bits per symbol, calculated by taking four bits divided by five code-word bits. It requires 1.25 clock cycles to transfer a bit of data.<sup>11</sup>T4 uses one dedicated pair for transmission or reception, alternating two pairs in either direction. The fourth pair required is primarily a control pair. T4 cannot do full-duplex.<sup>12</sup>The 100BASE-T4 scheme converts 8 bits (8B) into three different levels (3T): 0, +1, and -1.



4T1 Symbol te/Pair Data Rate / ding Efficiency)	7. Signal conditioning/ Line Encoding	8. Zero of Frequency Spectrum (=Clock Frequency) <sup>4</sup>	9. Nyquist Minimum Channel Bandwidth Required	10. Cable Specified (= Half of Frequency Spectrum) <sup>5</sup>	11. Cable Bandwidth
<i>UTP PHYs</i>					
Mbaud		20MHz	10MHz	Category 3	16MHz
5 Mbaud	MLT-3 <sup>20</sup>	125MHz	62.6MHz	Category 5	100MHz <sup>20</sup>
Mbaud			12.5MHz	Category 3	16MHz
Mbaud		25MHz	12.5MHz	Category 3	16MHz
5Mbaud	Trellis forward error correction	125MHz	62.5MHz	Category 5	100MHz
<i>Fiber PHYs</i>					
Mbaud	=10×2 =20Mbps	20MHz	10MHz	Category 3	16MHz
5Mbaud	=100×2 =200Mbps	125MHz	62.5MHz	Fiber	>Giga hertz
50Mbaud	=1000×2 =2000Mbps	1250MHz	625MHz	Fiber	>Giga hertz

<sup>13</sup>100BASE-T4 is better still. Due to its 8B/6T encoding that uses three different voltage levels, its efficiency is 1.33 bits per symbol. The 8B/6T encoding scheme efficiency is calculated by dividing eight bits into six three-level voltages:  $4/3=1.33$ .

<sup>14</sup>1000BASE-X uses the Fibre Channel 8B/10B encoding, which generates a 10bit code word from eight bits of data. The name of the encoding scheme is about to be changed from 8B/10B to 8B/1Q4.

<sup>15</sup>Both 100BASE-T2 and 1000BASE-T use simultaneous bidirectional transmission over same pairs, also known as dual-duplex technology.

<sup>16</sup>The 100BASE-T2 encoding is also known as PAM5x5 because it transmits PAM 5 simultaneously on two pairs of cable.

<sup>17</sup>The PAM5 encoding used by 100BASE-T2 and 1000BASE-T can transmit two bits per symbol (five different voltage levels). It requires only 0.5 clock cycles to transfer a bit.

<sup>18</sup>The 1000BASE-T encoding uses four pairs to transmit, and therefore could be called PAM5x5x5x5, a little wordy. It is better known as 4D-PAM5. Some people refer to 1000BASE-T's encoding as enhanced TX/T2 because it uses the PAM5 encoding from T2, plus the clock frequency and symbol rate of TX to transmit 1000Mbps. (The 10X-performance improvement is derived from the 5X-clock frequency improvement, multiplied by double the number of pairs used.)

<sup>19</sup>8B/10B encoding used by 1000BASE-X has the same efficiency as 4B/5B.

<sup>20</sup>If 100BASE-T would have used Manchester encoding, a 200MHz cable would have been required. MLT-3 reduces the bandwidth requirements so that 100MHz Category 5 can be used.

Columns 7 through 11 deal with the physical layer properties of a particular encoding method:

- The PHY design engineer also chooses the Signal Conditioning/line encoding technology shown in Column 7. Of particular interest is 100BASE-TX, which uses MLT-3.
- Column 8 refers to a variable that depends on the clock frequency. This variable is key for determining the bandwidth requirements of the cable to be used. If the cable bandwidth is equal to or greater than this "zero of frequency spectrum" number, the signal can be easily transmitted across a particular cable.
- The Nyquist minimum frequency number refers to the minimum cable bandwidth that is required to transmit the signal. Note that this is a theoretical minimum and in practice is almost impossible to achieve (column 9).
- The last two columns show you what cable the PHY design engineers have chosen (columns 10 and 11). Note how PHY design engineers have always chosen a cable that provides a bandwidth somewhere greater than Nyquist's number yet smaller than the "zero of frequency spectrum" number.

