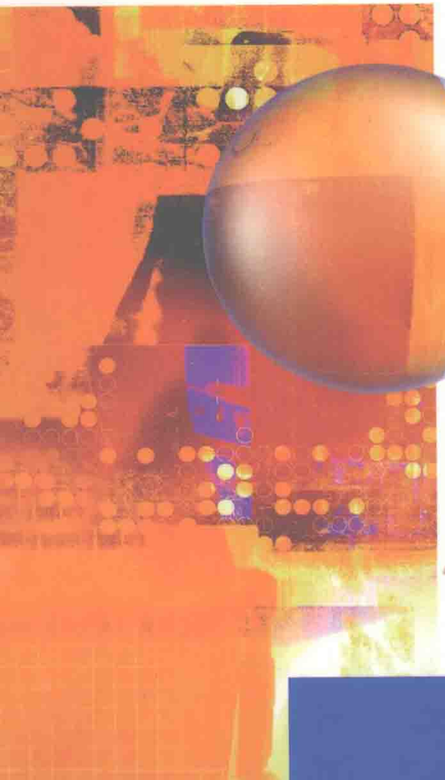
McGraw-Hill TELECOM
PROFESSIONAL

FIBRE CHANNEL for SANs



Step-by-step guidance through all 5 FC layers

Bandwidth-saving solutions for multimedia

Troubleshooting FC networks

ALAN F. BENNER

Fibre Channel for SANs

Alan F. Benner

McGraw-Hill
New York • Chicago • San Francisco
Lisbon • London • Madrid • Mexico City
Milan • New Delhi • San Juan • Seoul
Singapore • Sydney • Toronto

Cataloging-in-Publication Data is on file with the Library of Congress

McGraw-Hill

A Division of The ***McGraw·Hill*** *Companies*

Copyright © 2001 by The McGraw-Hill Companies, Inc. All rights reserved. Printed in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a data base or retrieval system, without the prior written permission of the publisher.

1 2 3 4 5 6 7 8 9 0 DOC/DOC 0 9 8 7 6 5 4 3 2 1

ISBN 0-07-137413-2

The sponsoring editor of this book was Steve Chapman. The editing supervisor was Steven Melvin, and the production supervisor was Sherri Souffrance.

Printed and bound by R. R. Donnelley & Sons, Inc.

This book is printed on recycled, acid-free paper containing a minimum of 50% recycled, de-inked fiber.

McGraw-Hill books are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. For more information, please write to the Director of Special Sales, Professional Publishing, McGraw-Hill, Two Penn Plaza, New York, NY 10121-2298. Or contact your local bookstore.

Information contained in this work has been obtained by The McGraw-Hill Companies, Inc. ("McGraw-Hill") from sources believed to be reliable. However, neither McGraw-Hill nor its authors guarantee the accuracy or completeness of any information published herein and neither McGraw-Hill nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that McGraw-Hill and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

List of Figures

Figure 1.1
Example of an Enterprise or Service Provider SAN+LAN Topology 7

Figure 2.1
Fibre Channel structural hierarchy. 15

Figure 2.2
FC-0 link. 16

Figure 2.3
Examples of Point-to-point, Fabric, and Arbitrated Loop topologies. 17

Figure 2.4
Transmitter and receiver FC-1 and FC-0 data flow stages. 19

Figure 2.5
Sample Data Frame + ACK Frame transmission, for bandwidth calculation. 22

Figure 2.6
Building blocks for the FC-2 Frame / Sequence / Exchange hierarchy. 23

Figure 3.1
Physical configuration of a host-to-disk Fibre Channel Arbitrated Loop connection. 44

Figure 3.2
Link initialization through Ordered Set Transmissions. 46

Figure 4.1
Nomenclature for describing FC-0 cable plant options. 63

Figure 4.2
Defined cable plant technology options. 63

Figure 4.3
Specifications for single-mode optical fiber links. 65

Figure 4.4
Duplex SC connector for single- and multi-mode optical links. 66

Figure 4.5 Eye diagram mask for the transmitter output. 67

Figure 4.6 Specifications for multi-mode fiber short wavelength laser and LED links. 68

