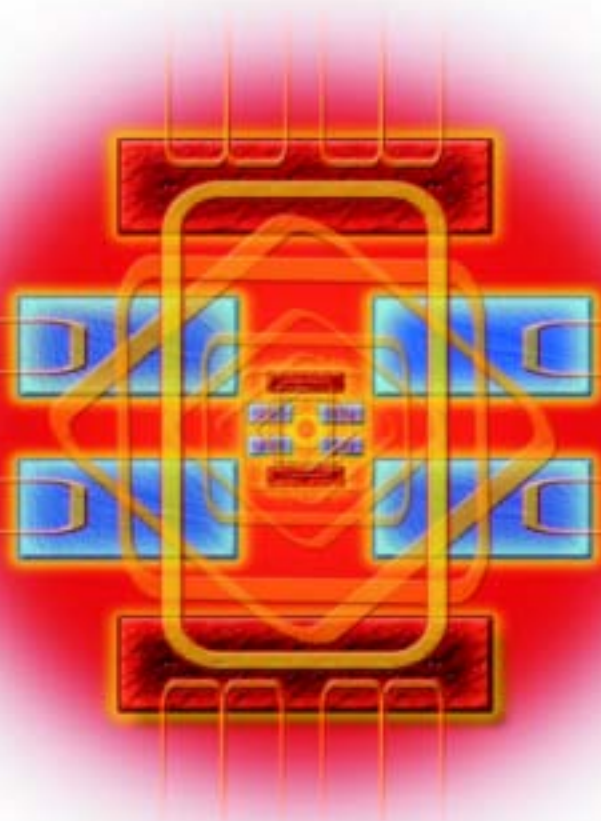


# Weaving High Performance Multiprocessor Fabric

Architectural insights into the Intel® QuickPath Interconnect

By Robert A. Maddox, Gurbir Singh and Robert J. Safranek



Intel  
PRESS

Books by Engineers, for Engineers



# Contents

---

**Preface ix**

<b>Chapter 1</b>	<b>A First Look at the Intel® QuickPath Interconnect 1</b>
	Evolution of Microprocessor Interconnect 1
	Solving the Cache Coherency Problem 4
	Write-Through Caching 5
	Write-Back Caching 6
	Tying It All Together 7
	Anatomy of a Multiprocessor System 8
	Evolution of a Link based System 10
	Anatomy of an Intel® QuickPath Interconnect System 13
	Terminology in Systems with the Intel® QuickPath Interconnect 18
	Layers of the Intel® QuickPath Interconnect 19
	The Physical Layer 21
	The Link Layer 22
	The Routing Layer 25
	The Protocol Layer 25
	Performance of the Intel® QuickPath Interconnect 28
	Reliability of the Intel® QuickPath Interconnect 29
	Deployment of the Intel® QuickPath Interconnect 29
	Designing with the Intel® QuickPath Interconnect 30
	Summary 31

**Chapter 2 Intel® QuickPath Interconnect in Operation 33**

- Life of an Intel® QuickPath Interconnect Transaction 33
  - System Address Spaces 34
  - Intel® QuickPath Interconnect Protocol Messages 35
  - The Notion of Node IDs 37
  - Directed versus Broadcast Messages 38
- End to End Sequence of Events 38
  - Request Generation 38
  - Operations of the Source Address Decoders 38
  - Routing over the Fabric 39
  - Delivery to the Target Device—the Target Address Decoder Function 43
  - Transaction Completion 44
  - Data and Non-Data Responses 44
  - Snoop Generation and Responses 45
- Examples of Intel® QPI Transactions 45
  - Feynman/Sequence Diagram Notation 48
  - Processor to Coherent Memory 49
  - Basic Flow Examples 51
  - Core to/from Non-Coherent Memory 57
  - Core to and from I/O 61
  - I/O to and from Memory 63
  - Interrupts 63
  - For a Deeper Dive into the Intel® QuickPath Interconnect Protocol 66
    - Conflict Resolution with Example 66
    - Using the Home Snoop Protocol 68
    - The Forward State Optimization 70
    - Invalidating Write Flow 71
    - Broadcast Transactions 74
    - Lock Flows 76
- Summary 79

**Chapter 3 Linking Two Devices 81**

- Enabling Communication 81
  - The Problem 81
  - Goals and Requirements 82
  - An Imperfect World 84
  - The Physical and Link Layers Exist to Provide These Functions 85
- Physical Layer Responsibilities 85