Figure 6.2 shows the different Ethernet encoding schemes used. The data line shows the raw unencoded data. The transmitter generates the clock signal, which provides timing information. The receiver requires a clock signal as well, and either the clock is sent on a separate line, or the transmitted data embeds the clock information. As a separate clock uses up another cabling pair, Ethernet uses self-clocking encoding methods. Over the years as Ethernet transmission speeds have increased significantly, IEEE engineers have resorted to more and more complex encoding schemes to allow Ethernet to be transmitted over the limited-bandwidth UTP cabling that is so pervasive today. Because Fiber optic cabling has so much more bandwidth, data transmitted over fiber is typically transmitted raw without any encoding.

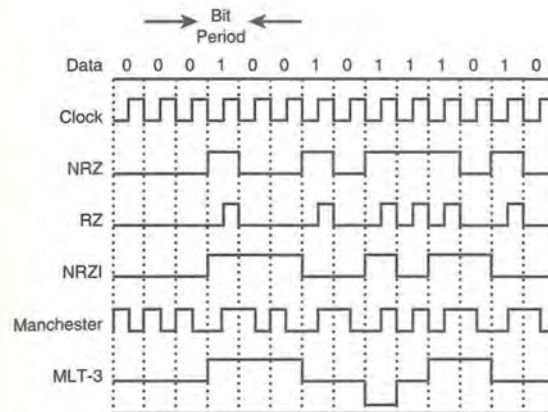


FIGURE 6.2 A digital data stream and a number of different line encoding schemes.

The analog signal present on the actual wire is a function of the block and subsequent line encoding. The frequency spectrum of the analog signal is important for determining the cabling requirements to transmit a particular signal. The cable acts like a filter: all signals that are injected into the cable but are at a frequency that are substantially beyond the rated bandwidth are effectively filtered out. The amount of energy that the entire frequency spectrum contains is called *power spectral density*. The cable used needs a bandwidth that permits a sufficient amount of the PSD to still be transmitted. Figure 6.3 shows the power spectral density of 10BASE-T, 100BASE-TX, and 1000BASE-T Ethernet. Notice how the PSD for each Ethernet version has been designed to decline rapidly at the bandwidth limits of the associated UTP cable.

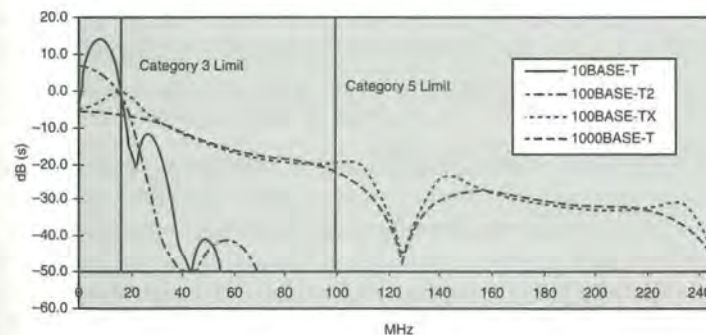


FIGURE 6.3 Power Spectral Density (PSD) is a measurement of how much energy is contained within a signal at a given frequency. Shown are the PSD for 10Mbps, 100Mbps, and 1000Mbps Ethernet. The dashed lines show the bandwidth limits of the respective Category 3 or 5 cable. Simulation data is courtesy of Level One Communications, Inc.