Figure 4.7 Specifications for electrical cable links. 72

Figure 4.8 Transmitter eye diagram mask for coaxial cable links. 73

Figure 5.1 Sample conversion of an FC-2 byte to FC-1 Transmission Character notation. 78

Figure 5.2 5B/6B and 3B/4B coding conversions. 79

Figure 5.3 Full 8B/10B code: Valid Special Characters. 80

Figure 5.4 Full 8B/10B code: Valid Data Characters, 1 of 3. 81

Figure 5.5 Full 8B/10B code: Valid Data Characters, 2 of 3. 82

Figure 5.6 Full 8B/10B code: Valid Data Characters, 3 of 3. 83

Figure 5.7 Example logical structure of an 8B/10B encoder. 84

Figure 5.8 Transmission order of bits and bytes in a word. 84

Figure 5.9 Example logical structure of a 10B/8B decoder. 85

Figure 6.1 Frame Delimiter Ordered Sets. 92

Figure 6.2 Primitive Signal and Primitive Sequence Ordered Sets. 96

Figure 6.3 Port state transition chart. 100

Figure 7.1 Frame and Frame Header formats. 108

Figure 7.2 Link Control Frame definitions. 113

试读结束：需要全本请在线购买： www.ertongbook.com

Figures

Figure 7.3
Data Frame definitions. 114

Figure 7.4
Frame Payload formats for Data Descriptor, Unsolicited Command, and Command Status Information Categories. 116

Figure 7.5
S_ID/D_ID address space partitioning. 117

Figure 7.6
CS_CTL Class Specific Control Field 119

Figure 7.7
Bit definitions for F_CTL Frame header field. 121

Figure 7.8
Summary of F_CTL bit usage and validity in Data and Link Control Frames. 126

Figure 7.9
Bit definitions for DF_CTL (Data Field Control) Frame Header field. 128

Figure 7.10
Frame reject reason codes. 141

Figure 8.1
N_Port- and Class-dependent conditions for considering Sequences as complete. 148

Figure 8.2
Sequence Status Block format. 154

Figure 8.3
Exchange Status Block format. 156

Figure 9.1
Pre-Login Default Service Parameters. 158

Figure 9.2
Frame flow during Fabric Login and N_Port Login. 160

Figure 9.3
Format of the N_Port Login and Fabric Login requests and their Accept replies. 161

Figure 9.4
Payload format of Logout (LOGO) Command and corresponding Accept (ACC). 166

Figure 9.5
LS_Command codes and Common Service Parameters for FLOGI and PLOGI Commands. 167

Figure 9.6
Bit field definitions for FLOGI and PLOGI Common Service Parameters 168

Figure 9.7
Formats of worldwide Port_Names and Node_Names. 170

Figure 9.8
F_Port and N_Port Class Service Parameters. 171

Figure 9.9
Bit field definitions for Class Service Parameters. 172

Figure 9.10
Multiple images are addressable behind a single Port using Process Associators. 175

Figure 9.11
Frames and Sequences transmitted in procedure for establishing streaming credit. 177

Figure 10.1
Examples of N_Port Login over a Fabric using the three Classes of service. 185

Figure 11.1
Basic Link Service commands. 197

Figure 11.2
Format of the Basic Accept to the ABTS Basic Link Service command. 198

Figure 11.2
Format of the Basic Accept to the ABTS Basic Link Service command. 198

Figure 11.3
Basic Reject BLS Response format, BA_RJT reason codes, and reason explanation codes. 199

Figure 11.4
Extended Link Service commands, with LS_Command codes and usage requirements. 201

