

InfiniBand™ Architecture Specification Volume 2

Release 1.3.1

November 2, 2016
Final Release

Copyright © 1999, 2001, 2002, 2003, 2004, 2005, 2006, 2011, 2012, 2013, 2014, 2015, 2016 by InfiniBandSM Trade Association. All rights reserved.

All trademarks and brands are the property of their respective owners.

Table 1 Revision History

Revision	Release Date	
1.0	2000-10-24	Released version
1.0.a	2001-6-19	Release 1.0 augmented with errata material. Updates only correct errors - no additional features have been added.
1.1	2002-11-06	Release 1.0.a augmented with additional features.
1.2	2004-10-1	Release 1.1 augmented with additional features, including Enhanced Signaling.
1.2.1	2006-10-3	Release 1.2 augmented with correction of errors - no additional features have been introduced
1.3	2012-11-6	<p>Release 1.2.1 augmented with additional features, including</p> <ul style="list-style-type: none"> - FDR & EDR signaling rates, 64b/66b data encoding, forward error correction, - electrical specifications for FDR interoperability, & - 4x and 12x (QSFP and CXP) pluggable managed cable connector interfaces. <p>Deprecated features have been appropriately labeled or moved into a separate document Volume 2-DEPR. Compliance statements have been moved into a separate document Volume 2-COMP.</p> <p>Annex A5, Pluggable Interfaces, 2007-3-6, has been incorporated into this document and is obsolete.</p> <p>Annex A6, CXP, 2009-9, has been incorporated into this document and is obsolete.</p> <p>Errata, 2009-12-18, has been incorporated into this document and is obsolete.</p>
1.3.1	2016-11-2	<p>Release 1.3 augmented with additional features, including</p> <ul style="list-style-type: none"> - electrical specifications for EDR interoperability - corrections to specifications for FDR interoperability and test methodologies - updates for QSFP28 and CXP28 memory map, mechanical and Tx/Rx Squelch specifications - updates for specifications for FEC support for FDR and EDR - description of improved testing methodology for EDR Limiting Active Cables

**LEGAL
DIS-
CLAIMER**

THIS VERSION OF THE IBTA SPECIFICATION IS PROVIDED “AS IS” AND WITHOUT ANY WARRANTY OF ANY KIND, INCLUDING, WITHOUT LIMITATION, ANY EXPRESS OR IMPLIED WARRANTY OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

IN NO EVENT SHALL IBTA OR ANY MEMBER OF IBTA BE LIABLE FOR ANY DIRECT, INDIRECT, SPECIAL, EXEMPLARY, PUNITIVE, OR CONSEQUENTIAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

TABLE OF CONTENTS

Chapter 1: Introduction	24	5
1.1 Document Conventions	25	6
1.1.1 Normative Terms.....	25	7
1.1.2 Informative Terms	25	8
1.1.3 Compliance Notation.....	26	9
1.1.4 Architecture Notes	26	10
1.1.5 Implementation Notes	26	11
1.1.6 Recommendation.....	26	12
1.2 References and Related Documents	27	13
1.3 Acknowledgments.....	30	14
1.3.1 Steering Committee	30	15
1.3.2 Technical Work Group.....	31	16
1.3.3 Volume Working Group.....	32	17
1.4 Disclaimer	36	18
Chapter 2: Glossary	37	19
2.1 A to C	37	20
2.2 D to H.....	39	21
2.3 I to L.....	43	22
2.4 M to P	47	23
2.5 Q to S.....	52	24
2.6 T to Z	55	25
Chapter 3: Physical Layer Overview	59	26
3.1 Introduction	59	27
3.2 Physical Port.....	60	28
3.2.1 Multi-porting	62	29
3.2.2 Active Cables	64	30
3.3 Link Electrical Signaling.....	64	31
3.4 Link Optical Signaling	64	32
3.5 Link Physical Layer.....	64	33
3.5.1 Speed Negotiation	65	34
3.5.2 Width Negotiation.....	65	35
3.6 Module Mechanical	66	36
3.7 Chassis Slot Mechanical	67	37
3.8 Power.....	69	38

3.9	Hardware Management	69	1
3.10	Operating System (OS) Power Management	70	2
Chapter 4: Port Signal Definitions		71	3
4.1	Signal Naming Conventions	71	4
4.2	Port Signal Summary	72	5
4.2.1	Backplane Port	73	6
4.2.2	Cable Port Using MicroGigaCN Connector.....	75	7
4.2.3	Fiber Optic Port.....	76	8
4.2.4	Active Cable Port Using MicroGigaCN Connector.....	77	9
4.2.5	Cable Port Using QSFP/QSFP+ Connector Interface	78	10
4.2.6	Cable Port Using CXP/CXP+ Connector Interface	78	11
4.3	Signaling Group	78	12
4.3.1	High Speed Electrical.....	78	13
4.3.2	High Speed Optical	78	14
4.4	Hardware Management Group	78	15
4.5	Bulk Power Group.....	79	16
4.6	Auxiliary Power Group	80	17
4.7	Active Cable Power Group	80	18
Chapter 5: Link/Phy Interface		81	19
5.1	Introduction	81	20
5.2	Symbol Encoding (8b/10b coding)	84	21
5.2.1	Notation conventions	84	22
5.2.2	Valid and invalid code-groups	86	23
5.2.3	Running disparity rules	93	24
5.2.4	Generating code-groups	96	25
5.2.5	Checking the validity of received code-groups	96	26
5.3	Block Encoding (64b/66b).....	97	27
5.3.1	Scrambling	98	28
5.3.2	Checking the validity of a received block	100	29
5.3.3	Block synchronization	101	30
5.3.4	Error Detection Per Lane	104	31
5.4	Forward Error Correction	106	32
5.4.1	Fire-Code FEC.....	107	33
5.4.2	Reed-Solomon Code FEC	114	34
5.5	Control Symbols/Blocks and Ordered-Sets	139	35
5.5.1	Control Symbols/Blocks	139	36
5.5.2	Control Ordered-sets 8b/10b encoding	142	37
5.5.3	Control Block 64b/66b Encoding.....	159	38
5.5.4	Logical Interfaces.....	164	39

5.6	Management Datagram Control and Status Interface.....	165	1
5.6.1	Control Inputs (MAD set)	165	2
5.6.2	Status Outputs (MAD get).....	169	3
5.6.3	Port Performance Counters	173	4
5.7	Packet Formats for Single and Multi Lane Support	175	5
5.7.1	Link Packet Ordering	176	6
5.7.2	Packet Formats.....	178	7
5.7.3	1x Packet Format.....	178	8
5.7.4	4x Packet Format.....	180	9
5.7.5	8x Packet Format.....	182	10
5.7.6	12x Packet Format.....	184	11
5.8	Link Initialization and Training.....	187	12
5.8.1	Link De-skew Training Sequence and SKIP ordered sets	188	13
5.8.2	Link Initialization and Training Options	191	14
5.8.3	Interactions With Other Entities	192	15
5.8.4	Link Training State Machine.....	193	16
5.8.5	State Machine Delays and Timeouts.....	225	17
5.8.6	Transmitter Interface and Behavior.....	226	18
5.8.7	Receiver Interface and Behavior.....	230	19
5.9	Link Physical Error Handling.....	235	20
5.9.1	Link Physical Errors Events	235	21
5.9.2	Minor Link Physical Errors Events	236	22
5.9.3	Link Physical Error Threshold Algorithm.....	237	23
5.9.4	Major Link Physical Errors Events	238	24
5.9.5	Heartbeat Error	238	25
5.10	Internal Serial Loopback	239	26
5.11	Clock Tolerance Compensation	239	27
5.11.1	Transmitter "SKIP" Requirements	240	28
5.11.2	Receiver "SKIP" Requirements.....	241	29
5.12	Retiming Repeaters	241	30
5.12.1	Retiming Repeater Functions	242	31
5.12.2	Clock Tolerance Compensation	243	32
5.12.3	Error Handling Considerations.....	244	33
5.12.4	Symbol Boundary Alignment.....	244	34
5.12.5	Multi-lane Repeater Considerations.....	245	35
5.12.6	Power State Considerations.....	245	36
5.13	Fine Tuning	246	37
5.13.1	Fine Tuning Block	246	38
5.13.2	Fine Tuning Rules.....	247	39
5.13.3	Fine Tuning State Machine	248	40
5.13.4	Fine Tuning Operation	250	41

5.14	Link Heartbeat	253	1
5.14.1	Operation of Link Heartbeats	253	2
5.14.2	Heartbeat Error Handling	254	3
5.14.3	Heartbeat Latency Calculation and Reporting	254	4
5.15	Max Packet Rate	255	5
5.16	FDR and Higher Rate Transmitter Equalization	256	6
5.17	Physical Layer Compliance Testing	256	7
5.17.1	Compliance Testing Overview	256	8
5.17.2	Compliance Testing Facilities	257	9
5.17.3	Example Compliance Test procedure	263	10
Chapter 6:	High Speed Electrical Interfaces	265	11
6.1	Introduction	265	12
6.1.1	Background and Reference Material	265	13
6.2	Electrical Topologies	266	15
6.2.1	Repeaters	268	16
6.2.2	Cable Types	270	17
6.3	General Requirements	273	18
6.3.1	Electrostatic Discharge (ESD)	273	19
6.4	Compliance Facilities	274	20
6.4.1	Compliance Points	274	21
6.4.2	Compliance Testing Hardware and Methodology	277	22
6.4.3	Link/Phy Compliance Provisions	278	23
6.5	Equalization Methodology	279	24
6.5.1	Equalization for InfiniBand Release 1.1 Devices	280	25
6.5.2	Equalization for InfiniBand Release 1.2.1 Devices	280	26
6.5.3	Equalization for InfiniBand Release 1.3 Devices	282	27
6.6	Differential Driver Outputs	282	28
6.6.1	General Requirements	282	29
6.6.2	Host Driver Output Characteristics for SDR	285	30
6.6.3	Host Driver Output Characteristics for DDR	287	31
6.6.4	Host Driver Output Characteristics for QDR	290	32
6.6.5	Host Driver Output Characteristics for FDR	293	33
6.6.6	Host Driver Output Characteristics for EDR	302	34
6.7	Differential Receiver Inputs	309	35
6.7.1	General requirements	309	36
6.7.2	Beacon Signaling	309	37
6.7.3	Signaling Cautions	310	38
6.7.4	Termination	310	39
6.7.5	Host Receiver input characteristics for SDR	311	40
6.7.6	Host receiver input characteristics for DDR	313	41
			42

6.7.7	Host Receiver input characteristics for QDR	315	1
6.7.8	Host Receiver input characteristics for FDR.....	317	2
6.7.9	Host Receiver input characteristics for EDR.....	320	3
6.8	Compliant Channels	323	4
6.8.1	DC Blocking	323	5
6.8.2	Linear Cable Electrical Requirements.....	324	6
6.8.3	Lane-to-lane Skew	325	7
6.8.4	SDR (2.5 Gb/s) compliant channel	326	8
6.8.5	DDR (5.0 Gb/s) linear channels and Limiting Active Cables.....	331	9
6.8.6	QDR (10.0 Gb/s) linear channels and Limiting Active Cables	334	10
6.8.7	FDR (14.0625 Gb/s) Linear Channels and Limiting Active Cables	337	11
6.8.8	EDR (25.78125 Gb/s) Linear Channels and Limiting Active Cables.....	344	12
Chapter 7:	Electrical Connectors for Modules and Cables	351	13
7.1	Introduction	351	14
7.1.1	Environmental Performance Requirements	351	15
7.1.2	Port and connector labeling	351	16
7.2	1X Interface	355	17
7.2.1	1X Board Connector	355	18
7.3	microGigaCN Interface	355	19
7.3.1	Mechanical requirements.....	355	20
7.3.2	4X Introduction.....	357	21
7.3.3	Electrical requirements	365	22
7.3.4	Environmental requirements	370	23
7.4	Pluggable Interface Connectors	370	24
7.4.1	Introduction	370	25
7.4.2	Obsolete Pluggable Interface Connectors	371	26
7.5	4X QSFP+ Interface connectors	373	27
7.5.1	Introduction	373	28
7.5.2	Mechanical requirements.....	375	29
7.5.3	Electrical Requirements	390	30
7.5.4	Environmental and thermal requirements	398	31
7.5.5	Memory Map.....	399	32
7.6	Optical 4x QSFP+ Modules	399	33
7.7	8X and 12XmicroGigaCN Interface	399	34
7.7.1	Introduction	399	35
7.7.2	Mechanical requirements.....	399	36
7.7.3	Electrical requirements	403	37
7.7.4	Environmental requirements	413	38
7.8	CXP Interface	413	39
7.8.1	Introduction	413	40
			41
			42

7.8.2	Mechanical requirements.....	415	1
7.8.3	Electrical Requirements.....	432	2
7.8.4	Environmental and thermal requirements	438	3
7.8.5	Thermal Performance Ranges.....	438	4
7.8.6	CXP Housing Assembly Thermal Interfaces.....	439	5
7.8.7	Mating of CXP Module & Host PCBs to CXP Electrical Connectors.....	440	6
7.8.8	CXP Memory Map.....	445	7
7.9	Electrical interface cable assemblies	445	8
7.9.1	Physical Requirements	445	9
7.9.2	Cable Shielding.....	448	10
7.9.3	1X Interface cable	449	11
7.9.4	MicroGigaCN interface cables	449	12
7.9.5	QSFP+ interface cables.....	465	13
7.9.6	CXP interface cables	472	14
Chapter 8:	Management Interface	480	15
8.1	Introduction	480	16
8.2	Voltage and Timing Specification	480	17
8.2.1	Management interface Voltage Specification.....	480	19
8.2.2	Management Interface Timing Specification	481	20
8.3	Memory Interaction Specifications	483	21
8.3.1	Timing for Memory Transactions.....	483	22
8.3.2	Timing for Control and Status Functions	483	23
8.3.3	Timing for Squelch and Disable Functions	485	24
8.4	Device Addressing and Operation	485	25
8.5	QSFP+ Memory Map	488	26
8.5.1	QSFP+ Memory Map - Lower Page.....	489	27
8.5.2	QSFP+ Memory Map - Upper Page 00.....	499	28
8.5.3	QSFP+ Memory Map - Upper Page 02.....	509	29
8.5.4	QSFP+ Memory Map - Upper Page 03.....	509	30
8.6	Read/Write Functionality for CXP	515	31
8.6.1	CXP Memory Address Counter (Read AND Write Operations).....	515	32
8.6.2	Read Operations (Current Address Read).....	515	33
8.6.3	Read Operations (Random Read).....	515	34
8.6.4	Read Operations (Sequential Read).....	516	35
8.6.5	Write Operations (Byte Write)	517	36
8.6.6	Write Operations (Sequential Write)	518	37
8.6.7	Write Operations (Acknowledge Polling)	519	38
8.7	CXP Memory Map	520	39
8.7.1	CXP Memory Map - TX Lower Page.....	522	40
8.7.2	CXP Memory Map - Rx Lower Page.....	529	41
			42

8.7.3	CXP Memory Map - Tx & Rx Common Upper Page 00h.....	536	1
8.7.4	CXP Memory Map - Tx Upper Page 01h	544	2
8.7.5	CXP Memory Map - Rx Upper Page 01h.....	545	3
8.7.6	CXP Memory Map - Tx and/or Rx Upper Page 02h.....	547	4
Chapter 9:	Fiber Attachment - 2.5 Gb/s, 5.0 Gb/s, & 10 Gb/s	548	5
9.1	Introduction	548	6
9.2	Scope.....	548	7
9.3	Fiber Attachment Technology Options	548	8
9.4	Fiber Attachment Overview.....	551	9
9.4.1	Fiber Optic System Overview	551	10
9.4.2	1x System Overview - SDR, DDR & QDR	552	11
9.4.3	4x System Overview	553	12
9.4.4	8x-SX Overview - SDR & DDR	557	13
9.4.5	12x System Overview - SDR, DDR, & QDR	559	14
9.5	Optical Specifications.....	560	15
9.5.1	Quiescent Condition.....	561	16
9.5.2	Optical Signal Polarity.....	562	17
9.5.3	Optical Transmitter Mask Compliance for links operating at 2.5 & 5.0 Gb/s.....	562	18
9.5.4	Optical Transmitter Mask Compliance for links operating at 10.0 Gb/s	565	19
9.5.5	Optical Jitter Specification for links operating at 2.5 Gb/s and 5.0 Gb/s	566	20
9.5.6	Optical Jitter Specifications for links operating at 10.0 Gb/s	570	21
9.5.7	Bit to Bit Skew.....	571	22
9.5.8	1x SDR Links - at 2.5 Gb/s	572	23
9.5.9	1x DDR Links - at 5.0 Gb/s	578	24
9.5.10	4x SDR Links - at 2.5 Gb/s	582	25
9.5.11	4x-DDR-SX Link.....	585	26
9.5.12	8x-SX Links - at 2.5 Gb/s.....	587	27
9.5.13	12x-SX Links - at 2.5 Gb/s.....	587	28
9.5.14	8x-DDR-SX and 12x-DDR-SX Links.....	590	29
9.5.15	1x QDR Links.....	592	30
9.6	Optical Receptacle and Connector	593	31
9.6.1	1x Connector - LC.....	594	32
9.6.2	4x-SX Connector - Single MPO	596	33
9.6.3	4x LX Connector - SC	598	34
9.6.4	8x-SX Optical Receptacle and Connector	599	35
9.6.5	12x-SX Connector - Dual MPO.....	600	36
9.7	Fiber Optic Cable Plant Specifications.....	602	37
9.7.1	Optical Fiber Specification	602	38
9.7.2	Modal Bandwidth	603	39
9.7.3	Optical Passive Loss of Fiber Optic Cable.....	604	40