The following are factors to keep in mind when selecting a desktop NIC:

- *Media types*—After making the decision to purchase 10/100, the next choice is between 100BASE-TX or 100BASE-T4. Desktop NICs must be chosen based on the type of wiring found in that workgroup. Some workgroups have only Category 3 UTP and require 10BASE-T or 100BASE-T4 clients. Some have Category 5 UTP and can support 100BASE-TX. Because 100BASE-T4 never really gained critical mass in the marketplace, we recommend that 100BASE-TX be used whenever possible.
- *Full-duplex and flow control*—As discussed in Chapter 4, “Layer 2 Ethernet Switching,” desktop NICs can support both full-duplex and half-duplex operation. Full-duplex does not typically enhance client performance more than 10% at 10Mbps or 100Mbps data rates, so it should not be considered as a criterion when buying the client NIC. You won’t pay extra for 10Mbps or 100Mbps full-duplex support on a client NIC. If a NIC does support full-duplex, make sure it supports the 802.3x full-duplex flow-control specification.
- *Auto-negotiation*—Another factor in selecting client NICs is IEEE 802.3 Auto-Negotiation support. An auto-negotiating client NIC automatically determines the transmission scheme (100BASE-TX, 10BASE-T, full-duplex, and so on) with the far end of the wire, thereby making desktop connections automatic and painless. Auto-negotiation is a standard feature on most Fast Ethernet NICs that ship today.

### Desktop Bus Type

When looking at the installed base of 10Mbps Ethernet NICs in 1995, the predominant bus architecture was the Industry Standard Architecture (ISA). When considering the ISA bus as a vehicle for delivering 10Mbps dedicated bandwidth (in the case of 10Mbps switching) or 100Mbps bandwidth, however, there were definite performance issues:

- The ISA bus is only 16 bits wide.
- The ISA bus runs at a clock rate of only 8MHz.
- The ISA bus doesn’t allow burst data transfers.
- Most ISA bus adapters are I/O mapped, causing data transfers to be slower.

These factors produce a theoretical ISA bus bandwidth of 5.33MBps, or 42.67Mbps. A more realistic ISA bus bandwidth is determined by actual



measurements of many different NICs in various ISA systems. The real-world ISA bus bandwidth available to the NIC is only about one-fourth of the theoretical amount, or 11Mbps—barely enough to fill a 10Mbps pipe, and definitely not enough for multiple 10Mbps or 100Mbps NICs.

In desktop systems built after 1995, however, the Peripheral Component Interconnect (PCI) bus started to become predominant. Extended Industry Standard Architecture (EISA), Micro Channel Architecture (MCA), S-Bus (Sparc Workstations), and NuBus (Apple Macintosh) systems were also shipping but were becoming smaller in numbers. In 1997, all Pentium processor class or higher systems had PCI slots.

PCI delivers a theoretical bandwidth of 132MBps and true plug-and-play features, much like Sun's S-Bus. Today, high-speed buses, such as PCI and S-Bus, are the logical slot for high-performance network adapters. Table 8.1 compares ISA, EISA, MCA, PCI, NuBus, S-Bus, and PCMCIA (PC Memory Card Interface Adapter) bus architectures. Note that the PCI bus is high performance, processor independent, and widely supported by computer manufacturers. Even Apple and Sun, two of the biggest non-PC-compatible computer manufacturers, have adopted the PCI bus.

TABLE 8.1 A COMPARISON OF BUS TYPES

Feature	ISA	PCMCIA	EISA
Supported architecture	PC	PC (Mobile)	PC
Bus speed (MHz)	8	8	8
Theoretical bus bandwidth (MBps)	5.33	5.33	32
Practical bus bandwidth for a NIC (MBps) <sup>2</sup>	1.4	1.4	8
Bus width (bits)	16	16	32
Burst mode data transfer?	No	No	Yes
Processor independent?	Yes	Yes	Yes
Widely accepted by major computer manufacturers?	Yes	Yes (in laptops)	Yes
Setup utilities	Manual or plug-and-play <sup>3</sup> BIOS	Card & socket services	EISA Config. Utility

<sup>1</sup> PCI and CardBus are inherently scalable beyond 33MHz. Work is currently underway to scale to PCI bus to 50 and 66MHz speeds.