Figure 11.5
Link Service Reject (LS_RJT) format and reason codes. 204

Figure 11.6
Link Service Reject (LS_RJT) explanation codes. 205

Figure 11.7
Payloads for the RCS, RES RSS, and RSI Requests. 208

Figures

Figure 11.8

Payloads for the RTV, RLS, ECHO, TEST, and RRQ ELS Requests and Replies. 211

Figure 12.1

Reply format to Read Link Error Status Block Request. 240

Figure 13.1

Placement of optional headers in Frame. 242

Figure 13.2

Summary of optional header usage. 243

Figure 13.3

Format of Network_Header. 244

Figure 13.4

Format of Association_Header. 245

Figure 14.1

Example topology for Class 1 Dedicated Connection discussion examples. 253

Figure 14.2

Possible responses to a SOFc1 connect-request Frame. 256

Figure 15.1

Applicability of the flow control mechanisms to different Classes of service. 264

Figure 15.2

End-to-end flow control management rules. 267

Figure 15.3

Buffer-to-buffer flow control management rules. 270

Figure 15.4

Summary of combined flow control management for Class 2 transmission. 272

Figure 16.1

FC-AL Hub, showing internal port bypass circuitry and loop topology. 278

Figure 16.2
Initialization sequences for Arbitrated Loops. 285

Figure 16.3
Mapping of the 127 neutral parity (F/N)L_Port AL_PA addresses onto Initialization Sequence bit positions. 287

Figure 16.4
Loop Port State Machine state diagram, simplified. 289

Figure 17.1
Example multi-switch Fabric 301

Figure 17.2
Traffic interaction in a multi-stage switch Fabric 304

Figure 18.1
Format of an IP Information Unit. 319

Figure 18.2
Format of ARP request and ARP response datagrams or Frame payloads. 320

Figure 18.3
Functional equivalence between SCSI-3 and FCP functions. 324

Figure 18.4
Information Units (IUs) defined for FCP. 326

Figure 18.5
FCP Information Unit Formats. 327

Figure 18.6
Information Unit transfers under FCP read and write operations. 330

Figure 19.1
InfiniBand System Area Network Topology 339

Preface

In the preface to the first edition of this book I wrote about the difficulty of keeping up with the flow of information in a book format, since the state of the technology changes so much during the time between when a book is started and when it actually appears. Since then, the pace of change in Fibre Channel, Storage Area Networks (SANs), and server networking has, if anything, accelerated. At least a dozen new standards have been approved since the first edition, and several more are either under development or have been superseded by better work.

Fortunately, however, most of the current progress is up in the network management and data management layers, so that the basic operation of the network, which is the primary subject of this book, has pretty well stabilized. It's possible, therefore, to expect that the subject matter here will stay stable for a reasonable period of time.

The most fundamental change in Fibre Channel since the first edition of this book, however, has been in where it gets used. When Fibre Channel was first created, it was not clear exactly which problems it would be most effective at solving, since it was such a widely applicable technology. Since then, Fibre Channel has become the de facto standard interconnect technology for storage networking. In fact, the term Storage Area Network, newly defined since that time, has currently become essentially synonymous with Fibre Channel.

Since Fibre Channel has become so tightly linked with storage networking, much of the progress in Fibre Channel has been towards solving specific storage networking problems. In this book, however, I've tried to keep a general focus on the whole of Fibre Channel technology, since restricting the discussion to only those aspects important in current-day SANs runs the risk of ignoring parts of the technology that will be extremely important in the future.

The intention of this book is two-fold. First, it's intended to be an overview guide to the concepts, the structures, and the goals of the Fibre Channel architecture at a fairly detailed level. A dedicated reader should be able to use it to understand most of the details of Fibre Channel before referring to the ANSI materials for authoritative information. Second, and perhaps more importantly, it's an attempt to show the reasons why the network is designed the way it is. Networking technology is consistently becoming more and more important, and new technologies are being developed regularly. As new networking technologies are invented, I hope that this book will help people exploit the goods features in the already-existing networks.