9.7.4	Fiber Optic Adapters and Splices	605	1
9.8	Signal Conditioner in Optical Transceiver.....	606	2
9.8.1	Motivation for Signal Conditioner	606	3
9.8.2	Signal Conditioner Implementation.....	607	4
9.9	Aux Power	612	5
9.9.1	Behavior in Aux Power Mode.....	612	6
9.9.2	Beaconing and Wake-up.....	612	7
9.10	Optical Pluggable Modules	612	8
9.10.1	1x Optical Pluggable Modules	612	9
9.10.2	4x Optical Pluggable Modules	612	10
9.10.3	8x and 12x Optical Pluggable Modules.....	612	11
Annex A1:	FDR and EDR Compliance Boards and Test Setups	614	12
A1.1	Background.....	614	13
A1.1.1	Compliance board design	614	14
A1.1.2	Compliance board calibration	614	15
A1.2	Test Setups	614	16
A1.2.1	Transmitter characterization	614	17
A1.2.2	Receiver characterization	615	18
A1.2.3	Cable characterization	617	19
A1.3	Compliance Boards - Electrical Specifications.....	618	20
A1.3.1	Host Compliance Boards	620	21
A1.3.2	Module Compliance Boards.....	621	22
A1.3.3	Combined Host & Module Compliance Boards.....	622	23
A1.4	QSFP28 compliance boards - Mechanical Description.....	625	24
A1.4.1	QSFP28 Host Compliance board.....	625	25
A1.4.2	QSFP28 Module Compliance board	626	26
A1.5	CXP compliance boards - Mechanical Description.....	627	27
A1.5.1	CXP Host Compliance board.....	627	28
A1.5.2	CXP Module Compliance board.....	629	29
			30
			31
			32
			33
			34
			35
			36
			37
			38
			39
			40
			41
			42

Annex A2: Cable Electrical parameters for FDR and EDR	630	1
A2.1 Insertion loss fitting	630	2
A2.2 Integrated Crosstalk Noise (ICN).....	631	3
A2.3 Integrated Common Mode Conversion Noise (ICMCN).....	633	4
A2.4 FDR Overall Link Budget for Linear Channels (Informative).....	634	5
A2.5 EDR Overall Link Budget for Linear Channels (Informative)	635	6
Annex A3: Management Interface Modifications for RoCE Support	636	7
A3.1 ROCE Management Interface - Overview	636	8
A3.2 InfiniBand vs. Ethernet Memory Map Differences - QSFP/QSFP+.....	636	9
A3.2.1 InfiniBand vs. Ethernet Memory Map Differences - Lower Page	636	10
A3.2.2 InfiniBand vs. Ethernet Memory Map Differences - Upper Page 00	636	11
A3.2.3 InfiniBand vs. Ethernet Memory Map Differences - Upper Page 03	637	12
A3.3.1 InfiniBand vs. Ethernet Memory Map Differences - CXP/CXP+	637	13
A3.3.2 InfiniBand vs. Ethernet Memory Map Differences - Tx & Rx Lower Page	637	14
A3.3.3 InfiniBand vs. Ethernet Memory Map Differences - Tx & Rx Upper Page 00	638	15
		16
		17
		18
		19
		20
		21
		22
		23
		24
		25
		26
		27
		28
		29
		30
		31
		32
		33
		34
		35
		36
		37
		38
		39
		40
		41
		42

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

IBTA

LIST OF FIGURES

FIGURE	TITLE	PAGE
Figure 1	Physical Layer Structure	59
Figure 2	Backplane Port - Single Physical Link	61
Figure 3	Cable Port	62
Figure 4	Fiber Optic Port	62
Figure 5	Backplane Port - Multiple Physical Links	63
Figure 6	Module Connector Location Designation	67
Figure 7	Chassis Connector Location Designation	68
Figure 8	8b/10b Example 4x Byte Striping Diagram	81
Figure 9	64b/66b Example 4x Byte Striping Diagram	82
Figure 10	64b/66b Example 4x Byte Striping Diagram - with RS-FEC Enabled	83
Figure 11	Link/Physical Interface Block Diagram	84
Figure 12	Transmit & Receive Data Ordering	86
Figure 13	Scrambler	98
Figure 14	De-scrambler	98
Figure 15	Scrambler Byte Stream Format	99
Figure 16	De-Scrambler Byte Stream Format	99
Figure 17	Scrambling order when RS-FEC is enabled	100
Figure 18	Block lock Finite StateMachine	102
Figure 19	EDPL calculator	104
Figure 20	EDPL block diagram	105
Figure 21	Fire-Code FEC layer block diagram	108
Figure 22	PN-2112 Scrambler	110
Figure 23	Fire-Code FEC Tx block diagram	111
Figure 24	Fire-Code FEC Rx block diagram	111
Figure 25	Fire-Code FEC block lock FSM	112
Figure 26	RS-FEC Functional Block Diagram	115
Figure 27	Transcoding Examples	118
Figure 28	Alignment Transcoding Block for x4 Link Width	121
Figure 29	Alignment Sequence After Lane Distribution for x4 Link Width	121
Figure 30	Alignment Transcoding Block for x8 Link Width	122
Figure 31	Alignment Transcoding Block for x12 Link Width	122
Figure 32	Reed-Solomon Encoder Functional Model	124
Figure 33	Symbol Distribution RS(528,514) for a 4x Link	129
Figure 34	Symbol Distribution RS(271,257) for a 4x Link	130
Figure 35	Transmit Bit Ordering	131
Figure 36	Lane Alignment Lock FSM	134
Figure 37	RS-FEC Codeword Lock FSM	136
Figure 38	Ordered-Sets	142
Figure 39	TS3 Ordered-Set: Detailed Format - Rev 0	144
Figure 40	TS3 Ordered-Set: Detailed Format - Rev 1	145

Figure 41	TS3 Ordered-Set: Detailed Format - Rev 2	146	1
Figure 42	HRTBT Ordered-Set: Detailed Format	155	2
Figure 43	TS-T Ordered-Set: Detailed Format	156	3
Figure 44	Packet Ordering Example 8b/10b	176	4
Figure 45	Packet Ordering Example 64b/66b	177	5
Figure 46	1x Packet Formats - 8b/10b	179	6
Figure 47	1x Packet Formats - 64b/66b	180	7
Figure 48	4x Packet Formats - 8b/10b	181	8
Figure 49	4x Packet Formats - 64b/66b	182	9
Figure 50	8x Packet Formats - 8b/10b	183	10
Figure 51	8x Packet Formats - 64b/66b	184	11
Figure 52	12x Packet Formats - 8b/10b	185	12
Figure 53	12x Packet Formats - 64b/66b	186	13
Figure 54	Link Training State Machine Interactions	192	14
Figure 55	Link Training State Machine - Legacy	195	15
Figure 56	Link Training State Machine - Enhanced Signaling	196	16
Figure 57	Polling Super State (Expanded)	197	17
Figure 58	Sleeping Super State (Expanded)	199	18
Figure 59	Configuration Super State (Expanded) - Legacy	202	19
Figure 60	Configuration Super State (Expanded) - Enhanced Signaling	208	20
Figure 61	Typical Control Flow through Configuration Super State - Enhanced Signaling	209	21
Figure 62	PRBS23 Pattern Generator	217	22
Figure 63	PRBS23 checker	218	23
Figure 64	PRBS11 Pattern Generator	218	24
Figure 65	PRBS11 checker	219	25
Figure 66	Recovery Super State (Expanded)	223	26
Figure 67	A Conceptual Retiming Repeater Block Diagram	243	27
Figure 68	Fine Tuning Flow Diagram	249	28
Figure 69	Responder State Machine	249	29
Figure 70	Initiator State Machine	250	30
Figure 71	PRBS31 Generator	261	31
Figure 72	PRBS31 checker	262	32
Figure 73	PRBS9 Generator	262	33
Figure 74	High-level topology block diagram	267	34
Figure 75	Optical Fiber Interconnect Topology	268	35
Figure 76	Full limiting active cable topology	272	36
Figure 77	Near-end limiting active cable topology	272	37
Figure 78	Far-end limiting active cable topology	272	38
Figure 79	Board/Backplane test points	275	39
Figure 80	Board/Cable test points	275	40
Figure 81	Host compliance board test points	275	41
Figure 82	Module compliance board test points, cable	276	42
Figure 83	Module compliance board test points, pluggable module	276	
Figure 84	Mated Module and Host Compliance board test points	276	
Figure 85	Waveform with De-emphasis	279	
Figure 86	Diamond-shaped eye mask	284	
Figure 87	Hexagonal eye mask	284	
Figure 88	Limits on SDD11 and SDD22 vs. Frequency for FDR Hosts and Cables	295	

Figure 89	Limits on SDC11 and SDC22 vs. Frequency for FDR Hosts and Cables	295	1
Figure 90	Tx Equalization FIR Representation.....	296	2
Figure 91	Limits on SDD11 and SDD22 vs. Frequency for EDR Hosts and Cables	304	3
Figure 92	Limits on SDC11 and SDC22 vs. Frequency for EDR Hosts and Cables	304	4
Figure 93	Termination and Signaling.....	310	5
Figure 94	Eye Opening at receiver for SDR, DDR, & QDR signaling (Differential)	312	6
Figure 95	I/O Plate to I/O Plate via Cable Topology.....	327	7
Figure 96	Eye Opening at receiving board connector pins, SDR rate (differential).....	329	8
Figure 97	Limits on SDD11 and SDD22 vs. Frequency for DDR Active Cables	333	9
Figure 98	Limits on SDD11 and SDD22 vs. Frequency for QDR Active Cables	336	10
Figure 99	Insertion Loss Deviation (ILD) for FDR linear cables	339	11
Figure 100	ICN vs. IL for FDR linear cables	340	12
Figure 101	Insertion Loss Deviation (ILD) for EDR linear cables	346	13
Figure 102	ICN vs. IL for EDR linear cables	347	14
Figure 103	Icons for SDR cables and ports.....	352	15
Figure 104	Icons for DDR cables and ports	353	16
Figure 105	Icons for QDR cables and ports	354	17
Figure 106	Icons for 4x and 12x FDR cables and ports	354	18
Figure 107	4X cable board connector	358	19
Figure 108	4X cable board connector footprint, top view	359	20
Figure 109	4X/12X cable board connector interface details	360	21
Figure 110	4X cable plug.....	363	22
Figure 111	4X/12X cable plug interface details	364	23
Figure 112	4X microGigaCN board and cable connector keying	365	24
Figure 113	Recommended active cable host board power supply filtering	368	25
Figure 114	QSFP Conceptual Model.....	373	26
Figure 115	QSFP+ pluggable modules and host board receptacle cages	374	27
Figure 116	QSFP+ Module Datum Definitions	376	28
Figure 117	QSFP+ module and host board connector contact assignment	377	29
Figure 118	QSFP+ host board connector.....	378	30
Figure 119	QSFP+ pluggable module cage dimensions	379	31
Figure 120	Sample QSFP+ host board connector footprint (Style A).....	381	32
Figure 121	Sample QSFP+ host board connector footprint (Style B).....	382	33
Figure 122	QSFP+ heat sink dimensions (part 1 of 2)	383	34
Figure 123	QSFP+ heat sink dimensions (part 2 of 2)	384	35
Figure 124	QSFP+ cage to bezel dimensions	385	36
Figure 125	QSFP+ bezel opening dimensions	386	37
Figure 126	QSFP+ module dimensions (basic).....	387	38
Figure 127	QSFP+ module dimensions (detail).....	388	39
Figure 128	QSFP+ paddle card dimensions.....	389	40
Figure 129	Example QSFP Host board schematic.....	396	41
Figure 130	Recommended QSFP Host Board Power Supply Filtering	397	42
Figure 131	12X cable board connector	400	
Figure 132	12X cable board connector footprint, top view	401	
Figure 133	12X cable plug.....	402	
Figure 134	12X board and cable connector keying	403	
Figure 135	CXP Conceptual Model	414	
Figure 136	CXP Cable Plug/module and Receptacle/Cage (without heat sink).....	415	

Figure 137	CXP interface datum definitions	417	1
Figure 138	CXP module and host board connector pin assignments	418	1
Figure 139	CXP host board connector footprint	421	2
Figure 140	CXP host board connector dimensions (part 1 of 2)	422	3
Figure 141	CXP host board connector dimensions (part 2 of 2)	423	4
Figure 142	CXP module and host board cage heat sink dimensions	424	5
Figure 143	CXP host board cage riding heat sink dimensions	425	6
Figure 144	CXP cage to bezel dimensions	426	7
Figure 145	CXP bezel opening dimensions	426	8
Figure 146	CXP module dimensions (basic)	428	9
Figure 147	CXP module dimensions (detail)	429	9
Figure 148	CXP module paddle card dimensions	430	10
Figure 149	Exemplary CXP module latch release mechanism	431	11
Figure 150	Receptacle Housing with Integrated Riding Heat Sink	440	12
Figure 151	Example CXP host board schematic	441	13
Figure 152	Recommended CXP Host Board Power Supply Filtering	443	14
Figure 153	microGigaCN cable assembly bend radius	446	14
Figure 154	QSFP+ cable assembly bend radius	447	15
Figure 155	CXP cable assembly bend radius	448	16
Figure 156	8x port using two 4X Pluggable devices	469	17
Figure 157	Example of incorrect cabling of an 8X port using two 4X Pluggable devices	470	18
Figure 158	12x port using three 4X Pluggable devices	471	19
Figure 159	Active Optical Cable	473	20
Figure 160	Passive or Active Copper Cable	473	21
Figure 161	12x to 3-4x cables	476	21
Figure 162	Management interface two wire serial interface timing diagram	482	22
Figure 163	QSFP+ Memory Map	488	23
Figure 164	Read Operation on Current Address	515	24
Figure 165	Random Read	516	25
Figure 166	Sequential Address Read Starting at Current Address	516	26
Figure 167	Sequential Address Read Starting with Random CXP Read	517	27
Figure 168	Write Byte Operation	518	28
Figure 169	Sequential Write Operation	519	29
Figure 170	Memory map two wire serial addresses A0x (Tx) & A8x (Rx)	521	29
Figure 171	1x Optical Link Overview	553	30
Figure 172	4x-SX Optical Link Overview	554	31
Figure 173	4x-LX Optical Link Overview	555	32
Figure 174	4x LX Transmitter Serialization	556	33
Figure 175	4x LX receiver de-serialization	557	34
Figure 176	8x-SX Optical Link Overview	558	35
Figure 177	12x-SX Optical Link Overview	560	35
Figure 178	Normalized Optical Transmitter Compliance Mask	564	36
Figure 180	Jitter Compliance Test Points	567	37
Figure 181	Jitter Tolerance Mask	569	38
Figure 182	1x-LX Trade-off between RMS Spectral Width and Center Wavelength	577	39
Figure 183	1x-DDR-LX Trade-off between OMA, RMS Spectral Width, and Center Wavelength ..	582	40
Figure 184	Dual LC Plug	594	41
Figure 185	1x-SX and 1x-LX Optical Receptacle orientation looking into the Optical Transceiver	595	42