<sup>2</sup> Includes measurements in buses with multiple adapters, such as SCSI, IDE, and VGA.

<sup>3</sup> Plug-and-play BIOS refers to the ISA plug-and-play specification.

Although PCI has become dominant, many systems still ship with both PCI and ISA buses. That said, performance isn't the only reason to move your desktops and servers away from the ISA architecture. ISA has also been traditionally difficult to configure with multiple adapters. In the PC world, true auto-configuration comes with a migration to PCI.

10/100Mbps NICs provide an affordable upgrade path from 10Mbps only, but how much should you be willing to pay for that upgradability? Currently, 10/100 PCI NICs cost about 10% more than 10-only NICs. This price difference is so small that you should start planning *when* to purchase 10/100Mbps client NICs instead of evaluating whether you should.

If you are not already doing so, look for PCI in your desktop systems and select PCI 10/100 LAN adapters to connect those systems. What do you do with all those old ISA systems? Continue to use them as an important part of your client base, but network them with 10Mbps switches as discussed in Chapter 9, "Deployment."

MCA	NuBus	S-Bus	Card Bus	PCI
PC	Apple	Sun	Mobile PC	All
8	8	33	33	16–33 <sup>1</sup>
40	5.33	132	132	132
10	1.4	30–100	30–60	30–100
32	16	32	32	32–64
Yes	Yes	Yes	Yes	Yes
Yes	No	No	Yes	Yes
No	No	No	Yes	Yes
POR Register Conf Utility	True plug-n- play	True plug-n- play	True plug-n- play BIOS	True plug-n- play BIOS

Work is currently under way to extend PCI from a 32-bit data path to 64 bit in high-end workstations. The operating speed of PCI, currently 33MHz, is also being pushed to 50MHz and even 66MHz in some prototype workstations. So PCI in the desktop has a lot of headroom left.

Even so, far be it from the computer industry to sit still. In 1997, Compaq and Intel announced the Dual Independent Bus (DIB) architecture, which is even higher performance than PCI and allows for hot-plug capability in the desktop.

#### **LAN on Motherboard**

*LAN on motherboard* (LOM) is defined as a desktop, workstation, or server that has networking silicon right on the motherboard, obviating the need for an add-in NIC. Apple and Sun pioneered this concept, which was later implemented by Compaq, IBM, Hewlett-Packard, Siemens-Nixdorf, and other PC vendors. Traditionally, the LOM idea wasn't popular in the volume PC desktop market, mostly because there were too many different networking options to justify locking in one design. Now that 10/100 Ethernet has been integrated into a single chip (Intel's 82559), however, more PC vendors may start shipping products with this silicon on the motherboard. LOM may end up saving you a lot of money; therefore, depending on your network architecture and budget, you may want to consider LOM in your purchasing decision.

#### **Network Interface Card Drivers**

A *driver* is the piece of software that allows the desktop operating system to interface with the NIC hardware. Therefore a NIC has one driver for each kind of operating system. Common names for NIC drivers are NDIS (for Microsoft) and ODI (for Novell). Since the advent of Windows 95, it is common for a particular NIC driver to come included on the Windows 95 CD-ROM, thereby making it extremely easy to install the NIC.