A few words on notation. Determining a consistent notation in this type of work is not trivial, since the subject matter bridges both computer and communications arenas, which have traditionally used slightly different notations. For example, communications data rates are generally measured in megabits per second, where "mega" means 10^6, while computer data is measured in megabytes, where "mega" means 2^{20}. The text mixes both somewhat, using "b" to represent bits, as in Gbps, and "B" to represent bytes, as in MBps. In recognition of the communications-oriented nature of the subject, the prefixes "mega (M)" and "giga (G)" will mean 10^6 and 10^9 here, rather than 2^{20} and 2^{30}. All numbers in the book are written in binary (b'0110 0101'), hexadecimal (x'FF FFFD'), or decimal (65,532) formats. Single bits are written as 1 or 0.

A number of common words, such Sequence, Exchange, and Connection, have specific meanings in Fibre Channel that are quite distinct from their common usage. In this book, words with specific Fibre Channel meanings are generally capitalized to distinguish them from the common usage. This capitalization generally matches the format used in the ANSI standards documents. Information provided here is in the public domain, through generally available books, articles, ANSI documents, or other reference material.

Several terms used in this book, such as ATM and HIPPI, are taken from other architectures. Any trademarks used are properties of their rightful owners. Ethernet is a trademark of the Xerox Corporation. ESCON, FICON, and SBCON are trademarks or registered trademark of the IBM Corporation in the United States or other countries or both. InfiniBand is a registered trademark of the InfiniBand Trade Association.

The book is organized as follows. The first few chapters give an overview of the features and goals for the Fibre Channel architecture, along with an example of how data is transmitted under a Fibre Channel network.

The middle chapters cover the concepts and structures of Fibre Channel in a fair amount of detail. These include chapters on all of the Fibre Channel physical components and logical constructs, supported functions, flow control, and error recovery.

The final chapters cover configuration and operation of the Arbitrated Loop topology, mapping of Fibre Channel constructs to upper level protocols such as SCSI and the IP level of TCP/IP. These chapters show how Fibre Channel fits in with currently existing software and operating system levels. In the last chapter I have taken the opportunity to make some predictions for the future of Fibre Channel, SANs, and server networking, as far as I dare.

Thanks are due to far too many people to list here, but I'll try — I apologize in advance to those I may have missed. Many thanks to Carl Zeitler, Ki Won Lee, Mike Yang, Dan Eisenhower, Ron Cash, Roger Weekly, Giles Frazier, Jerry Chapman, Jerry Rouse, Jonathan Thatcher, Bill George, Al Widmer, Tom McConathy, Casey Cannon, Gary Nutt, R. Bryan Cook, Paul Green, Dal Allan, Martin Sachs, Horst Truestedt, Richard Taborek, Schelto Van Doorn, and Roger Cummings, who helped especially during the writing of the first edition. Thanks to Herman Presby, Ivan Kaminow, and Jon Sauer,

试读结束：需要全本请在线购买： www.ertongbook.com

for helping me get involved in optical networking. Thanks to Frank Kampf, Bob Stucke, Harish Sethu, Doug Joseph and Bob Cypher for helping me understand some of the many issues in computer communications. Thanks also to Steve Chapman and Caroline Levine for help in putting the book together and getting the project finished, and to Steve again for help the second time around.

I particularly want to thank Joe Mathis, for teaching me a lot about how to actually get work of this kind done successfully, and Renato Recio, for always demonstrating how to do it well. I also want to thank the InfiniBand crowd, you know who you are, for the chance to do some great work with great people. Hopefully we'll get to do more in the future.

I hope it's clear from this book that I love figuring out and explaining how to build better communications systems — I hope it's helpful to people.

This book is dedicated to Micah, for putting up with the time I spend on this stuff when I should be doing things more important, and to Ashley and Denali, who are busy building their own futures — it's wonderful to watch it happening.