Figure 186	MPO Plug and Receptacle	597	1
Figure 187	4x-SX Optical Receptacle orientation looking into the Transceiver.....	598	2
Figure 188	4x-LX Optical Receptacle orientation	599	3
Figure 189	8x-SX Optical Receptacle orientation looking into the Transceiver.....	600	4
Figure 190	Double MPO Optical Receptacle Configuration	601	5
Figure 191	12x-SX and 8x-SX Optical Receptacle orientation looking into the Transceiver.....	602	6
Figure 192	Recommended 1x, 1x DDR, and 1x QDR Optical Link Implementation	608	7
Figure 193	Recommended 4x-SX & 4x-DDR-SX Optical Link Implementation.....	608	8
Figure 194	Recommended 4x-LX Optical Link Implementation	609	9
Figure 195	Recommended 8x-SX and 8x-DDR-SX Optical Link Implementation	610	10
Figure 196	Recommended 12x-SX & 12x-DDR-SX Optical Link Implementation.....	611	11
Figure 197	Host transmitter output characterization setup using HCB.....	615	12
Figure 198	Host receiver input characterization setup	616	13
Figure 199	Cable characterization setup using MCB	617	14
Figure 200	HCB trace pair reference through response.....	620	15
Figure 201	MCB trace pair reference through response	621	16
Figure 202	Mated HCB/MCB trace pair through response.....	622	17
Figure 203	Limits on Return Loss for QSFP+ Host Compliance Boards.....	623	18
Figure 204	Limits on HCB/MCB common-mode/differential return loss	623	19
Figure 205	Limits on HCB/MCB differential to common-mode conversion	624	20
Figure 206	QSFP28 Host Compliance Board layout	625	21
Figure 207	QSFP28 Module Compliance Board layout.....	626	22
Figure 208	CXP Host Compliance Board layout, side A	627	23
Figure 209	CXP Host Compliance Board layout, side B	628	24
Figure 210	CXP Module Compliance Board layout	629	25
Figure 211	FDR Overall Link Budget (Informative)	634	26
Figure 212	EDR Overall Link Budget (Informative)	635	27
			28
			29
			30
			31
			32
			33
			34
			35
			36
			37
			38
			39
			40
			41
			42

IBTA

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

LIST OF TABLES

TABLE	TITLE	PAGE
Table 1	Revision History	2
Table 2	Symbol Substitution	7
Table 3	Type Notation.....	8
Table 4	Backplane Port Signal Summary	9
Table 5	Cable Port Signal Summary.....	10
Table 6	Fiber Optic Port Signal Summary	11
Table 7	Active Cable Port Signal Summary.....	12
Table 8	Valid Data Code Groups	13
Table 9	Valid Special Code Groups	14
Table 10	5b/6b Coding for Data Characters	15
Table 11	3b/4b Coding for Data Characters	16
Table 12	5b/6b Coding for Special Characters	17
Table 13	3b/4b Coding for Special Characters	18
Table 14	64b/66b Block Format.....	19
Table 15	Fire-Code FEC Format	20
Table 16	Scrambled Lane Stream	21
Table 17	FEC (2112, 2080).....	22
Table 18	Example of PN2112 Scrambler Output.....	23
Table 19	Coefficients of the Generator Polynomial (Decimal)	24
Table 20	Transcoder Function Input 80 66-Bit Blocks	25
Table 21	RS-FEC Encoder Output for RS(528,514).....	26
Table 22	RS-FEC Encoder Output for RS(271,257).....	27
Table 23	Link Control Symbols	28
Table 24	Lane Identifiers	29
Table 25	DDSV, ADD and MOD Encoding	30
Table 26	TS3 Rev1 and 2 Variables	31
Table 27	64b/66b Control Blocks.....	32
Table 28	64b/66b SDP Block.....	33
Table 29	64b/66b SLP Block	34
Table 30	64b/66b EGP3 Block	35
Table 31	64b/66b EGP7 Block	36
Table 32	64b/66b EBP3 Block	37
Table 33	64b/66b EBP7 Block	38

Table 34	64b/66b Idle Block	162	1
Table 35	64b/66b Skip Block	162	2
Table 36	64b/66b TS1 Ordered-Set.....	162	3
Table 37	64b/66b TS2 Ordered-Set.....	163	4
Table 38	Fine Tuning Block	163	5
Table 39	64b/66b Heartbeat Ordered-Set	163	6
Table 40	Vendor-specific Block.....	164	7
Table 41	Test Time Values.....	191	8
Table 42	Transmitter and Receiver Interface.....	193	9
Table 43	PRBS11 polynomials when Different Polynomials are used for Each Lane	219	10
Table 44	PRBS23 Polynomials when Different Polynomials are used for Each Lane	219	11
Table 45	InfiniBand Link Data Rates	265	12
Table 46	InfiniBand Signal Test Points	274	13
Table 47	Equalization responsibilities.....	281	14
Table 48	Host Driver Characteristics for 2.5 Gb/s (SDR)	285	15
Table 49	Host Driver Characteristics for 5.0 Gb/s (DDR)	287	16
Table 50	Host Driver Characteristics for driving 5.0 Gb/s (DDR) limiting cables	289	17
Table 51	Host Driver Characteristics for 10 Gb/s (QDR).....	290	18
Table 52	Host Driver Characteristics for driving 10.0 Gb/s (QDR) limiting cables.....	292	19
Table 53	Host Driver Characteristics for 14.0625 Gb/s (FDR).....	294	20
Table 54	Tx FIR Filter Coefficients and Amplitudes for 14.0625 Gb/s (FDR)	299	21
Table 55	FDR host output specifications at Preset 0, for Limiting Active Cables	301	22
Table 56	Host transmitter output characteristics for 25.78125 Gb/s (EDR).....	303	23
Table 57	Tx FIR Filter Coefficients and Amplitudes for 25.78125 Gb/s (EDR).....	306	24
Table 58	EDR host output specifications at Preset 0, for Limiting Active Cables	308	25
Table 59	Host Receiver Characteristics for 2.5 Gb/s (SDR).....	311	26
Table 60	Host Receiver Characteristics for 5.0 Gb/s (DDR).....	313	27
Table 61	Host Receiver Characteristics for 10 Gb/s (QDR)	315	28
Table 62	Host Receiver Characteristics for 5.0 Gb/s and 10 Gb/s limiting cables.....	316	29
Table 63	Host Receiver characteristics for 14.0625 Gb/s (FDR).....	317	30
Table 64	FDR stressed input signal specifications - linear cables.....	318	31
Table 65	FDR receiver stressed input signal specifications - limiting cables.....	319	32
Table 66	Host Receiver characteristics for 25.78125 Gb/s (EDR).....	320	33
Table 67	EDR stressed input signal specifications - linear cables.....	321	34
Table 68	EDR receiver stressed input signal specifications - limiting cables	322	35
Table 69	Cable Assembly Electrical Requirements	324	36
Table 70	Lane-to-lane skew maximum values.....	325	37
Table 71	Lane-to-lane skew maximum values for separable fiber optic cables.....	326	38
Table 72	SDR Loss Parameter Maximum Values.....	327	39

Table 73	InfiniBand SDR Signal eye opening.....	328	1
Table 74	SDR Cable Assembly Electrical Requirements	329	2
Table 75	Linear Channel S Parameter Requirements for 5.0 Gb/s (DDR).....	331	3
Table 76	SERDES to SERDES pin S Parameters for 5.0 Gb/s (Informative).....	331	4
Table 77	DDR active cable input electrical specifications.....	332	5
Table 78	DDR active cable output electrical specifications	332	6
Table 79	Linear Channel S Parameter Requirements for 10 Gb/s (QDR).....	334	7
Table 80	SERDES pin to SERDES pin S Parameters for 10 Gb/s (QDR) (Informative) ..	335	8
Table 81	QDR active cable input electrical specifications	335	9
Table 82	QDR active cable output electrical specifications	336	10
Table 83	FDR compliant linear cable specifications	338	11
Table 84	FDR limiting active cable input electrical specifications.....	341	12
Table 85	FDR limiting active cable output electrical specifications.....	342	13
Table 86	EDR compliant linear cable specifications	345	14
Table 87	EDR limiting active cable input electrical specifications.....	348	15
Table 88	EDR limiting active cable output electrical specifications	349	16
Table 89	microGigaCN cable connector physical requirements	356	18
Table 90	Recommended microGigaCN interface connector physical requirements	357	19
Table 91	4X passive board connector signal assignment.....	361	20
Table 92	4X active board connector signal assignment	362	21
Table 93	MicroGigaCN Cable connector electrical performance requirements	365	22
Table 94	Power supply specifications	368	23
Table 95	QSFP+ connector physical requirements	375	24
Table 96	Recommended QSFP+ connector physical parameters.....	375	25
Table 97	Contact Assignment for 4x QSFP+ Interface	380	26
Table 98	QSFP/QSFP+ connector electrical performance requirements	390	27
Table 99	Power Mode Truth Table	392	28
Table 100	Power Supply Specification	397	29
Table 101	Power Budget Classification	398	30
Table 102	12X passive board connector signal assignment.....	404	31
Table 103	12X active board connector signal assignment	406	32
Table 104	12X board connector signal assignment for 12X to 3-4X cables	408	33
Table 105	8X passive board connector signal assignment.....	410	34
Table 106	8X active board connector signal assignment	412	35
Table 107	CXP connector physical requirements.....	416	36
Table 108	Recommended CXP connector physical parameters	416	37
Table 109	Contact Assignments for 12x Pluggable-CXP Interface	419	38
Table 110	CXP Module & Receptacle Connector Electrical Performance Requirements ..	432	39
Table 111	CXP Board Connector Signal Assignment.....	433	40
			41
			42

Table 112	12X Board Connector Signal Assignment for 12X CXP to 3-4X QSFP Cables	.434	1
Table 113	Module and Receptacle Mechanical and Environmental Requirements.....	439	2
Table 114	Temperature Classification of Module Case	439	3
Table 115	Power Supply Specification	444	4
Table 116	Power Budget Classification	444	5
Table 117	QSFP+ bend radius limits	447	6
Table 118	4X passive cable connector signal assignment	450	7
Table 119	4X active cable connector signal assignment	451	8
Table 120	8X passive cable connector signal assignment	452	9
Table 121	8X active cable connector signal assignment	455	10
Table 122	12X passive cable connector signal assignment	457	11
Table 123	12X active cable connector signal assignment	460	12
Table 124	12X to 3-4X cable connector signal assignment.....	463	13
Table 125	QSFP cable connector signal assignment	465	14
Table 126	Signal connections between 4x-QSFP+ and 4x microGigaCN connectors	467	15
Table 127	Signal connections between 4x-QSFP+ and 4x microGigaCN connectors	468	16
Table 128	12X CXP to CXP cable signal assignment	475	17
Table 129	12X CXP to 3-4X microGigaCN passive cable signal assignment	477	19
Table 130	12X CXP to 3-4X microGigaCN active cable signal assignment	478	20
Table 131	12X CXP to 3-4X QSFP cable signal assignment	479	21
Table 132	Low Speed Control and Sense Signal Specifications	481	22
Table 133	Management interface two wire serial interface timing specifications	482	23
Table 134	Management interface memory transaction timing specification	483	24
Table 135	I/O Timing for Control and Status Functions	483	25
Table 136	I/O Timing for Squelch and Disable	485	26
Table 137	CXP Device Addresses.....	487	27
Table 138	QSFP+ lower page memory map	489	28
Table 139	QSFP+ upper page 00 memory map	499	29
Table 140	QSFP+ upper page 02 memory map	509	30
Table 141	QSFP+ upper page 03 memory map	510	31
Table 142	CXP Tx Lower Page Memory Map	522	32
Table 143	CXP Rx Lower Page Memory Map (Optional)	529	33
Table 144	CXP Tx & Rx Upper Page 00h Memory Map.....	536	34
Table 145	CXP Tx Upper Page 01h Memory Map	544	35
Table 146	CXP Rx Upper Page 01h Memory Map	545	36
Table 147	CXP Rx Upper Page 02h Memory Map	547	37
Table 148	Fiber Optic Attachment Option for SDR (2.5 Gb/s).....	549	38
Table 149	Fiber Optic Attachment Options for DDR (5.0 Gb/s).....	550	39
Table 150	Fiber Optic Attachment Options for QDR (10.0 Gb/s).....	551	40

Table 151	Equivalent Response of Reference O/E Converter	563	1
Table 152	Attenuation Tolerance of Reference O/E Converter.....	564	2
Table 153	Maximum Jitter of Optical Links for SDR	568	3
Table 154	Maximum Jitter of Optical Links for DDR	568	4
Table 155	Maximum Optical Bit to Bit Skew Values for 2.5 Gb/s	571	5
Table 156	Maximum Optical Bit to Bit Skew Values for 5.0 Gb/s	572	6
Table 157	Maximum Optical Bit to Bit Skew Values for 10.0 Gb/s	572	7
Table 158	Link Parameters - 1x-SX.....	573	9
Table 159	Optical Transmitter Parameters - 1x-SX.....	574	10
Table 160	Optical Receiver Parameters - 1x-SX.....	574	11
Table 161	Link Parameters - 1x-LX	575	12
Table 162	Optical Transmitter Parameters - 1x-LX	576	13
Table 163	Optical Receiver Parameters - 1x-LX	577	14
Table 164	Link Parameters - 1x-DDR-SX.....	578	15
Table 165	Optical Transmitter Parameters - 1x-DDR-SX	579	16
Table 166	Optical Receiver Parameters - 1x-DDR-SX	579	17
Table 167	Link Parameters - 1x-DDR-LX	580	18
Table 168	Optical Transmitter Parameters - 1x-DDR-LX	581	19
Table 169	Optical Receiver Parameters - 1x-DDR-LX	581	20
Table 170	Link Parameters - 4x-SX.....	583	21
Table 171	Optical Transmitter Parameters - 4x-SX.....	584	22
Table 172	Optical Receiver Parameters - 4x-SX.....	584	23
Table 173	Link Parameters - 4x-DDR-SX.....	586	24
Table 174	Optical Transmitter Parameters - 4x-DDR-SX	587	25
Table 175	Optical Receiver Parameters - 4x-DDR-SX	587	26
Table 176	Link Parameters - 12x-SX and 8x-SX.....	588	27
Table 177	Optical Transmitter Parameters - 12x-SX and 8x-SX	589	28
Table 178	Optical Receiver Parameters - 12x-SX and 8x-SX	589	29
Table 179	Link Parameters - 8x-DDR-SX and 12x-DDR-SX	591	30
Table 180	Optical Transmitter Parameters - 8x-DDR-SX and 12x-DDR-SX	591	31
Table 181	Optical Receiver Parameters - 12x-DDR-SX and 8x-DDR-SX	592	32
Table 182	Optical Fiber Specifications	604	33
Table 183	Compliance board response	619	34
Table 184	ICN Calculation Parameters for FDR and EDR cables.....	632	35

CHAPTER 1: INTRODUCTION

This document comprises the specification for the physical aspects of InfiniBand Architecture and addresses the following areas:

- Glossary of terms and acronyms used in this specification
[Chapter 2: Glossary](#)
- Overview of the Physical Layer and the aspects that make it up
[Chapter 3: Physical Layer Overview](#)
- Signal Definitions that comprise the defined connection interfaces
[Chapter 4: Port Signal Definitions](#)
- Link/Phy layer that defines the encoding and lane striping functions
[Chapter 5: Link/Phy Interface](#)
- High Speed Electrical Signaling defines the parameters for the basic link signals
[Chapter 6: High Speed Electrical Interfaces](#)
- Cable and connector physical and mechanical specifications
[Chapter 7: Electrical Connectors for Modules and Cables](#)
- Electrical and Logical specification of the management interface for passive and active cables
[Chapter 8: Management Interface](#)
- Optical fiber interface specifications
[Chapter 9: Fiber Attachment - 2.5 Gb/s, 5.0 Gb/s, & 10 Gb/s](#)

Several annexes specify methods for testing and verifying compliance with specifications for cables and devices.

- Test fixtures and methods for measuring FDR and EDR cable and device compliance with specified parameters
[Annex A1: FDR and EDR Compliance Boards and Test Setups](#)
- Methods for calculating various parameters that are derived from an array of basic parameter measurements
[Annex A2: Cable Electrical parameters for FDR and EDR](#)

Note that this document does not include description of a variety of functions from prior releases of the specification that have since been deprecated. These deprecated functions (e.g., 3U and 6U high modules to fit in InfiniBand™ I/O slots, backplane connectors for modules operating at 2.5 Gb/s bit rate, power and O/S management functions for modules, etc.) are moved into a separate document, named [InfiniBand™ Architecture Specification Volume 2-DEPR Release 1.3](#).

The summaries of compliance statements for various InfiniBand™ products have also been moved to a separate document, named [InfiniBand™ Architecture Specification Volume 2-COMP Release 1.3](#).

The following list provides an overview and summary of changes since Release 1.2.1.

- 1) Added transmitter, receiver, and channel parameters for FDR data rate
- 2) Added specification for 64b/66b link encoding for FDR and EDR data rates, vs. the 8b/10b link encoding used for SDR, DDR, and QDR data rates
- 3) Added new specifications for transmitter pre-emphasis and receiver equalization training
- 4) Added new compliance measurement methodology for FDR, that is also be applicable to EDR data rate for linear channels
- 5) Added required compliance and characterization functions
- 6) Added new specifications for active cable connector receptacles and memory maps for 4x (QSFP+) and 12x (CXP) interfaces for QDR/FDR and EDR (QSFP28, CXP28) speeds
- 7) Added definition of Host and Module Compliance Boards and characterization methodology
- 8) Added specifications for several forward error correction codes (FEC), supporting various tradeoff levels of complexity and latency vs.error correction capability
- 9) Added new specifications and measurement methodology for EDR limiting active cables (active optical cables and active electrical cables)

1.1 DOCUMENT CONVENTIONS

The following conventions are used in this specification.