The following are key issues when you are considering NIC drivers:

- *Driver support*—Brand-name NIC vendors typically develop much of their driver suite internally. This is important because if a driver wasn't developed in-house, it is possible that the vendor will not be able to support it very well. Many vendors strike a successful compromise with the network operating system (NOS) vendors, enlisting them to write a driver for that particular NOS. Although support for this kind of driver won't be the same as for one developed by the NIC vendor itself, it is still better than a complete third-party driver. As a hypothetical example, if 3Com and a third-tier NIC vendor both ask IBM to develop OS/2 drivers, it is likely that IBM will focus its efforts on 3Com. When using a



th to 64 bit in  
MHz, is also  
rkstations. So

Compaq and  
hich is even  
in the desktop.

or server that  
eed for an  
iter imple-  
d other PC  
ne PC desktop  
g options to  
integrated  
start shipping  
saving you a  
and budget,

g system to  
for each kind  
for Microsoft)  
non for a par-  
A, thereby

ers:

o much of  
river wasn't  
be able to sup-  
mise with the  
o write a dri-  
of driver  
itself, it is still  
example, if  
OS/2 drivers,  
ing a

non-brand-name NIC vendor, ask where the driver was developed. If you cannot get this information, ask where the customer support for that driver will come from.

Another point to consider when questioning driver support is hardware and software compatibility. Any brand-name vendor should be able to provide a list of OS versions tested. Likewise, any vendor that cannot produce a list of PC systems tested compatible with the NIC should be scrutinized.

- *Driver certification*—Most NIC vendors typically don't ship a NIC unless it has passed all major certifications. These include certification tests with NOS vendors, such as Novell, Microsoft, SCO, and IBM. Some NIC vendors sometimes bypass this expensive and time-consuming task. Without a certified NIC and driver, you have no way of confirming that the NIC vendor has tested the NIC and driver in a variety of configurations. Other tests that major NIC vendors perform are FCC emissions testing and environmental testing (temperature and mechanical stress tests).
- *Technical support and warranty*—A thorough review of a NIC vendor's technical support model should be as critical a factor in your NIC purchase decision as price or performance. If problems arise, you will be spending time with technical support long after the salespeople have disappeared.

As far as product quality goes, you must consider several key factors when selecting a NIC. NICs by vendors such as 3Com, Intel, and SMC all support lifetime warranties as well as 90-day money-back guarantees. Reliability is also a concern for mission-critical client/sever applications. NIC reliability is sometimes measured in *mean time between failures* (MTBF). Although this usually doesn't give you much information, you should be suspicious of any value below 180,000 hours. This indicates that a failure may occur, on average, every 180,000 hours, or 20 years. Most brand-name NIC vendors meet this level of reliability.

### Setup Wizards

In the majority of cases, NIC installation at the desktop goes smoothly, but sometimes it is very helpful to have a setup wizard. A setup wizard is a program that will help in cases where system conflicts occur or when you have problems getting connected to the network. Intel's EtherExpress PRO/100+ LAN Adapter Setup software is particularly good in this regard and can even run cross-network diagnostics in case the server is not responding.



### Software Support

Software support for laptop NICs becomes very important because the interface between laptop hardware and software is buried with the operating system. For this reason, it is always a good idea to buy a mobile NIC that has its driver included in your mobile operating system. If you are using Windows 95, Windows 98, or Windows NT, check whether the NIC driver is listed in the supported hardware list.

### Manageability

WfM applies to mobile systems as well. The upcoming WfM 2.0 specification will include several laptop extensions to the standard, not the least of which is advanced power-management capabilities. Because power usage is more important for battery-driven laptops, support for ACPI has to be implemented in conjunction with Wake-on-LAN capabilities. Because both standards are part of the WfM standard, make WfM-compliance a must for your laptop purchases.

### LAN on Notebook Motherboard

Several laptop vendors, including Toshiba and Dell, have elected to design the network connection right in the laptop itself. This allows connectivity to the network without the need for an add-in adapter. If you go with this strategy, make sure the LAN on motherboard solution you choose works in the necessary environments. Some laptops may need an additional PCMCIA or CardBus NIC for dialup access anyway, so a combination LAN/modem card might be just what the doctor ordered.