Table of Contents

Chapter 1. Fibre Channel and Storage Area Networks 1
Fibre Channel Features 3
Storage Area Networks 5
SAN topologies 6
SANs, LANs, and NAS 8
LANs vs. SANs: Differences in Network Design 8
LANs vs. SANs: Differences in Protocol Design 9
Network-attached Storage (NAS) and Storage Area Networks (SAN) 10
Goals of This Book 12

Chapter 2. Overview 13
FC-0 General Description 16
FC-1 General Description 17
FC-2 General Description 18
Physical Model: Nodes, Ports, and Topologies 20
Bandwidth and Communication Overhead 21
Building Blocks and Their Hierarchy 22
Link Control Frames 27
General Fabric Model 27
Classes of Service 31
Basic and Extended Link Service Commands 34
Arbitrated Loop Functions 35
Protocols 35
Segmentation and Reassembly 36
Data Compression 37
Error Detection and Recovery 37
FC-3 General Description 38
FC-4 General Description 39
IP over Fibre Channel 39
SCSI over Fibre Channel 40
FICON, or ESCON over Fibre Channel 41

Chapter 3. Initialization and Data Transfer 43
Link Initialization 45
Loop Initialization 47
Fabric and N_Port Login 49
Initial Sequence Transmission 51
Completion of an Exchange 54
Protocol Not Covered by the Example 55
Error Detection and Recovery 56
Class 1 and Fabric Operation 57
Other Protocol Operations 57

Chapter 4. FC-0: Physical Interface 59

General Characteristics 60
Transmitter and Receiver 61
Intentional Transmission of Invalid Code 62
FC-0 Nomenclature and Technology Options 62
Long-Wavelength Laser Single-Mode Link 64
Cable Plant 65
Transmitter 66
Receiver 67
Short-Wavelength Laser Multi-Mode Fiber Links 68
Cable Plant 69
Transmitter 69
Receiver 70
Open Fiber Control Safety System for SW Laser Links 70
LED Multi-Mode Fiber Links 71
75-Ohm Coaxial Cable Electrical Links 71
150-Ohm Shielded Twisted Pair Electrical Links 74

Chapter 5. FC-1: 8B/10B, Transmitters, and Receivers 75
8B/10B Transmission Code Overview 76
Notation Conventions 77
Character Encoding and Decoding 78
Transmission, Reception, and Error Detection 80
Transmitter and Receiver States 85

Chapter 6. Ordered Sets 89
Ordered Sets 91
Start of Frame Delimiters 91
End-of-Frame Delimiter 93
Primitive Signals — Idle and R_RDY 95
Primitive Sequences 97
Port States on Non-Loop Topologies 98
Active State 99
Link Recovery States (LR1, LR2, and LR3) 99
Link Failure States (LF1 and LF2) 102
Offline States (OL1, OL2, and OL3) 103
Primitive Sequence Protocols 104
Link Recovery Protocols 104
Online to Offline Protocol 105

Chapter 7. FC-2: Frames 107
Frame Format 109
Idle and R_RDY Primitive Signals 109
Start of Frame Delimiter 109
Frame Header Fields 110
Optional Headers 110
Payload 110
CRC Field 111

End-of-Frame Delimiter 111
Frame Header Fields 112
Routing Control Field: R_CTL 113
Address Identifiers: S_ID and D_ID 115
Class Specific Control: CS_CTL 119
Data Structure Type: TYPE 120
Frame Control: F_CTL 120
Sequence ID: SEQ_ID 127
Data_Field Control: DF_CTL 127
Sequence Count: SEQ_CNT 128
Originator Exchange_ID: OX_ID 128
Responder Exchange_ID: RX_ID 129
Parameter 130
Data Frames and Link Control Frames 131
Data Frame Types and Handling 132
Link Credit Reset (LCR) Frame 134
Detailed Responses to Data and Link Control Frames 135
R_RDY Response 135
ACK Frames 135
F_BSY Frames 138
P_BSY Frames 139
F_RJT and P_RJT Frames 140
Notify (NTY) and End (END) Frames 142

Chapter 8. FC-2: Sequences and Exchanges 143
Sequence Management 144
Sequence Initiation 145
Sequence Handling Validity 146
Sequence Count Management 147
Normal Sequence Completion 147
Exchange Management 149
Exchange Origination and X_ID Assignment 150
Exchange Management 151
Exchange Termination 152
Sequence and Exchange Status Blocks 152
Sequence Status Block Rules 153
Exchange Status Block Rules 154