1.1.1 NORMATIVE TERMS

- Shall

The use of the word shall indicates a mandatory requirement that must be implemented to claim compliance to this specification.

- Shall not

The use of the word shall not indicates a mandatory requirement to not implement a given aspect in order to claim compliance to this specification.

1.1.2 INFORMATIVE TERMS

- Should

The use of the word should indicates flexibility of choice in an implementation with a strongly preferred preference.

- **May**
The use of the word may indicates flexibility of choice in an implementation with no implied preference.

1.1.3 COMPLIANCE NOTATION

This specification has statements of compliance within the body of the document. These are identified using two notational conventions:

- **Mandatory compliance:** statements in the form of "CXX-YY: statement text" where 'XX' is the chapter number and 'YY' is the individual statement number. These statements are required to be satisfied depending on the compliance level being claimed.
- **Optional compliance:** statements in the form of "oXX-ZZ: statement text" where 'XX' is the chapter number and 'ZZ' is the individual statement number. These statements are required to be satisfied if a given optional feature is implemented within a compliance level being claimed.

The 'ZZ' numbers for Optional compliance statements increment independently from the 'YY' numbers used for Mandatory compliance statements.

[Volume 2C, Chapter 1: Volume 2 Compliance Summary](#) defines the compliance levels for Volume 2. Within that chapter, a summary listing of all the Mandatory compliance and Optional compliance statements that apply for that level are shown.

1.1.4 ARCHITECTURE NOTES

Architecture Note

This information appears in-line to clarify architected items.

1.1.5 IMPLEMENTATION NOTES

Implementation Note

This information appears in-line and describes hints on implementations.

1.1.6 RECOMMENDATION

Recommendation to Someone

This information appears in-line and describes a recommended approach of a feature.

1.2 REFERENCES AND RELATED DOCUMENTS

The following are documents to which this specification refers.

- [1] InfiniBand Architecture Specification, Volume 1
InfiniBand™ Trade Association, <http://www.infinibandta.org>
- [2] InfiniBand Architecture Specification, Volume 2-COMP
Summary of Vol. 2 compliance requirements for various product categories
InfiniBand™ Trade Association, <http://www.infinibandta.org>
- [3] InfiniBand Architecture Specification, Volume 2-DEPR
Deprecated features from prior releases of Vol. 2. These are features which may not be compatible with current Vol. 2 features and functions
InfiniBand™ Trade Association, <http://www.infinibandta.org>
- [4] ANSI/TIA/EIA-492AAAA-B-2009 - Detail Specification for 62.5- μm Core Diameter/125- μm Cladding Diameter Class 1A Graded-Index Multimode Optical Fibers Jan. 1, 1998
- [5] ANSI/TIA/EIA-492AAAB - Detail Specification for 50- μm Core Diameter/125- μm Cladding Diameter Class 1A Graded-Index Multimode Optical Fibers, Nov., 2009 (OM2)
- [6] ANSI/TIA/EIA-492AAC - Detail Specification for 850-NM Laser-Optimized, 50- μm Core Diameter/125- μm Cladding Diameter Class 1A Graded-Index Multimode Optical Fibers, Nov. 2009 (OM3)
- [7] ANSI/TIA/EIA-492AAAD - Detail Specification for 850-NM Laser-Optimized, 50- μm Core Diameter/125- μm Cladding Diameter Class 1A Graded-Index Multimode Optical Fibers Suitable for Manufacturing OM4 Cabled Optical Fiber, Sept. 2009 (OM3)
- [8] ANSI/TIA/EIA-492CAAA-98 – Detail Specification for Class IVa Dispersion-Unshifted Single-Mode Optical Fibers May 1, 1998
- [9] ANSI/TIA/EIA-604-10: FOCIS 10 – Fiber Optic Intermateability Standard Type LC, 1999
- [10] TIA/EIA-364-1000-2009, "Environmental Test Methodology for Assessing the Performance of Electrical Connectors and Sockets Used in Controlled Environment", February 12, 2009
- [11] TIA/EIA-364-70B-2007, "Temperature Rise Versus Current Test Procedure for Electrical Connectors and Sockets", June 8, 2007
- [12] TIA/EIA-364-108-2007, "Impedance, Reflection Coefficient, Return Loss, and VSWR Measured in the Time and Frequency Domain Test Procedure for Electrical Connectors, Cable Assemblies or Interconnection Systems", March 1, 2007
- [13] TIA/EIA-364-107-2007, "Eye Pattern and Jitter Test Procedure for Electrical Connectors, Sockets, Cable Assemblies or Interconnection Systems", March 1, 2007

[14] Center for Devices & Radiological Health	1
[15] Fibre Channel - Methodologies for Jitter Specification revision 10	2
[16] Fibre Channel Physical Interface Standard revision 8	3
[17] Fibre Channel Physical Interface Standard revision 8, Annex A	4
[18] FOTP-107 (TIA/EIA-107A) – Return Loss for Fiber Optic Components: 1st Ed. Feb. 1989, 2nd Ed. Mar. 1999	5
[19] IEC 1754-7-3 - Push/Pull MPO Adapter Interface Standard	6
[20] IEC 1754-7-4 - Push/Pull MPO Female Plug Connector Interface	7
[21] IEC 1754-7-5 – Push/Pull MPO Male Plug Connector Interface	8
[22] IEC 60793-2 Type A1a Dec. 1998	9
[23] IEC 60793-2 Type A1b Dec. 1998	10
[24] IEC 60793-2 Type B1 Dec. 1998	11
[25] IEC 60825-1 - Radiation Safety of Laser Products - Equipment Classification, Requirements and User's Guide, 1st Ed. Nov. 1993, Amended Sep. 1997	12
[26] IEC 60825-1 Class I, IIIA, IIIB – Nov. 1993	13
[27] IEC 512-5-1	14
[28] IEEE 802.3-2015 Standard for Ethernet	15
[29] ITU-T G.957 – Optical Interfaces for Equipment and systems Relating to the Synchronous Digital Hierarchy June, 1999	16
[30] TIA2.2.1 working specification TIA/EIA-455-204-FOTP204 Measurement Method for Multimode Fiber Bandwidth – to be published	17
[31] ISO 8601 Date/Time Format, http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=40874	18
[32] IEC 61280-4-1 ed2.0 Fibre-optic communication subsystem test procedures - Part 4-1: Installed cable plant - Multimode attenuation measurement, Publication date 2009-06-10	19
[33] System Management Bus Specification, Revision 1.1, December 11, 1998. Copyright(c)1996, 1997, 1998, Benchmark Microelectronics Inc., Duracell Inc., Energizer Power Systems, Intel Corporation, Linear Technology Corporation, Maxim Integrated Products, Mitsubishi Electric Corporation, National Semiconductor Corporation, Toshiba Battery Co., Varta Batterie AG	20
[34] The I2C-Bus Specification, Version 2.0, December 1998, Philips Semiconductors	21
[35] IEEE Standard 181-2011, "IEEE Standard for Transitions, Pulses, and Related Waveforms" [supersedes 181-2003]	22
[36] Richard E. Blahut, "Algebraic Codes for Data Transmission", Cambridge University Press, 2012	23
[37] Shu Lin and Daniel J. Costello, "Error Control Coding", Prentice Hall, 2004	24

[38] ITU-T Recommendation O.153 - Basic parameters for the measurement of error performance at bit rates below the primary rate, 1992	1
[39] NEBS-GR-63-CORE, "NEBS Requirements: Physical Protection", Issue Number 04, Apr 2012	2
[40] ANSI/ESDA/JEDEC JS-001-2015, "For Electrostatic Discharge Sensitivity Testing Human Body Model (HBM) - Component Level" [supersedes JESD22-A114:B2000]	3
[41] IEC / EN 61000-4-2:2009, "Electromagnetic compatibility (EMC). Testing and measurement techniques. Electrostatic discharge immunity test"	4
[42] OIF-CEI-03.0, "Common Electrical I/O (CEI) - Electrical and Jitter Interoperability agreements for 6G+ bps, 11G+ bps and 25G+ bps I/O", 1st September 2011	5
[43] SFF-8024 "SFF Committee Cross Reference to Industry Products", Rev 4.1 June 27, 2016	6
[44] SFF / INF-8074, "Specification for SFP (Small Formfactor Pluggable) Transceiver", Rev 1.0 May 12, 2001 and related documents from the SFF TA TWG of the SNIA (Small Form Factor Technology Affiliate Technical Working Group of the Storage Networking Industry Association), including	7
SFF-8084 SFP+ 1X 4 Gb/s Pluggable Transceiver Solution	8
SFF-8083 SFP+ 1X 10 Gb/s Pluggable Transceiver Solution (SFP10)	9
SFF-8081 SFP+ 1X 16 Gb/s Pluggable Transceiver Solution (SFP16)	10
SFF-8402 SFP+ 1X 28 Gb/s Pluggable Transceiver Solution (SFP28)	11
SFF-8419 SFP+ Power and Low Speed Interface	12
SFF-8418 SFP+ 10 Gb/s Electrical Interface	13
SFF-8472 Diagnostic Monitoring Interface for Optical Transceivers	14
SFF-8071 SFP+ 1X 0.8mm Card Edge Connector	15
SFF-8432 SFP+ Module and Cage	16
SFF-8433 SFP+ Ganged Cage Footprints and Bezel Openings	17
[45] SFF-8436, "Specification for QSFP+ 10 Gbs 4X Pluggable Transceiver", Rev 4.8 October 31, 2013, and related, updated documents:	18
SFF-8635 QSFP+ 10 Gb/s 4X Pluggable Transceiver Solution (QSFP10)	19
SFF-8636 Common Management Interface	20
SFF-8679 QSFP+ 4X Electrical Specification	21
SFF-8682 QSFP+ 14 Gb/s 4X Connector	22
SFF-8683 QSFP+ 14 Gb/s Cage	23
SFF-8685 QSFP+ 14 Gb/s 4X Pluggable Transceiver Solution (QSFP14)	24
SFF-8661 QSFP+ 28 Gb/s 4X Pluggable Module	25
SFF-8662 QSFP+ 28 Gb/s 4X Connector (Style A)	26
SFF-8663 QSFP+ 28 Gb/s Cage (Style A)	27
SFF-8665 QSFP+ 28 Gb/s 4X Pluggable Transceiver Solution (QSFP28)	28
SFF-8672 QSFP+ 28 Gb/s 4X Connector (Style B)	29
[46] SFF-8642/EIA-965, "MiniMultilane 10 Gb/s 12x Shielded Connector" (CXP10)", Rev 2.9 September 16, 2012, and related, updated documents:	30
SFF-8647 Mini Multilane 14 Gb/s 12x Shielded Cage/Connector (CXP14)	31
SFF-8648 Mini Multilane 12x 28 Gb/s Shielded Cage/Connector (CXP28)	32

The IBTA Compliance and Interoperability Working Group (CIWG) maintains a set of "MOI" (Method of Implementation) documents, which describe in detail the procedures and specific test equipment used for compliance and interoperability tests. These MOI documents are available in the CIWG section of the InfiniBand Trade Association web site at <http://www.infinibandta.org>.

1.3 ACKNOWLEDGMENTS

This specification represents the collaboration of a number contributing companies and individuals. Special thanks to those who contributed.

1.3.1 STEERING COMMITTEE

1.3.1.1 Co-CHAIRS - DIRECTORS

The following individuals served as directors of the InfiniBand™ Trade Association through the Steering Committee during the creation of Release 1.0 and 1.1 or creation of Releases 1.2 and 1.2.1 of this specification:

Tom Bradicich

Tom Macdonald

The following individuals served as co-chairs of the InfiniBand™ Trade Association Steering Committee during the creation of Release 1.3 of this specification:

Mark Atkins

Jim Pappas

The following individuals served as co-chairs of the InfiniBand™ Trade Association Steering Committee during the creation of Release 1.3.1 of this specification:

Barry Barnett

Mark Atkins

Jim Pappas

1.3.1.2 MEMBERS

The following individuals served as members of the InfiniBand™ Trade Association Steering Committee during the creation of Releases 1.0 and 1.1 or creation of Releases 1.2 and 1.2.1 of this specification:

Jacqueline Balfour

Ken Jansen

Jim Pinkerton

Kevin Deierling

Michael Krause

Martin Whittaker

Balint Fleisher

Todd Matters

Bob Zak

Dr. Alfred Hartmann

Ed Miller

40

David Heisey

John Pescatore

41

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

The following individuals served as members of the InfiniBand™ Trade Association Steering Committee during the creation of Release 1.3 of this specification:

David Brean	Skip Jones	Marc Sultzbaugh
Paul Grun	Scott Misage	

The following individuals served as members of the InfiniBand™ Trade Association Steering Committee during the creation of Release 1.3.1 of this specification:

David Brean	Ranjit Henry	Alex Nicolson
Kobby Carmona	Skip Jones	Marc Sultzbaugh
Kevin Dierling	Dale Kaisner	Pat Thaler
Albert Greenberg	Mark Miller	
Paul Grun	Scott Misage	

1.3.2 TECHNICAL WORK GROUP

1.3.2.1 CO-CHAIRS

The following individuals served as chairs of the Technical Work Group (TWG) with responsibility for the oversight of the technical production during the creation of Release 1.0 and 1.1 or creation of Releases 1.2 and 1.2.1 of this specification:

Dwight Barron	Jeff Hilland	David Wooten
Paul Grun	Irving Robinson	

The following individuals served as chairs of the Technical Work Group (TWG) during the creation of Release 1.3 and 1.3.1 of this specification:

Diego Crupnicoff	Bill Magro
------------------	------------

1.3.2.2 MEMBERS

The following individuals served as members of the Technical Work Group (TWG) with responsibility for the oversight of the technical production during the creation of Releases 1.0 and 1.1 or creation of Releases 1.2 and 1.2.1 of this specification:

Dr. Alan Benner	Dr. Al Hartmann	Dr. Greg Pfister
Mark Bradley	Michael Krause	Greg Still
Wolfgang Cristl	Bill Lynn	Ken Ward
Diego Crupnicoff	Ed Miller	

The following individuals served as members of the Technical Work Group (TWG) during the creation of Release 1.3 of this specification:

Alan Benner	Arlin Davis	Ira Wein
Brad Benton	Paul Grun	Mike Woodacre
Ariel Cohen	Ted Kim	Eitan Zahavi
David Craddock	Todd Rimmer	
Rupert Dance	Tom Sand	

The following individuals served as members of the Technical Work Group (TWG) during the creation of Release 1.3.1 of this specification:

Alan Benner	Paul Grun	Jim Ryan
David Craddock	Jimmy Hill	Tom Sand
Rupert Dance	Ted Kim	Pat Thaler
Freddy Gabbay	Bill Lee	

1.3.3 VOLUME WORKING GROUP

This volume is principally authored by the membership the Electrical/Mechanical Work-group (EWG) of the InfiniBand Trade Association, with significant input from members of the othe IBTA Working Groups. The other working groups that are part of the InfiniBand Trade Association are listed in *InfiniBand Architecture Specification, Volume 1* and *Infini-Band Architecture Specification, Volume 3*. The EWG is indebted for their efforts in consultation and review.