### Server NICs

A *server* is broadly defined as a specific kind of system built to be powerful, upgradable, and reliable. These qualities are also important in a good server NIC. One server may support hundreds of clients, which can put enormous stress on the server NIC. Unlike a desktop computer running Windows 95, where the CPU spends most of its time idle, the server is constantly working, trying to use every ounce of CPU power available. How hard a server CPU works is referred to as its *CPU utilization*. Imagine a server that spends 60% of its time providing data to a server NIC. That server only has 40% of its cycles left to perform other activities. In some Fast Ethernet adapters, the NIC can take up to 50% of the server CPU resources. Add a Gigabit Ethernet server NIC, and this number immediately pegs to 100%.

This is why the CPU utilization of a server NIC is very important. The less time the server CPU has to spend providing for the NIC, the more time it has to run applications, perform file transfers, and run NOS software. Servers also

typically support more expansion slots than do desktops, which allows a wide variety of configurations, including additional hard drives and network adapters. In addition, many high-end servers offer fault-tolerant features, such as redundant power supplies and hot-swappable hard drives.

### Server NIC Bus Types

With Fast Ethernet now widely installed in servers, many network managers have their sights set on Gigabit Ethernet. We discuss later why an investment in a high-speed expansion bus will help prepare your server for Gigabit Ethernet networking.

As discussed previously, ISA servers just won't support Fast Ethernet data rates. EISA, the historical server expansion bus of choice, supports Fast Ethernet data rates, but does not do so efficiently, both in terms of CPU utilization and price. EISA does not even come close to supporting Gbps data rates. PCI offers higher bandwidth and a path to the future through bus-width extensions and higher-frequency data transfers. When considering upgrades to current EISA servers, choose 10/100 NICs that offer lower CPU utilization, even at the expense of throughput. When considering new PC server purchases, you can choose from any number of PCI server NICs.

Most new servers support a 32-bit PCI bus running at 33MHz. This provides a total bus bandwidth of just over 1Gbps. Therefore, it is not very promising for a network manager looking to install a Gigabit server adapter. Fortunately, the PCI bus has been extended to 64-bit data paths and work is being done to increase its operating frequency to 66MHz. The combination of these two effects effectively quadruples the available bandwidth of PCI to just over 4Gbps.

In addition, servers that support the PCI hot-swap feature allow you to add or remove peripherals while the server is powered on. This is obviously beneficial because many mission-critical servers have to be in operation 24 hours a day. Compaq has led the charge in this area of server development.

### Wire Speed: 10, 100, or 1000?

Fast Ethernet, and specifically 10/100 NICs, started finding their way into servers as early as 1994. This means that some of those servers are ready for replacement today, and many network managers are considering Gigabit Ethernet NICs for the replacement systems.

Whereas Fast Ethernet server NICs ran at 10 or 100Mbps speeds and supported UTP and fiber, many early Gigabit NICs will run only at 1000Mbps and support fiber only. Because the NIC can't revert back to 100Mbps operation, the network manager should only selectively install Gigabit Ethernet NICs. It



host CPU. This is referred to as peer-to-peer transfer and allows, for instance, files to be sent directly from the hard disk to the LAN controller without first passing through the host CPU or main memory.

The memory and processing requirements to meet the I<sub>2</sub>O specification are expensive, but the benefits can be great, especially in lowering the CPU utilization of the server. Still, the prudent move is to evaluate I<sub>2</sub>O on a server NIC first before deploying it in production servers. You can obtain more information about the I<sub>2</sub>O specification from the Intel Web site at [www.intel.com](http://www.intel.com).

### The Impact of Gigabit NICs on Servers

Gigabit Ethernet server adapters are just starting to become available at the time of this book's publication. NICs from Alteon, 3Com, Intel, and several smaller startup companies have all been announced and/or started shipping. Figure 8.6 shows Alteon's ACEnic Gigabit Ethernet adapter.