Chapter 9. Login and Logout Services 157
Login and Logout Overview 158
Fabric Login 161
N_Port Login 163
Logout 165
N_Port and F_Port Service Parameters 166
Common Service Parameters 166
N_Port_Name or F_Port_Name 169
Node_Name or Fabric_Name 170
N_Port and F_Port Class Service Parameters 171

Vendor Version Level 173
Overview of Process Login/Logout 174
Procedure to Estimate End-to-End Credit 176
Procedure Details 178

Chapter 10. Classes of Service 181
Class 1 Service 184
Variations on Class 1 Service 187
Class 2 Service 188
Class 3 Service 189
Performance and Reliability with Class 3 Service 190
Intermix 191
Class 4 — Fractional 192
Class 6 — Uni-Directional Dedicated Connection 194

Chapter 11. Link Services 195
Basic Link Service Commands 196
No Operation (NOP) 197
Remove Connection (RMC) 197
Basic Accept (BA_ACC) 198
Basic Reject (BA_RJT) 198
Abort Sequence (ABTS) Command 199
Extended Link Service Command Overview 200
Types of Extended Link Service Commands 202
Extended Link Service Replies 203
Link Service Reject (LS_RJT) 203
Accept (ACC) 203
ELS Requests: Login, Logout, and Abort Exchange 203
N_Port Login (PLOGI) 204
F_Port Login (FLOGI) 206
Logout (LOGO) 206
Abort Exchange (ABTX) 206
ELS Requests: Status Determination and Initiative Request 207
Read Connection Status (RCS) 207
Read Exchange Status Block (RES) 207
Read Sequence Status Block (RSS) 209
Request Sequence Initiative (RSI) 209
ELS Requests: Credit Determination 209
Establish Streaming (ESTS) 209
Estimate Credit (ESTC) 210
Advise Credit (ADVC) 210
ELS Requests: Miscellaneous Functions 210
Read Timeout Value (RTV) 212
Read Link Error Status Block (RLS) 212
Echo (ECHO) 212
Test (TEST) 212
Reinstate Recovery Qualifier (RRQ) 213
ELS Requests for Alias IDs 213

Get Alias_ID (GAID) 213

Fabric Activate and Deactivate Alias_ID (FACT, FDACT) 214

N_Port Activate and Deactivate Alias_ID (NACT, NDACT) 214

ELS Requests for Class 4 Service 214

Quality of Service Request (QoSR) 214

Read Virtual Circuit Status (RVCS) 215

FC-4 Link Services 215

Chapter 12. Error Detection and Recovery 217

Timeout Periods 219

R_T_TOV 219

E_D_TOV 219

R_A_TOV 220

CR_TOV 221

FC-AL Loop Timeout 221

Usage of Timeouts 221

Sequence Timeout 222

Link Error Detection and Recovery 224

Link Recovery: Secondary Effects 225

Frame Reception and Frame Validity 226

Detection of Missing Frames 226

Exchange and Sequence Integrity 228

Exchange Error Policies for Class 1 and Class 2 229

Rules Common to All Discard Policies 229

Abort, Discard Multiple Sequences Error Policy 230

Discard Multiple with Immediate Retransmission 231

Abort, Discard a Single Sequence Error Policy 232

Process with Infinite Buffering Error Policy 232

Class 3 Sequence Error Handling 233

Sequence Recovery 233

Error Recovery Strategy 234

Abort Sequence Protocol for Sequence Recovery 234

Sequence Retransmission for Class 1 Recovery 238

Stop Sequence Protocol Overview 239

Link Error Status Block Rules 240

Chapter 13. Optional Headers and Special Functions 241

Optional Headers 242

Network_Header 244

Association_Header 244

Device_Header 246

Special Functions 246

Data Compression 246

Encryption 246

Chapter 14. Class 1 Connection Management 247

Overview: Normal Procedures 249

试读结束：需要全本请在线购买： www.ertongbook.com