1.3.3.1 Co-CHAIRS

The following individuals served as co-chairs of the EWG during the creation of Releases 1.0 and 1.1 of this specification:

Mike Chastain	David Moss
---------------	------------

The following individual served as chair of the EWG during the creation of Releases 1.2 and 1.2.1 of this specification:

Moises Cases

The following individual served as chair of the EWG during the creation of Release 1.3 and Release 1.3.1 of this specification:

Alan Benner

The following individuals served as co-chair of the EWG-EE High-Speed Electrical Subteam of the EWG during the creation of Release 1.3 and Release 1.3.1 of this specification:

Jay Diepenbrock Brent Rothermel

1.3.3.2 VOLUME EDITOR

The following individual served as editor of Release 1.0 and 1.1 of this volume during its creation:

Greg Still

The following individual served as editor of Release 1.2, 1.2.1, 1.3, and Release 1.3.1 of this volume during its creation:

Alan Benner

1.3.3.3 SUBTEAM LEADERS

The following individuals served as subteam leaders and specific chapter editors of Release 1.0 and 1.1 of this volume during its creation:

Bill Bunton	Dominic Goodwill	Trevor Williams
Del Cecchi	John Leder	
Jay Diepenbrock	Siamak Tavallaei	

The following individuals served as subteam leaders and specific chapter editors of Release 1.2 and Release 1.2.1 of this volume during its creation:

Alan Benner	Jay Diepenbrock
Del Cecchi	Trevor Williams

The following individuals served as subteam leaders and specific chapter editors of Release 1.3 and Release 1.3.1 of this volume during its creation:

Alan Benner	Brent Rothermel
Rupert Dance	Oren Sela
Jay Diepenbrock	

1.3.3.4 INDIVIDUAL CONTRIBUTORS

The following individuals provided active contributions to Release 1.0 and 1.1 of this volume during its creation:

Andrew Alduino	Ken Gross	Eddie Reid	1
Paul Artman	Art Kimmel	Harry Rogers	2
Brian Beaman	Don Lenz	Art Rousmaniere	3
Dave Brown	Dennis Miller	Richard Schumacher	4
Ray Clemo	J. P. Miller	Kim Sides	5
Steve Contreras	Mike Miller	Tom Slaight	6
Casimer DeCusatis	Mark Myers	Bill Stanley	7
Jay Diepenbrock	Istvan Novak	Dan Stigliani	8
Hal Dozier	Shlomo Novotny	Farrukh Syed	9
Pat Egan	Tom Osten	Pat Thaler	10
Stillman Gates	Ali Oztaskin	Chris Trudeau	11
Ali Ghiasi	Ken Privitt	Jim Warren	12
Lowell Good	Kevin Przybylski		13
			14
			15
			16
			17
			18

The following individuals provided active contributions to Release 1.2 and Release 1.2.1 of this volume during its creation:

Moises Cases	Andy Kayner	Greg Pfister	19
John Calvin	Tadashi Kumamoto	Daniel Reed	20
Del Cecchi	Todd Leonard	Eddie Reid	21
William Cerreta	Jacques Longueville	Michael Rost	22
Hillel Chapman	Gregory Mann	Peter Smith	23
Ian Colloff	Padraig McDaid	Mike Sorna	24
Casimer DeCusatis	J. P. Miller	Greg Still	25
Christy Devonport	Keith Murr	Antonius Susanto	26
Jay Diepenbrock	Dale Murray	Trevor Williams	27
Ronen Eckhouse	Jay Neer		28
Shahar Eytan	John Petrilla		29
			30
			31

The following individuals provided active contributions and/or commentary to Release 1.3 of this volume during its creation:

David Bahr	Roman Inozemtsev	Ran Ravid	32
Carlos Castil	Kyle Klinger	Hal Rosenstock	33
Rupert Dance	Keith Lang	Michael Rost	34
Matt Davis	Catalin Muntean	Olindo Savi	35
Piers Dawe	Jay Neer	John Sawdy	36
Mike Dudek	Mark Owen	Wei Tang	37
Jason Ellison	Tom Palkert	Jim Vana	38
Jørgen Høg	John Petrilla		39
			40
			41
			42

The following individuals participated in discussions regarding Release 1.3 of this volume during its creation:

Shahela Ali	Galen Fromm	Shai Raphaeli	1
Abel Almazan	Dariusz Gorzkiewicz	Michael Ressl	2
Michael Annand	Mitch Ito	Yoav Rozenberg	3
Pervez Aziz	Ken Jackson	Jeff Rubin	4
Vittal Balasubramanian	Yonatan Malkiman	Katharine Schmidtke	5
Peter Baughman	Greg McSorley	Atul Sharma	6
Bob Brown	Idan Mizrahi	Xueliang Song	7
Brad Brubaker	Merrick Moeller	Don Thompson	8
Ed Cady	Glenn Moore	Ola Torudbakken	9
Patrick Decker	Thinh Nguyen	Sen Velmurugan	10
Eric Dube	David Nidelius	Dean Vermeersch	11
Hecham Elkhatab	Andrew Nowak	Michael Walmsley	12
Bob Elliot	Vit Novak	Bob Wagner	13
Jason Ellison	Gourgen Oganessyan	Dean Wallace	14
Zvika Eyal	Hans Opheim	A.J. Yang	15
Frank Flens	Troy Oxby		16
Daniel Fradkin	Bruce Pecor		17
			18
			19
			20
			21
			22
			23
			24
			25
			26
			27
			28
			29
			30
			31
			32
			33
			34
			35
			36
			37
			38
			39
			40
			41
			42

The following individuals provided active contributions and/or commentary to Release 1.3.1 of this volume during its creation:

David Bahr	Mitch Ito	Michael Ressl	1
Noah Beltran	Ken Jackson	Brent Rothermel	2
Carlos Castil	John Lenthal	Alexander Rysin	3
John Calvin	Yonatan Malkiman	Oren Sela	4
Randy Clark	Tomer Michaeli	John Sawdy	5
Cliff Cole	Idan Mizrahi	Robert Sleigh	6
Rupert Dance	James Morgante	Xueliang Song	7
Matt Davis	Glenn Moore	Wei Tang	8
Piers Dawe	Catalin Muntean	Jim Vana	9
Jay Diepenbrock	Jay Neer	Oded Wertheim	10
Mike Dudek	Thinh Nguyen	Xin Wu	11
Jason Ellison	Hans Opheim	Eitan Zahavi	12
Frank Flens	Tom Palkert	Jie Zheng	13
Hiroshi Goto	John Petrilla		14
Alex Haser	Zvi Rechtman		15
			16
			17
			18
			19
			20

The following individuals participated in discussions regarding Release 1.3.1 of this volume during its creation:

Ed Cady	Bernie King-Smith	Takeshi Nishimura	20
Roy Cideciyan	Daniel Kuchta	Joe O'Brien	21
Patrick Decker	Jeff Lapak	Hal Rosenstock	22
Curtis Donahue	Ryan Lott	Chris Roth	23
Mitch Fields	Jeff Mason	Katharine Schmidtke	24
Dariusz Gorzkiewicz	Erdem Matoglu	Sandy Shirk-Heath	25
Jørgen Høg	Gerald Monaco	Pavel Zivny	26
Thad Kadela	Frank Morana		27
			28
			29
			30

1.4 DISCLAIMER

Like any document, this specification is subject to errata for correctness, clarity, and enhancements. The InfiniBand Trade Association hosts a web site at <http://www.infiniband.org>. Please visit this site to check for errata and updates to this specification.

31
32
33
34
35
36
37
38
39
40
41
42

CHAPTER 2: GLOSSARY

The terms in this chapter are divided as follows: [Section 2.1. "A to C." on page 37](#), [Section 2.2. "D to H." on page 39](#), [Section 2.3. "I to L." on page 43](#), [Section 2.4. "M to P." on page 47](#), [Section 2.5. "Q to S." on page 52](#), [Section 2.6. "T to Z." on page 55](#).

2.1 A TO C

Address Handle	An object that contains the information necessary to transmit messages to a remote port over Unreliable Datagram service.	10
Actively Managed Chassis	An Actively Managed Chassis provides an InfiniBand™ specified GUID and physical Slot Information to every IB Module on the Module's unique IB-ML . In addition, it provides a Chassis Management Entity on at least one IB Module's IB-ML. In an actively managed chassis, the Slot Information for every Module provides information identifying all of the Slot(s) that have access to the CME.	12
Adapter	Also called I/O adapter. Please see Host Channel Adapter or Target Channel Adapter . A form of TCA which conforms to the InfiniBand™ Architecture form factor definition. The term adapter by itself is overloaded due to its general use in the computer industry and should not be used by itself. Typically the term adapter pertains to the channel adapter but is also used in the context of an IO adapter.	17
Anycast	An identifier for a set of interfaces (typically belonging to different nodes). A packet sent to an anycast address is delivered to one of the interfaces identified by that address (the "nearest" one, according to the routing protocols' measure of distance).	22
Attribute	The collection of management data carried in a Management Datagram .	26
Aux Power	The 5V auxiliary power input to an IB Module that is provided on the VA_In pin. This is the standby power source used when Bulk Power is off.	28
Average Optical Power	The optical power measured using an average reading power meter when transmitting a specified code sequence as defined in the test procedure.	30
B_Key	Please see Baseboard Management Key .	33
Backplane	Physical PCB into which an IB Module plugs, this may be into a Server, Switch, I/O Chassis, etc.	35
Backplane Connector	The connector located in a Server , Switch , I/O Chassis , etc. that mates with an IB Modules Edge Connector .	37
Base LID	The numerically lowest Local Identifier that refers to a Port .	40

Baseboard Managed Unit	Any Unit which provides InfiniBand™ specification defined information about itself by a Baseboard method MAD operation through the InfiniBand™ link.	1
Baseboard Management Key	A construct that is contained in IBA management datagrams to authenticate that the sender is allowed to perform the requested operation.	2
Baseboard Management	Management of physical hardware (environmentals, Vital Product Data, LEDs, etc.). Considered synonymous with System Management as opposed to "Subnet Management".	3
Baseboard Management Proxy	A Unit which provides the means to translate InfiniBand™ Baseboard Management Datagram (MAD) requests and responses to IB-ML or other transports on behalf of one or more other entities. The proxy must be addressable by the InfiniBand™ subnet to allow operations to be targeted to it (aka it has a LID).	4
baud (Bd)	A unit of measure of signaling rate on a lane, expressed as the number of possible transitions per second.	5
Beacon Sequence	The periodic transmission of the Training Sequence 1 Control Ordered-set . The Beacon sequence is used to wake or begin Link Training with the device on the opposite end of a Link that may be in a power managed state or has been just attached to the link.	6
BER	Bit error ratio.	7
BTH	Base Transport Header.	8
Board	Physical, pluggable entity that is defined by the InfiniBand™ Architecture.	9
Bulk Power	The 12V main power input to an IB Module that is provided on the VB_In pins.	10
Burst-BER	Number of bit errors measured within a sliding window.	11
Byte Striping	The byte stream representing the packet is sent by distributing the bytes sequentially across the available Physical Lanes , one byte per Lane , in order. The first byte goes in lane 0, the second in lane 1, the third in lane 2 and so on. When the last physical lane is reached, the process starts again with lane 0.	12
CA	Please see Channel Adapter .	13
Carrier Module	The carrier module, designed specifically for each size of IB Module defines the thermal, EMI, ESD and vibration/shock interface to a chassis.	14
CDR	Clock Data Recovery unit	15
Center Wave-length	The nominal value of the central wavelength of the operating, modulated laser. This is the wavelength (see FOTP-127) where the effective optical power resides.	16

Channel	The association of two queue pairs for communication.	1
Channel Adapter	Device that terminates a link and executes transport-level functions. One of Host Channel Adapter or Target Channel Adapter .	2
Chassis	The collection of IB Modules, and their associated power and cooling resources housed within a single mechanical package. This may be a Server, Switch, I/O Chassis, etc.	3
Chassis GUID	8 bytes of Globally Unique ID for every Chassis	4
Chassis Management Entity	The Chassis Management Entity may or may not include a processor. In its simplest form the CME is merely an IB-ML de/MUX. The CME may provide proxy access to the IB Modules IB-ML.	5
Clock Compensation	The SKIP ordered-set is used for clock compensation. A device may add or drop a SKP symbol from the Control Ordered-set to prevent its link input/output buffer from over-running or under-running.	6
CME	Please see Chassis Management Entity .	7
COM	The Comma symbol is transmitted to identify the start of a Training Sequence 1 , Training Sequence 2 or SKIP ordered-set .	8
Compliance Channel	A worst-case connection from driver to receiver. The compliance channel and the transmitter define the minimum acceptable inputs which a receiver shall be capable of receiving at a specified bit error ratio.	9
Compliant Channel	A compliant channel is any channel which provides a signal at the receiver which is better than the Compliance Channel .	10
Control Ordered-set	Control Ordered-sets are used for Link Training and Clock Compensation . The first symbol of all ordered-sets is the COM symbol, additional symbols are unique to the set type.	11
Cover	The protective cover which mates with the carrier module defining the outer surface of an IB Module .	12
CRC	Please see Cyclic Redundancy Check .	13
Cyclic Redundancy Check	A number derived from, and stored or transmitted with, a block of data in order to detect corruption. By recalculating the CRC and comparing it to the value originally transmitted, the receiver can detect some types of transmission errors.	14
2.2 D TO H		15
Data Payload	The data, not including any control or header information, carried in one packet.	16
		17
		18
		19
		20
		21
		22
		23
		24
		25
		26
		27
		28
		29
		30
		31
		32
		33
		34
		35
		36
		37
		38
		39
		40
		41
		42

Data Dependent Pulse Width Shrinkage	The difference between 1 UI and the shortest interval between threshold crossings for a PRBS9 after averaging.	1
DDPWS	Please see Data Dependent Pulse Width Shrinkage	2
DDJ	Please see Jitter, Data Dependent .	3
DETH	Datagram Extended Transport Header.	4
de/MUX	Multiplexer / Demultiplexer	5
DFE	Decision Feedback Equalizer.	6
DGID	Destination Globally Unique Identifier .	7
DJ	Please see Jitter, Deterministic	8
Disparity	The difference between the number of ones and number of zeros transmitted in a Physical Lane . The running disparity is a binary parameter with a positive or negative value.	9
Dispersion	A term used to denote pulse broadening and distortion. For fiber optic links, the two general categories of dispersion are modal dispersion, due to the difference in the propagation velocity of the propagation modes in a multimode fiber, and chromatic dispersion, due to the difference in propagation of the various spectral components of the optical source.	10
DLID	Destination Local Identifier	11
Duty Cycle Distortion	The absolute value of the difference between the average time of rising edges and the average time of falling edges for a mixed-frequency pattern such as PRBS9 or PRBS31. It may be measured by the absolute value of the difference between the pulse width of a '1' or a '0' pulse and the ideal unit interval in a 10101010 sequence embedded within a mixed-frequency pattern.	12
EBP	Please see End of Bad Packet Delimiter .	13
Edge Connector	The connector on an IB Module that mates with a Backplane Connector .	14
EGP	Please see End of Good Packet Delimiter .	15
EMC	Electro-Magnetic Compatibility.	16
EMC Gasket	The name of the gasket used between IB Modules to shield for electromagnetic emis-	17

	sions.	1
Enhanced Signaling	All signaling functions specified in Rel. 1.2 and later releases, and not specified in Rel. 1.1 or earlier releases. These functions include DDR, QDR, FDR and EDR signaling rates as well as enhanced functions such as (from Rel. 1.2) TS3 ordered-sets, heartbeat ordered-sets, LinkRoundTripLatency measurement, TS-T ordered-sets, adaptive driver de-emphasis, and Phy Test compliance testing, as well as (from Rel. 1.3) 64b/66b encoding, Rev 1 TS3, improved transmitter equalization functionality, max packet rate, and forward error correction coding.	2 3 4 5 6 7 8
End of Good Packet Delimiter	The End of Good Packet Delimiter (EGP) symbol is used to mark the end of each packet as it is transmitted by the originating node.	9 10 11
End of Bad Packet Delimiter	The End of Bad Packet Delimiter (EBP) symbol is used to mark the end of a bad packet forwarded by a switch or router node.	12 13 14 15
Endnode	An endnode is any node that contains a Channel Adapter and thus it has multiple queue pairs and is permitted to establish connections, end to end context, and generate messages. Also referred to as Host Channel Adapter or Target Channel Adapter , two specific types of endnodes.	16 17 18 19
Endpoint	A Port which can be a destination of LID -routed communication within the same Subnet as the sender. All Channel Adapter ports on the subnet are endpoints of that subnet, as is Port 0 of each Switch in the subnet. Switch ports other than Port 0 may not be endpoints. When <i>port</i> is used without qualification, it may be assumed to mean <i>endpoint</i> whenever the context indicates that it is a destination of communication.	20 21 22 23 24
ESD	Electro-Static Discharge.	25 26
Error Event	A single root cause can result in multiple error events.	27 28
Error Propagation	A special character in the end of packet delimiter is used to delimit packets in which a transmission error has already been detected and reported.	29 30 31
Even Alignment	A special character in the code set used on the link to establish a given disparity.	32 33
Extinction Ratio	The ratio (in dB) of the average optical energy in a logic one level to the average optical energy in a logic zero level measured under modulated conditions at the specified signaling rate.	34 35 36
Fabric	The collection of Links , Switches , and Routers that connects a set of Channel Adapters .	37 38
Fall Time	The time interval for the falling edge of a pulse to transition from its 80% amplitude level to its 20% amplitude level.	39 40 41 42