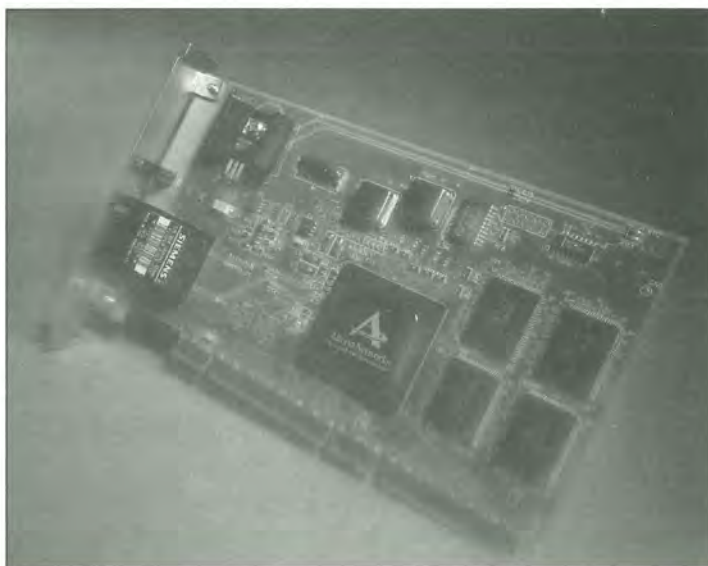


FIGURE 8.6 Alteon's Gigabit Ethernet NIC.

In general, Gigabit Ethernet NICs are hard on today's servers in that they can provide a bigger pipe than today's servers can fill. This will be rectified over time as server system technology steadily improves to match the speed of a

Gigabit connection. And when that happens, there will be a new push to fill a full-duplex Gigabit pipe, or who knows, maybe even 10Gbps!

### Throughput and CPU Utilization

The latest data out of Intel labs is that a Gigabit server adapter can send about 670Mbps of data at close to 100% CPU utilization. This translates to a 6.7 Performance/Efficiency Index. Not bad for a first try. The performance battles between Gigabit NIC vendors have only just begun. Engineers are, at this very moment, trying to increase Gbps NIC throughput and reduce CPU utilization. Unfortunately, most of the improvements will have to take place in the server rather than on the Gigabit NIC.

### Server Subsystems

Most new servers support a 32-bit PCI bus running at 33MHz. This provides a total bus bandwidth of just over 1Gbps. Not very promising for a network manager looking to install a Gigabit server adapter. Fortunately, the PCI bus has been extended to 64-bit data paths and work is being done to increase its operating frequency to 66MHz. The combination of these two efforts effectively quadruples the available bandwidth of PCI to just more than 4Gbps. Figure 8.7 shows this.

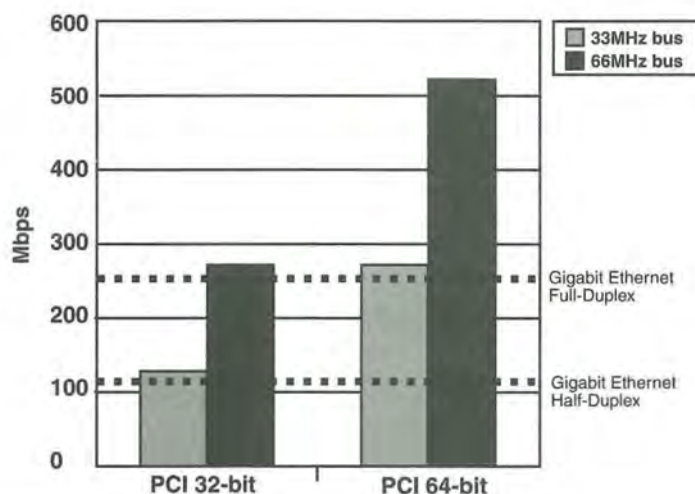


FIGURE 8.7 Theoretical bandwidth of the PCI bus.

The first recommendation for the Gigabit-inclined network manager is to buy servers with 64-bit PCI slots, and use 64-bit Gbps server NICs in those slots.