- Quickly Getting 80-Bit Flits from Point A to Point B 86
- Dealing with an Imperfect Path between A and B 86
- Transition from the Intel FSB 88
  - FSB Characteristics 88
  - Intel QPI Characteristics 89
- Logical and Electrical Sub-Blocks 92
- Physical Layer Electrical Characteristics 93
  - General Features 93
  - Transmitter Equalization 94
  - Equalization Example 96
  - Perfection Is Not Required 101
- Physical Layer Logical Functions 101
  - Basic State Machine 101
  - Reset State 102
  - Finding Another Device (Detect State) 102
  - Finding Bits and Phits (Polling State) 103
  - Exchanging Parameters (Polling State) 104
  - Establishing the Link Width (Config State) 105
  - Parameters and Flit Lock (Config State) 107
  - Transition to Operational State LO 107
- Physical Layer Miscellaneous Topics 109
  - Taking Down a Link—Inband Reset 109
  - Link Speed 110
    - Core Speed and Link Speed 110
- Link Layer Responsibilities 110
  - Higher Layer Services 111
  - Flow Control 111
  - Error Detection and Recovery 112
- Details of Link Layer Functions 112
  - Link Layer Initialization 112
  - Services to Higher Layers 113
  - Flow Control 118
  - Link Reliability 120
  - Interleaving 122
- Packet Formats 124
  - Strict Encapsulation Is Not Used 124
  - Intel QPI Packets—One or More Flits 124
  - Defined Packets and Their Contents 125
  - Specific Fields 127
  - Putting It All Together 129
- Controlling the Features 130

- Physical Layer Configuration 130
- Physical Layer Device Registers 132
- Link Layer Configuration 132
- Link Layer Device Registers 134
- Summary 135

## **Chapter 4 System Initialization 137**

- Basic Concepts 137
  - Before Software Runs 138
  - Selecting Processors for Specific Tasks 141
  - What Is Out There? Topology Discovery 142
  - Intel® QPI Layer Setup 142
  - Link Speed Transition 142
  - Rest of BIOS 142
- Actions Prior to Boot Firmware Execution 143
  - Multiple Reset Domains 143
  - Boot Modes 145
  - Physical Layer Actions 146
  - Link Layer Initialization 146
  - Node IDs 147
  - PBSP Selection and CSR Access 149
  - Firmware Discovery 149
  - Handing Over Control 150
- Initial Boot Firmware Actions 150
  - System Bootstrap Processor (SBSP) Selection 151
  - Early Link Actions 151
  - Running Configuration Cycles 152
  - Link Layer Configuration 152
- Topology Discovery and Configuration 154
  - Example Topologies 154
  - Discovery Process 156
  - Communication Infrastructure Programming 157
- Full Speed Link Transition 159
- Higher Layer Initialization 160
  - Coherent Transaction Snooping Modes 160
  - Transaction Pool Allocations 161
  - Interrupt Configuration 162
  - Broadcast Lists 162
- Preparing for the Operating System 163
  - Processor Readiness 163
  - Memory Configuration 164

- Address Decoding 166
- Additional Functional and RAS Modes 170
- Hand-off to the Operating System 170
- Summary 171

## **Chapter 5    Advanced System Considerations 173**

- Dynamic Reconfigurations 174
- Partitioning 174
  - Static Partitioning 176
  - Dynamic Partitioning 176
- Quiesce and Dequiesce 177
- On Line Addition or Deletion (OL\_\*) 178
- Additional Memory Subsystem Features 178
  - DIMM Sparing 179
  - Memory Mirroring 183
  - Memory Migration 187
- Power Management 188
  - Link Power Management 188
  - Platform Power Management 195
  - PM Messages 197
- Fault Detection and Reporting 197
  - Error Reporting Methods 197
  - Fault Diagnosis 197
- Interrupts 200
  - Delivery Methods 202
  - Example Interrupt Flows 202
- System Management 203
  - Configuration Space Access 204
- Summary 205

## **Chapter 6    RAS and Dfx Features 207**

- CRC and Retries 207
  - Standard CRC Protection 208
  - Rolling CRC Protection 209
  - Link Layer Retry 212
- Hard Failure-Tolerant Links 216
  - Data Lane Failover 217
  - Clock Failover 218
- Vital Indication 219
- Data Poisoning 220
- Timeouts 221

- Hot Plug Capabilities 222
- Dynamic Reconfiguration 223
- Link Layer Dfx Hooks 224
  - Debug Packets 225
  - Credit Status and Defeathering 227
- Physical Layer Dfx Hooks 227
  - Compliance State 228
  - Loop Back 228
  - Freeze on Initialization Abort 231
  - State Machine Single Step 232
  - Latency Fixing 232
- Debug Tools and Examples 233
  - Probing 233
  - In System Link Margining 243
- Summary 246