FFE	Feedforward Equalizer.	1
Fiber Optic Adapter	A device into which two optical connectors plug, joining two optical segments.	2
Fiber Optic Cable	A jacketed optical fiber or fibers.	3
Fiber Optic Segment	An unbroken length of optical fiber with an optical connector on each end. The fiber may contain splices. A fiber optic segment shall not contain a fiber optic adapter.	4
Fiber Optic Test Procedure	EIA/TIA standards developed and published by the Electronic Industries Association (EIA) and Telecommunications Industry Association (TIA) under the EIA-RS-455 series of standards. Please see FOTP .	5
Form Factor	The definition of physical packaging that specifies mechanical limits, connectors, and placement of connectors and connector plates.	6
FOTP	Please see Fiber Optic Test Procedure .	7
Fully Managed Repeater	A Repeater which provides VPD information about itself by a GetInfo MAD operation targeted to it and allows IB-ML access to defined information. A repeater cannot fulfill this definition.	8
Fully Managed TCA	A TCA which provides Module Information about itself when a GetInfo MAD operation is targeted to it and allows IB-ML access to defined Module information.	9
Fully Managed Unit	Any Unit which is both a Baseboard Managed Unit and an IB-ML Managed Unit .	10
Gb/s	Giga-bits per second (10^9 bits per second)	11
GB/s	Giga-bytes per second (10^9 bytes per second)	12
General Services Interface	An interface providing management services (e.g., connection, performance, diagnostics) other than subnet management. QP1 is reserved for the GSI, which may redirect requests to other QPs.	13
GID	Please see Globally Unique Identifier .	14
Globally Unique Identifier	A software-readable number that uniquely identifies a device or component.	15
GMP	General Management Packet.	16
Graceful Hot Removal	The removal of an IB Module that has first been placed in a quiescent state. V_{Bulk} may or may not be on.	17

GSI	Please see General Services Interface .	1
GT/s	Giga-transfers per second (10^9 transfers per second).	2
GUID	Please see Globally Unique Identifier .	3
HCA	Please see Host Channel Adapter .	4
Host	One or more Host Channel Adapters governed by a single memory/CPU complex.	5
Host Channel Adapter	A Channel Adapter that supports the Verbs interface.	6
Host Connector	The connector interface associated with the mating of a pluggable device to a printed circuit board (PCB).	7
Host Node	A host node is a type of endnode consisting of one or more host channel adapters governed by a single memory/CPU complex. A host node supports one or more software processes, which use the subnet for communication between peer processes (IPC) and I/O services. A host may support processes that provide an I/O service (e.g., console service). Such a process is considered an I/O controller and the host enumerates that service the same as an I/O unit enumerates I/O controllers.	8
Hot Add	The insertion of an IB module into a backplane that has both V_{Bulk} and V_{Aux} present. The IB module powers up and initiates a training sequence.	9
2.3 I TO L		10
IBA	InfiniBand™ Architecture.	11
IB-ML	InfiniBand™ Management Link.	12
IB-ML Managed Unit	Any Unit which provides InfiniBand™ specification defined information about itself through accesses from the IB-ML .	13
IB-ML Management Proxy	A Unit which serves as a means to translate IB-ML operations (requests and responses) to the InfiniBand™ Subnet link(s) on behalf of one or more other entities.	14
IB-ML Master	A device on the IB-ML that initiates operations and provides the clock for the transfer to or from an IB-ML Slave device.	15
IB-ML Slave	A device on the IB-ML that responds to an operation that is addressed to it by an IB-ML Master . The clock used for the transfer is provided by the IB-ML Master .	16
IB Board	The PCB assembly inside an IB Module .	17
IB Module	A unit that conforms to any of the form factors defined in Volume 2 of the InfiniBand™	18

specification. The IB module minimally consists of the following: at least one [IB Board](#), a [Carrier Module](#), and a [Cover](#). There are two defined module heights and two defined module widths. Double width modules occupy two IB chassis slots.

	Standard Module	1
	Standard Double Wide Module	2
	Tall Module	3
	Tall Double Wide Module	4
IBT	InfiniBand™ Technology	5
IHV	Independent Hardware Vendor	6
Idle Data	Data symbols transmitted to fill idle time on a link. These symbols are not part of a packet and do not have delimiters symbols to mark their start and end.	7
In-band Management	Refers to the monitoring and control of InfiniBand™ components using the IB Subnet Link.	8
Independent Hardware Vendor	Any vendor providing hardware. Used synonymously at times with Hardware Vendor.	9
Inter IB-ML Management Proxy	A unit which serves to translate IB-ML operations from one unit to another unit's IB-ML. (This is currently not specified in the InfiniBand™ Architecture but is included for completeness).	10
Intersymbol Interference	The effect on a sequence of symbols in which the symbols are distorted by transmission through a limited bandwidth medium to the extent that adjacent symbols begin to interfere with each other. Also referred to as ISI .	11
Invalid Key	A Key is invalid if it numerically different from the correct Key value associated with an IBA-defined resource. Please see Key .	12
I/O	Input/Output.	13
I/O Adapter	An I/O controller and TCA, usually implemented as an IB Module, that provides access to one or more I/O devices possible attached via a secondary peripheral bus of network.	14
I/O Chassis	The collection of Slots and their associated power and cooling resources housed within a single mechanical package.	15
I/O Controller	One of the two architectural divisions of an I/O Unit . An I/O controller (IOC) provides I/O services, while a Target Channel Adapter provides transport services.	16
I/O Hierarchy	An I/O unit contains one or more I/O controllers. Each I/O controller provides access to one or more I/O devices or I/O ports. This is roughly equivalent to a PCI card that might	17

	contain up to 8 I/O functions.	1
	On the host side, the equivalent of a controller is a process. The part of the I/O process that accesses an I/O controller is referred to as the I/O driver. An I/O driver controls one or more I/O controllers, thus there is an instance of an I/O device for each I/O controller.	2
I/O Node	A type of endnode that provides I/O functions.	3
I/O Plate	Physical end of an IB Module where the I/O media cables are mounted for exit.	4
I/O Unit	An I/O unit (IOU) provides I/O service(s). An I/O unit consists of one or more I/O Controllers attached to the fabric through a single Target Channel Adapter .	5
IOC	Please see I/O Controller .	6
IOU	Please see I/O Unit .	7
ISI	Please see Intersymbol Interference .	8
J2	Please see Jitter, J2 .	9
J9	Please see Jitter, J9 .	10
Jitter	The difference between the time of occurrence of the actual threshold crossing of a signal and the time of occurrence of the threshold crossing of an ideal reference signal.	11
Jitter, Data Dependent	The time difference between the latest and earliest threshold crossings for a PRBS9 after averaging.	12
Jitter, Deterministic	Timing distortions, defined as Total Jitter (TJ) minus Random Jitter (RJ). Deterministic Jitter may be described as the dual-Dirac estimate of high probability jitter, which is found from a Gaussian fit to the tails of the jitter distribution of a signal, or as being composed of other components, including DCD (please see Duty Cycle Distortion), DDJ (please see Jitter, Data Dependent), SJ (Sinusoidal Jitter, caused by interaction with low-frequency oscillators), and UJ (Uncorrelated Jitter, caused by other effects). DJ is caused by the interaction of the limited bandwidth of the transmission system components and the symbol sequence.	13
Jitter, J2	The time interval that includes all but 10^{-2} of the jitter distribution.	14
Jitter, J9	The time interval that includes all but 10^{-9} of the jitter distribution.	15
Jitter, Pattern Dependent	Please see Jitter, Data Dependent .	16
Jitter, Random	Jitter which may be modeled as a Gaussian probability distribution function (PDF). The peak-to-peak value of RJ is of a probabilistic nature and thus any specified value yields an associated BER. Random Jitter is found from a Gaussian fit to the tails of the jitter	17

	distribution of a signal. Sometimes referred to as Gaussian Jitter.	1
Jitter, Total	The difference between the two sampling times before and after the majority of the transitions of a signal at which the error ratio at these sampling times is equal to the specification error ratio. Total jitter is customarily subdivided into random jitter (RJ) and deterministic jitter (DJ) components.	2 3 4 5 6 7 8 9 10 11 12 13
Key	A construct used to limit access to one or more resources, similar to a password. Security is provided by (1) limiting the ability to generate keys and check keys to well-trusted levels of function in the network, (2) making the keys large numbers, and managing how they are changed, so that each key should be unique within the system at once, and (3) the detection of invalid key usage is raised to high levels of software where the source of the invalid key can be identified and dealt with. The following keys are defined by the InfiniBand™ Architecture:	7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
	Baseboard Management Key .	14
	Management Key .	15
	Queue Key .	16
	Partition Key .	17
L_Key	Please see Local Route Header .	19
Lane	Please see Physical Lane	21
Lane Identifier	The second symbol of a training sequence that identifies the Physical Lane(s) of a 1x, 4x or 12x Link . It is used to recognize physical lane reversal and to determine link width, etc.	22 23 24 25 26 27 28 29 30 31 32 33 34
LED	Light Emitting Diode.	26 27 28 29 30 31 32 33 34
'Let-Through' Failure	A Failure mechanism of non-isolated voltage converters which allows the primary input power of the converter, through a DC path, onto the output.	28 29 30 31 32 33 34
LID	Please see Local Identifier .	30 31 32 33 34
Link	A full duplex transmission path between any two network fabric elements, such as Channel Adapters or Switches .	32 33 34 35 36 37 38 39 40 41 42
Link Heart-beat Ordered-Set	The sixteen symbol ordered-set composed of a COM symbol, Lane Identifier , three D1.2 data symbols, an OpCode symbol, a reserved symbol, a PortNum symbol (only used for switch ports), and an eight-symbol Globally Unique ID (GUID), used for ensuring link aliveness and determining link round-trip latency.	35 36 37 38 39 40 41 42
Link Training	Link Training is the process of establishing link synchronization between two Link Endpoints. The Link Training State Machine controls the transition between the Link Down and Link UP state. This process includes but is not limited to:	40 41 42

1)	Bit synchronization	1
2)	Symbol synchronization	2
3)	Width and Speed negotiation	3
4)	Physical Lane ordering	4
5)	Physical Lane polarity	5
6)	Training Sequence handshake	6
7)	Error recovery	7
Local Identifier	An address assigned to a port by the Subnet Manager , unique within the subnet, used for directing packets within the subnet. The Source and Destination LIDs are present in the Local Route Header .	9
Local Route Header	Routing header present in all InfiniBand™ Architecture packets, used for routing through switches within a subnet.	10
Logic Ground	The voltage to which the logic signals (high speed and low speed) are referenced. This is synonymous with Signal Ground .	11
Longitudinal Airflow	As it pertains to the IB Module , longitudinal airflow is defined as airflow across the module predominantly in the direction normal to the backplane and connector housing.	12
LRH	Please see Local Route Header .	13
2.4 M TO P		14
M_Key	Please see Management Key .	15
MAD	Please see Management Datagram .	16
Managed Unit	A Unit which provides Module Information about itself to an external source.	17
Management Datagram	Refers to the contents of an Unreliable Datagram packet used for communication among the HCAs , Switches , Routers , and T to Zs to manage the network. InfiniBand™ Architecture describes the format of a number of these management commands.	18
Management Key	A construct that is contained in IBA Management Datagrams to authenticate the sender by the receiver.	19
Management Proxy	A Unit which serves as a means to allow for system management operations to get from one “band” to the other (i.e. an InfiniBand™ Link to IB-ML or IB-ML to an InfiniBand™ link)	20
MB/s	Mega-bytes per second (10^6 bytes per second)	21
Message	A transfer of information between two or more Channel Adapters that consists of one or more packets.	22

MN	Please see Modal Noise .	1
Modal Noise	Noise in a laser based optical communication system caused by the incomplete collection of the spatially correlated interference pattern.	2
Mode-Partition Noise	Noise in a laser based optical communication system caused by the changing distribution of laser energy partitioning itself among the laser modes (or lines) on successive pulses in the data stream. The effect is a different center wavelength for the successive pulses resulting in arrival time jitter attributable to chromatic dispersion in the fiber.	3
Mode-Partition Noise k Factor	Empirically derived factor linking mode-partition noise to system penalty.	4
Modifiers	In a verb definition, the list of input and output objects that specify how, and on what, the verb is to be executed.	5
Module Information	Information provided to Management Software about an IB Module .	6
MPN	Please see Mode-Partition Noise .	7
MPN-k	Please see Mode-Partition Noise k Factor .	8
Numerical Aperture	The sine of the radiation or acceptance half angle of an optical fiber, multiplied by the refractive index of the material in contact with the exit or entrance face.	9
O/E	Optical/Electrical.	10
OFSTP	Please see Optical Fiber System Test Practice .	11
OMA	Please see Optical Modulation Amplitude .	12
ORL	Please see Optical Return Loss .	13
Operating System Vendor	The software manufacturer of the operating system that is running on the node under discussion.	14
Optical Cable Plant	All passive communications elements (e.g., optical fiber, connectors, splices, etc.) between a transmitter and a receiver.	15
Optical Connector	An optical connector connects the optical media to the receptacle of an optical transmitter, to the receptacle of an optical receiver, or to a fiber optic adapter.	16
Optical Connector Loss	The optical power lost between two optical connectors mated in a fiber optic adapter.	17

Optical/Electrical Converter	Also referred to as “ O/E Converter”. A device that converts electrical signals on a board to/from IB-compliant optical signals. It optionally contains CDR and/or de/MUX functionality, but it does not contain data buffers and is not protocol aware. The electrical signals into / out of an Optical/Electrical Converter may be vendor-specific.	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
Optical Eye Opening	For a fiber optic link, the time interval across the eye, measured at the 50% normalized eye amplitude which is error free to the specified BER.	
Optical Fiber	Any filament or fiber, made of dielectric material, that guides light.	
Optical Fiber System Test Practice	Standards developed and published by the EIA/TIA under the EIA/TIA-526 series of standards. This term is also referred to as OFSTP .	
Optical Modulation Amplitude	The absolute difference between the optical power of a logic one level and the optical power of a logic zero level. This term is abbreviated OMA .	
Optical Passive Loss	The insertion loss resulting from connections (adapters or splices), and attenuation attributable to the fiber cable plant.	
Optical Receiver	The part of an Optical/Electrical Converter which receives an optical signal.	
Optical Receiver Bandwidth	High frequency 3dB roll-off frequency of the optical receiver.	
Optical Receiver Overload	The maximum acceptable value of the received average optical power at the receiver input to achieve the specified BER.	
Optical Receiver Sensitivity	The minimum acceptable value of average received signal, at the defined optical test point, to achieve the specified BER . It takes into account power penalties caused by use of a transmitter with a worst-case output. In the case of an optical path it does not include power penalties associated with Dispersion , Jitter , effects related to the modal structure of the source or reflections from the optical path. These effects are specified separately in the allocation of maximum optical path penalty.	
Optical Receptacle	The part of the Optical Transmitter or Receiver into which an optical connector plugs.	
Optical Return Loss	The ratio (expressed in units of dB) of optical power incident upon a component port or an assembly to the optical power reflected by that component when that component or assembly is introduced into a link or system. This term is abbreviated ORL .	

Optical Stressed Receiver Sensitivity	Minimum receiver sensitivity required in order to perform PLL locking on the data. This measure takes a transmitter with worst-case output, and adds the channel loss and the EYE losses. The stressed EYE is used principally used to know how the jitter affect the system and how low the low pass filter has to be in the receiver.	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
Optical System Penalty	An optical link penalty to account for those effects other than optical passive loss.	
Optical Transceiver	A device that converts IB-compliant electrical signals on a board to/from IB-compliant optical signals. An optical transceiver may be 1x, 4x or 12x wide. In general, an Optical Transceiver will contain one or more Optical/Electrical Converters , CDR 's, and data buffers.	
Optical Transmitter	The part of an Optical/Electrical Converter which transmits an optical signal.	
Ordered Set	Please see Control Ordered-set .	
Out-of-band Management	Management messages which traverse a transport other than the InfiniBand™ fabric.	
P_Key	Please see Partition Key .	
Packet	The indivisible unit of IBA data transfer and routing, consisting of one or more headers, a Packet Payload , and one or two Common Completion Entries.	
Packet Data	Data symbols transmitted as part of a Data Packet Payload or a Link Packet Payload. Packet data is delimited by start symbols SDP and SLP and terminated by end symbols EGP and EBP.	
Packet Payload	The portion of a Packet between (not including) any Transport header(s) and the CRCs at the end of each packet. The packet payload contains up to 4096 bytes.	
PAD	The PAD symbol is transmitted only on a 12x link to maintain packet framing alignment.	
Partition Key	A value carried in packets and stored in Channel Adapters that is used to determine membership in a partition.	
Partition Manager	The entity that manages partition keys and membership.	
Passively Managed Chassis	A Passively Managed chassis provides an InfiniBand™ specified GUID and physical Slot Information to every IB Module on the Modules unique IB-ML .	
PCB	Printed Circuit Board	

PD	Photodiode, used for converting optical signals to electrical at a receiver.	1
PDF	Probability Distribution Function	2
Permanent Errors	A permanent error is surfaced when a continuous and irreversible fault is present within an IBA component.	3
Physical Lane	A set of one transmit and one receive differential pairs. A 1x, 4x and 12x link is composed of one, four and twelve physical lanes respectively.	4
Pluggable	A description for a transceiver module which may be unplugged for repair or replacement.	5
PM	Please see Partition Manager .	6
PN	Processor Node	7
Polling	A port state where the transmitter is generating a Beacon Sequence and the receiver is waiting to respond to a Beacon Sequence.	8
Port	Location on a Channel Adapter or Switch to which a link connects. There may be multiple ports on a single Channel Adapter , each with different context information that must be maintained. Switches /switch elements contain more than one port by definition.	9
Port Type 1	An electrical interface type that utilizes the parameters of the original InfiniBand specification (1.0, 1.0a).	10
Port Type 2	An electrical interface type that utilizes more stringent parameters than Port Type 1 . Port Type 2 is currently associated with pluggable devices.	11
Power Management	The ability for an operating system to control the power consumption of InfiniBand™ Architecture compliant devices (Endnode devices and Switches).	12
Power Sub-system Management	The ability to monitor and control the power supplies of a system or chassis.	13
PRBS	Please see Pseudo Random Binary Sequence .	14
Processor Node	One or more general-purpose processors running under a single cache coherent memory model that uses a HCA to connect a processor bus to one or more Ports. Additionally, a PN is capable of performing initialization and configuration.	15
Proxy Managed Repeater	A Repeater which must have a proxy in-place to respond to GetInfo MAD operations but allows IB-ML access to defined VPD information. All repeater boards fulfill this definition.	16

Proxy Managed TCA	A TCA which must have a proxy in-place to respond to GetInfo MAD operations but allows IB-ML access to defined VPD information. This type of TCA is not defined by the InfiniBand™ specification.	1 2 3
Proxy Managed Unit	Any Unit which must have a proxy in-place because Baseboard information is not accessible through its native InfiniBand™ fabric link(s).	4 5
PSE	Protocol State Engine (Phy, Link, Frame)	6 7
Pseudo Random Binary Sequence	A binary sequence of bits which is deterministically generated, but approximates the distribution of a randomly-generated bit sequence. Various lengths of PRBS sequences (PRBS9, PRBS11, PRBS23, PRBS31) are used, e.g., for transmitter and receiver testing.	8 9 10 11 12
2.5 Q TO S		
QP	Please see Queue Pair .	13 14
Queue Pair	Consists of a Send and a Receive Work Queue. Send and receive queues are always created as a pair and remain that way throughout their lifetime. A Queue Pair is identified by its Queue Pair Number .	15 16 17
Queue Pair Number	Identifies a specific Queue Pair within a Channel Adapter.	18 19 20
RDETH	Reliable Datagram Extended Transport Header.	21 22
Receive Queue	One of the two queues associated with a Queue Pair . The receive queue contains Work Queue Elements that describe where to place incoming data.	23 24 25
Relative intensity noise	Laser noise in dB/Hz with 12 dB optical return loss, with respect to the optical modulation amplitude.	26 27 28
Retiming Repeater	A device which recovers and retransmits data, using a local oscillator to eliminate jitter transfer, and hence perform a jitter reset. (SKIP ordered-set dependent)	29 30
RIN12-OMA	Please see Relative intensity noise .	31 32
Rise Time	The time interval for the rising edge of a pulse to transition from its 20% amplitude level to its 80% amplitude level.	33 34
RJ	Please see Jitter, Random .	35 36
RMS	Root Mean Square.	37 38
RNR Nak	Receiver Not Ready. A response signifying that the receiver is not currently able to accept the request, but may be able to do so in the future.	39 40 41 42

Router	A device that transports packets between IBA subnets.	1
Run Length	Maximum number of consecutive identical bits in the transmitted signal e.g., the pattern 001111010 has a run length of five (5).	2
Running Disparity	A binary parameter indicating the cumulative disparity (positive or negative) of all previously issued bits.	3
SA	Please see Subnet Administration .	4
SDP	Please see Start of Data Packet Delimiter .	5
Send Queue	One of the two queues of a Queue Pair . The Send queue contains WQEs that describe the data to be transmitted.	6
Server	<ol style="list-style-type: none"> 1) The passive entity in a connection establishment exchange. 2) An entity (e.g., a process) that provides services in response to requests from clients. 3) The class of computers that emphasize I/O connectivity and centralized data storage capacity to support the needs of other, typically remote, client computers. 	7
SGID	Source Globally Unique Identifier .	8
Signal Ground	Please see Logic Ground .	9
Skew	The timing difference between Physical Lanes as measured using the zero crossings of the differential voltage of the COM s present in the training sequences TS1 and TS2 or in the SKIP ordered-set .	10
SKP	The SKP symbol is transmitted as part of a SKIP ordered-set .	11
SKIP ordered-set	The SKIP ordered-set consists of a COM symbol followed by three SKP symbols when initially transmitted. It may consist of a COM symbol followed by one to five SKP symbols when received.	12
Sleeping	A port state where the transmitter is quiescent and the receiver is waiting to respond to a Beacon Sequence .	13
SLID	Source Local Identifier	14
Slot	An InfiniBand™ specified volumetric envelope with a specified backplane connector into which one of the defined IB Modules plug.	15
Slot ID	A Slot Designation provided in Slot Information	16

Slot Information	Information provided by the Chassis about individual Slots	1
SLP	Please see Start of Link Packet Delimiter .	2
SM	Please see Subnet Manager .	3
SMA	Please see Subnet Management Agent .	4
SMA GUID	The Globally Unique Identifier common to all resources within the scope of a single Subnet Management Agent .	5
SMD	Please see Subnet Management Data .	6
SMP	Please see Subnet Management Packet .	7
Solicited Event	A facility by which a message sender may cause an event to be generated at the recipient when the message is received.	8
Spectral Width	The weighted root mean square width of the optical spectrum (see FOTP-127).	9
Start of Data Packet Delimiter	The Start of Data Packet Delimiter symbol is transmitted to identify the start of a end-to-end data packet.	10
Start of Link Packet Delimiter	The Start of Link Packet Delimiter symbol is transmitted to identify the start of a link control packet.	11
Stressed Receiver	In a Stressed Receiver test the worst case transmitter eye opening is applied to the optical receiver under test. The eye closure typically is generated with a combination of coaxial cable and/or B4th orderessel-Thomson filter driving a laser diode directly.	12
Stress Receiver ISI Penalty	Vertical eye closure penalty due to ISI.	13
Stressed Receiver DCD component of DJ	Horizontal eye closure caused by Data Dependent Pulse Width Shrinkage measured at average optical power.	14
Subnet	A set of InfiniBand™ Architecture Ports , and associated links, that have a common Subnet ID and are managed by a common Subnet Manager . Subnets may be connected to each other through routers.	15
		16
		17
		18
		19
		20
		21
		22
		23
		24
		25
		26
		27
		28
		29
		30
		31
		32
		33
		34
		35
		36
		37
		38
		39
		40
		41
		42

Subnet Ad- ministration	The architectural construct that implements the interface for querying and manipulating subnet management data.	1
Subnet Man- ager	One of several entities involved in the configuration and control of the subnet. Active Subnet Manager: Any subnet manager currently exercising control over all or part of the subnet. An active subnet manager may be the master subnet manager, or an alternate subnet manager acting on the behalf of the master. This is sometimes referred to as the formal subnet manager. Alternate Subnet Manager: Any subnet manager that is acting on the behalf of the master subnet manager, but is not the master subnet manager. Master Subnet Manager: The subnet manager that is authoritative, that has the reference configuration information for the subnet. Standby Subnet Manager: A subnet manager that is currently quiescent, and not in the role of a master SM, by agency of the master SM. Standby Isms are dormant managers.	2
Subnet Man- agement Agent	An entity present in all IBA Channel Adapters and Switches that processes Subnet Management Packets from Subnet Manager(s) .	3
Subnet Man- agement Data	Vital Product Data required by the Subnet Manager .	4
Subnet Man- agement Packet	The subclass of Management Datagrams used to manage the subnet. SMPs travel exclusively over Virtual Lane 15 and are addressed exclusively to Queue Pair Number 0.	5
Surprise Hot Removal	The removal of an IB Module from a backplane that has both V _{Bulk} and V _{Aux} present without first being placed in a quiescent state.	6
Switch	A device that routes packets from one link to another of the same Subnet , using the Destination Local Identifier field in the Local Route Header.	7
Switch Man- agement Port	A virtual port by which a Switch may be managed.	8
Symbol Time	The transmit time for 1 symbol. With 8b/10b encoding, a symbol is 10 bits long; thus, a symbol time is 10*UI (e.g., 4 ns for 2.5Gb/s signaling, and 1 ns for 10Gb/s signaling).	9
2.6 T TO Z		10
TCA	Please see Target Channel Adapter .	11
Target Channel Adapter	A Channel Adapter typically used to support I/O devices. TCAs are not required to support the Verbs interface. See also I/O Unit .	12

Termination mismatch	The difference between the low frequency single-ended impedances of the two legs of a differential pair.	1 2 3 4 5 6 7 8 9 10 11 12 13 14
TJ	Please see Jitter, Total	15 16 17 18 19 20 21
Training Sequence 1	The sixteen symbol ordered-set which consists of a COM symbol, Lane Identifier and fourteen D10.2 data symbols.	22 23 24 25 26 27 28
Training Sequence 2	The sixteen symbol ordered-set which consists of a COM symbol, Lane Identifier and fourteen D5.2 data symbols.	29 30 31 32 33 34 35
Training Sequence 3	The sixteen symbol ordered-set which consists of a COM symbol, Lane Identifier , six D13.2 data symbols, a bit map of the active speeds, a bit map requesting transmitter driver de-emphasis and/or link heartbeat enabling, a symbol describing the driver de-emphasis setting which should be used, and five reserved symbols. TS3 is used for negotiating link configuration between two peer ports on a link.	36 37 38 39 40 41 42
Training Sequence for Test	The sixteen symbol ordered-set which is used by test equipment to place a port's Link/Phy state machine into one of several states used for testing transmitter and receiver compliance with physical layer specifications.	43 44 45 46 47 48 49
Transparent Retiming Repeater	A device that recovers and retransmits data, to eliminate jitter transfer, and hence perform a jitter reset. (SKIP symbols are not added or deleted to SKIP ordered-set)	50 51 52 53 54 55 56
Transport Service Type	Describes the reliability, sequencing, message size, and operation types that will be used between the communicating Channel Adapters .	57 58 59 60 61 62 63
	Transport service types that use the IBA transport and that pertain to Volume 2 are:	64 65 66 67 68 69 70
	• Unreliable Datagram	71 72 73 74 75 76 77
	See Volume 1 for other Transport Service Types.	78 79 80 81 82 83 84
Transverse Airflow	As it pertains to the IB Module , transverse airflow is defined as airflow that might predominantly enter the module parallel to the backplane. In a vertical module orientation this direction would be an upper or lower entrance point. Transverse airflow may be directed longitudinally within the module, but the predominant entrance and/or exitpoint for the airflow is in a direction parallel to the backplane.	85 86 87 88 89 90 91
TS1	Please see Training Sequence 1 .	92 93 94 95 96 97 98
TS2	Please see Training Sequence 2 .	99 100 101 102 103 104 105
TS3	Please see Training Sequence 3 .	106 107 108 109 110 111 112
TS-T	Please see Training Sequence for Test .	113 114 115 116 117 118 119

UI	Please see Unit Interval .	1
Unit	One or more sets of processes and/or functions attached to the fabric by one or more channel adapters. Please see Host and I/O Unit .	2
Unit Interval	The time interval between possible transitions on a lane; the inverse of the signaling rate.	3
Unmanaged Chassis	An Unmanaged chassis does not implement any IB-specified GUID or Slot Information and leaves the Module IB-ML unconnected.	4
Unreliable Datagram	A Transport Service Type in which a Queue Pair may transmit and receive single-packet messages to/from any other QP. Ordering and delivery are not guaranteed, and delivered packets may be dropped by the receiver.	5
VA_In	The Auxiliary Voltage Input pin (5V nominal) that is defined on the IB Module edge connector and backplane connector.	6
VB_In	The Bulk Voltage Input pins (12V nominal) that are defined on the IB Module edge connector and backplane connector.	7
Verbs	An abstract description of the functionality of a Host Channel Adapter . An operating system may expose some or all of the verb functionality through its programming interface.	8
Virtual Lane	A method of providing independent data streams on the same physical link.	9
Virtual Memory	The address space available to a process running in a system with a memory management unit (MMU). The virtual address space is usually divided into pages, each consisting of 2^{**N} bytes. The bottom N address bits (the offset within a page) are left unchanged, indicating the offset within a page, and the upper bits give a (virtual) page number that is mapped by the MMU to a physical page address. This is recombined with the offset to give the address of a location in physical memory	10
Vital Product Data	Device-specific data to support management functions.	11
VL	Please see Virtual Lane .	12
VPD	Please see Vital Product Data .	13
Wake Request Event	Events that can be produced by an IB Module that desires to return from a Sleeping or Polling state to an operation state.	14
Wander	Deviations at <10 kHz from the ideal timing of an event.	15
Work Queue Element	The Host Channel Adapter 's internal representation of a Work Request . The consumer does not have direct access to Work Queue Elements .	16

Work Queue Pair	Please see Queue Pair .	1
Work Request	The means by which a consumer requests the creation of a Work Queue Element .	2
Workstation, or Client Computer	The class of computers that emphasize numerical and/or graphic performance and provide an interface to a human being.	3
WRE	Please see Wake Request Event	4
		5
		6
		7
		8
		9
		10
		11
		12
		13
		14
		15
		16
		17
		18
		19
		20
		21
		22
		23
		24
		25
		26
		27
		28
		29
		30
		31
		32
		33
		34
		35
		36
		37
		38
		39
		40
		41
		42

CHAPTER 3: PHYSICAL LAYER OVERVIEW

3.1 INTRODUCTION

This volume defines the low level physical interface protocols, electrical and mechanical specifications for developing applications based on the InfiniBand Architecture. The InfiniBand Architecture supports a range of applications from the backplane interconnect of a single host, to a complex system area network consisting of multiple independent and clustered hosts and I/O components.

In keeping with the layered nature of the InfiniBand Architecture, [Figure 1](#) depicts the structure within the Physical Layer itself.

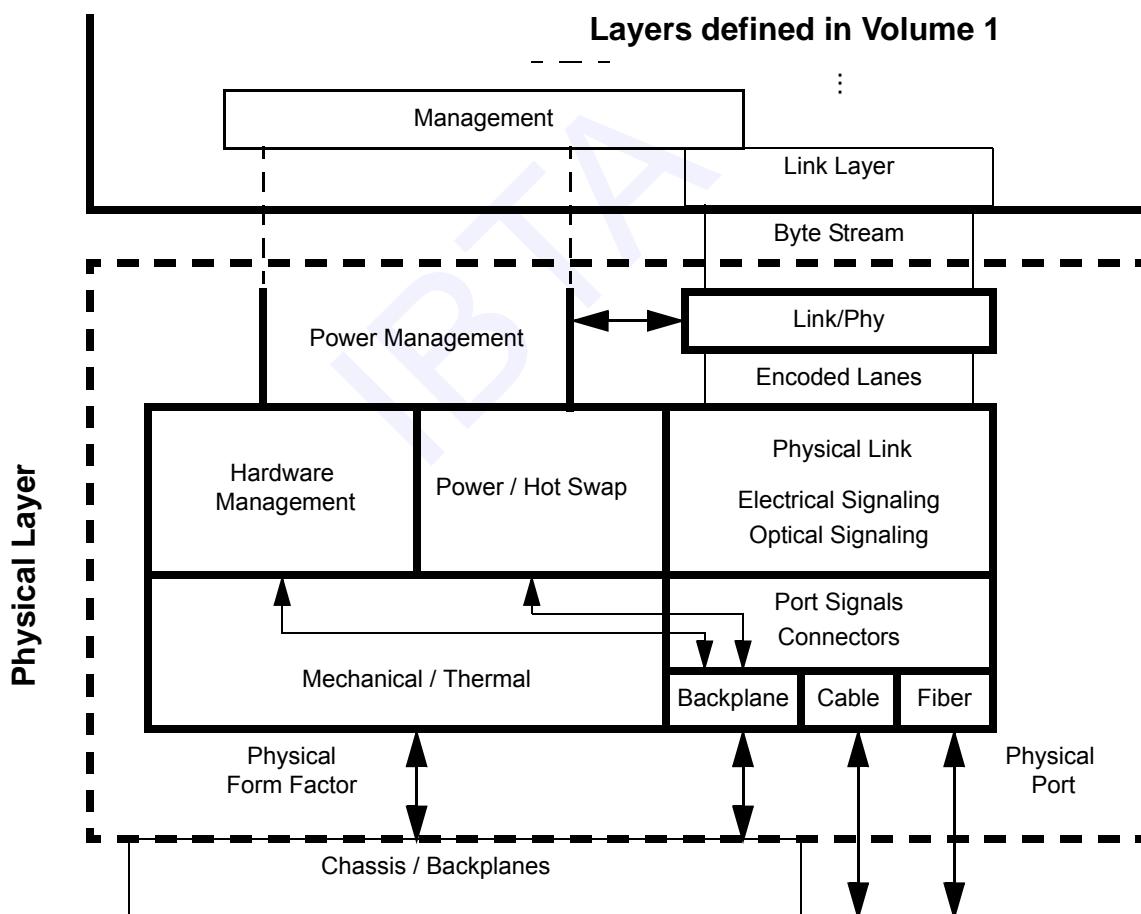


Figure 1 Physical Layer Structure

The basic interconnect of the InfiniBand Architecture is a link (or “physical link”), which is a full duplex transmission path between any two fabrics elements. A fabric is a collection of links, switches, repeaters and routers that connects a set of end nodes. A link physically terminates at a port.

The physical attach point to a port is either:

- 1) A Cable Connector, which is defined for use for copper cables.
- 2) A Fiber Connector, which is defined for use for optical cables.
- 3) A Backplane Connector, which is defined for accepting a specified form factor that houses a function (Channel Adapter, Switch, etc.).

For the form factors, elements of Power and Hardware Management are also specified.

The remaining chapters in this volume provide the detailed specifications to be met in order for devices to properly function on an InfiniBand fabric.

3.2 PHYSICAL PORT

A physical “Port” is a set of signals as seen on a connector interface identified by this specification. The physical ports defined are:

- Backplane Port
- Cable Port
- Fiber Optic Port

Some physical ports contain all signals (e.g. backplane) while others contain a subset (e.g. cable and optics).

A physical port consists of up to four groups of signals which serve different purposes. These groups are:

- Signaling Group
- Hardware Management Group
- Bulk Power Group
- Auxiliary Power Group

[Figure 2](#) depicts a Backplane Port and the signal groupings containing a single physical link made of a number of physical lanes dependent on the link width. Similarly, [Figure 3](#) and [Figure 4](#) depict a Cable Port and a Fiber Optic Port, respectively.

See [Chapter 4: Port Signal Definitions](#) for further details on these port types.

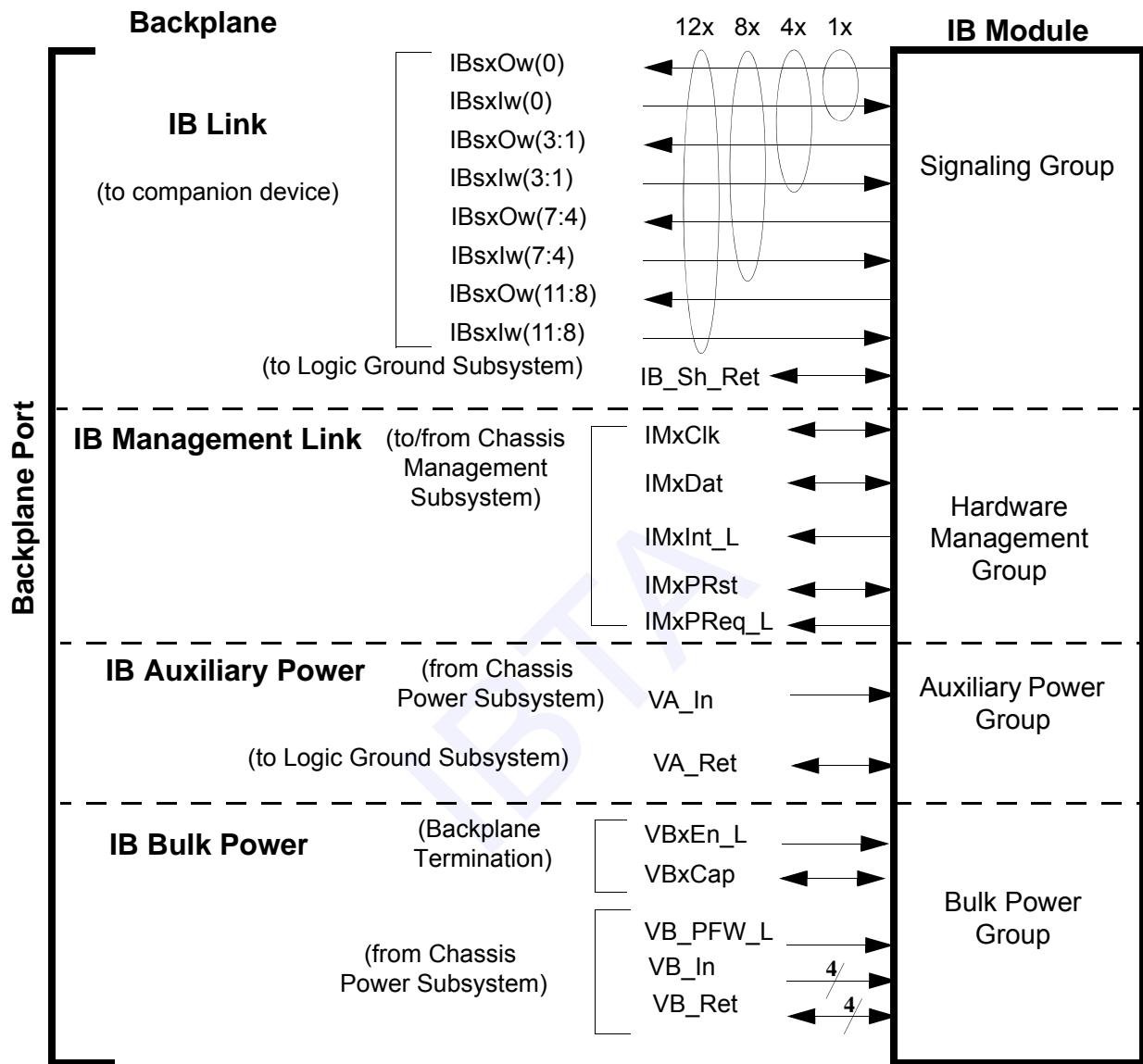


Figure 2 Backplane Port - Single Physical Link

Implementation Note

IB Management Link, IB Auxiliary Power, and IB Bulk Power signals are features which were deprecated in Rel. 1.3. They may not be compatible with current InfiniBand architecture features and functions.

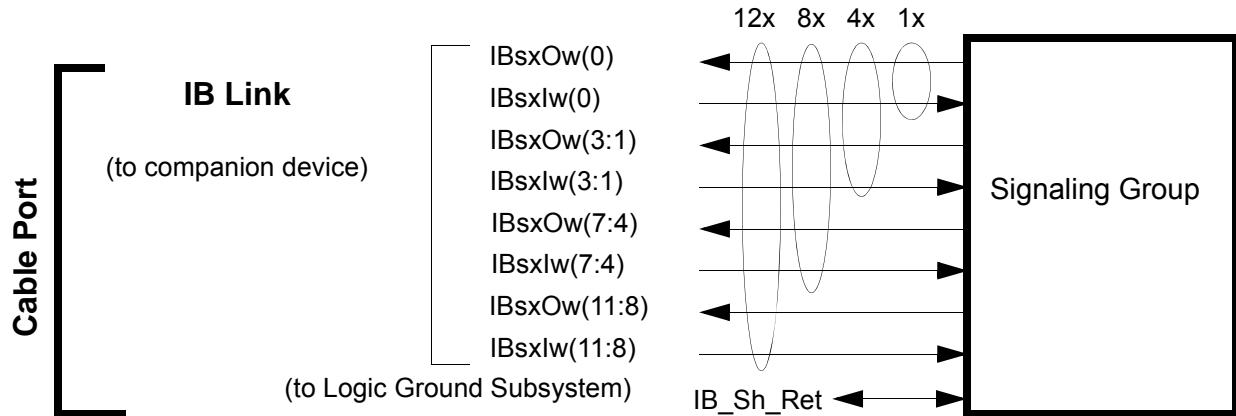


Figure 3 Cable Port

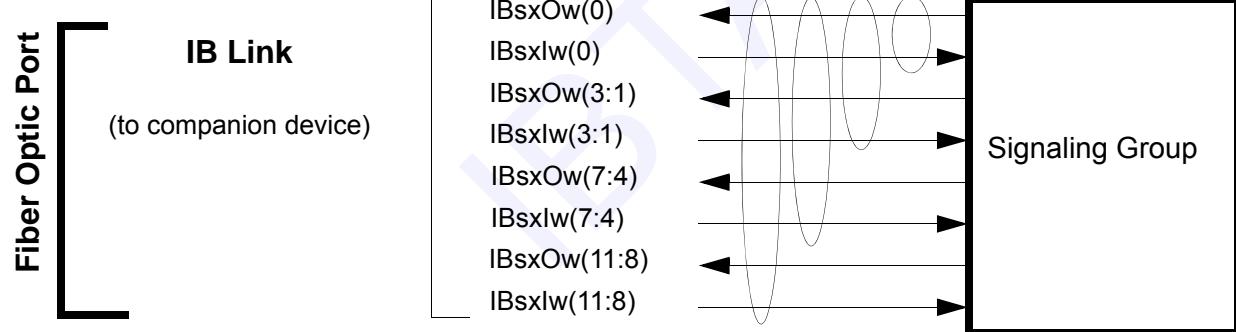


Figure 4 Fiber Optic Port

3.2.1 MULTI-PORTING

This version of the InfiniBand specification supports the utilization of the backplane connector on the defined form factors in a “multi-ported” manner. As the physical connector, as defined in [Volume 2-DEPR: Chapter 2: Backplane Connector Specification](#), supports up to 12 physical lanes, it is possible to implement modules with multiple IBA links (i.e. ports) using only a single connector. Since some of the form factors defined in [Volume 2-DEPR: Chapter 1: Mechanical Specification](#) allow for the presence of multiple physical connectors, it is additionally possible to have multiple links on each of the physical connectors present.

Figure 5 depicts a Backplane Port containing multiple independent physical links in a single physical connector; this is also referred to in this specification as being “multi-ported”.

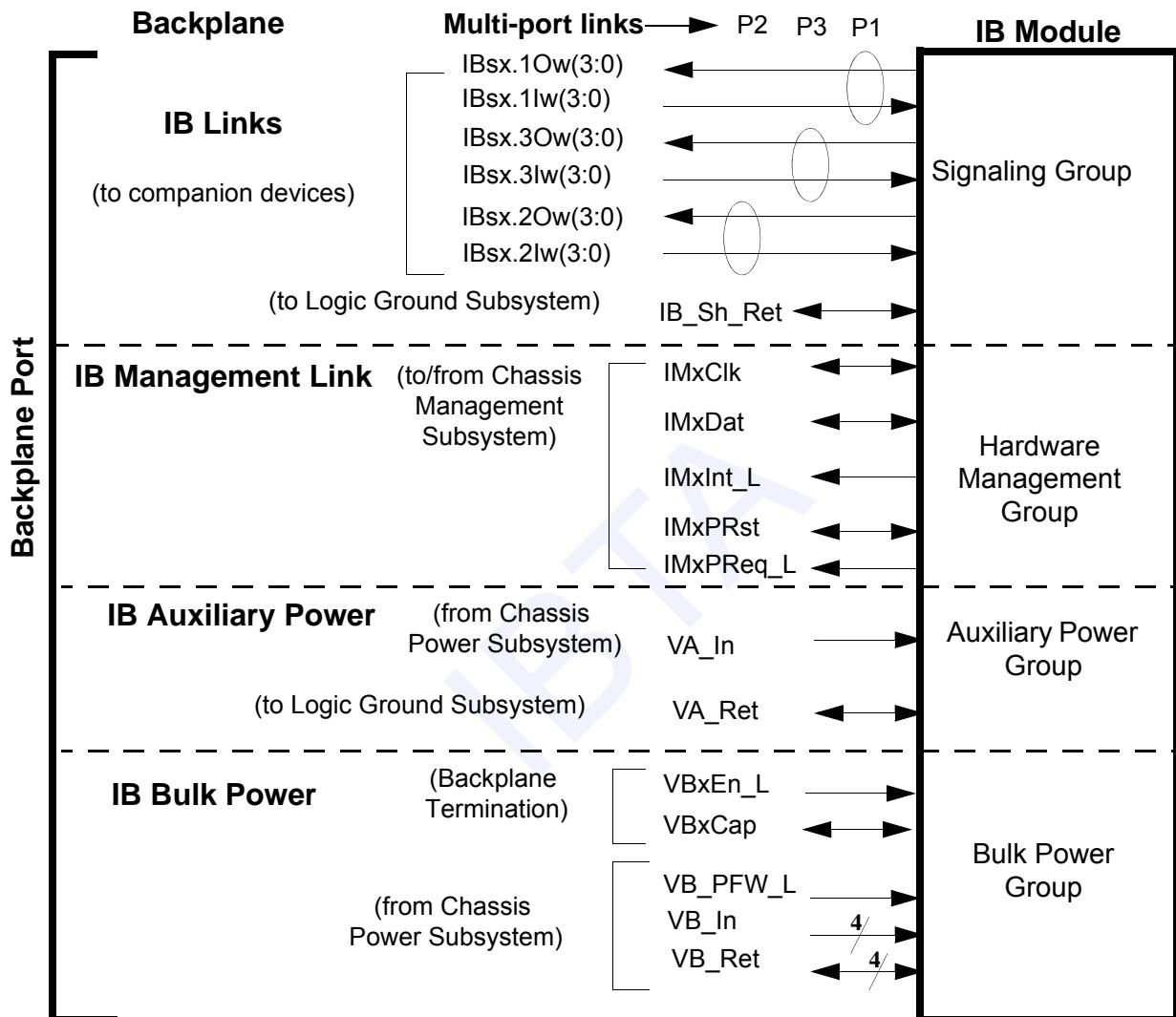


Figure 5 Backplane Port - Multiple Physical Links

Implementation Note

IB Management Link, IB Auxiliary Power, and IB Bulk Power signals are features which were deprecated in Rel. 1.3. They may not be compatible with current InfiniBand architecture features and functions.

A similar configuration of multiple 4x physical links in a 12x connector applies for cables also, as described in [Section 7.9.4.7, “12X to 3-4X microGigaCN passive cables,” on page 42](#).

[page 462](#) and [Section 7.9.6.2, "12x CXP to 3-4x cables." on page 476.](#)

3.2.2 ACTIVE CABLES

Release 1.2 and following releases of this specification allow provision of power and management interface signals in cable connectors, to allow incorporation of active components into cables. Various types of active cables are described in [Section 6.2.2, "Cable Types." on page 270.](#)

3.3 LINK ELECTRICAL SIGNALING

This specification defines the characteristics required to communicate between the output of a port of one InfiniBand node and the input of a port of another InfiniBand node using copper printed wiring and passive or active cabling. The signaling rate for encoded data on the media for the original release is 2.5 Gb/s which results in a data rate that can be considered to be 250 MB/s per physical lane, at the "SDR" (Single Data Rate) speed. Additional enhanced signaling rates defined in Release 1.2 and 1.2.1 are 5.0 Gb/s for the "DDR" (Double Data Rate) speed, and 10.0 Gb/s for the "QDR" (Quad Data Rate) speed. Additional enhanced signaling rates defined in Release 1.3 and 1.3.1 are 14.0625 Gb/s for the "FDR" (Fourteen Gigabit per second Data Rate) speed, and 25.78125 Gb/s for the "EDR" (Extended Data Rate) speed. The connections for all rates are point to point and signaling is full duplex, unidirectional.

See [Chapter 6: High Speed Electrical Interfaces](#) for details of SDR (2.5Gb/s), DDR (5.0Gb/s), QDR (10.0 Gb/s), FDR (14.0625 Gb/s), and EDR (25.78125 Gb/s) operation. See [Chapter 7: Electrical Connectors for Modules and Cables](#) for details of the copper cable and connectors used for electrical links. See [Chapter 8: Management Interface](#) for details of the interface for management of cables.

3.4 LINK OPTICAL SIGNALING

This specification defines the characteristics required to communicate between the output of a port of one InfiniBand node and the input of a port of another InfiniBand node using optical fiber. The only signaling rate for encoded data on the media for prior releases was 2.5 Gb/s which results in a data rate that can be considered to be 250 MBytes/second per physical lane, at the "SDR" (Single Data Rate) speed. Additional enhanced signaling rates described in Release 1.2 are 5.0 Gb/s for the "DDR" (Double Data Rate) speed, and 10.0 Gb/s for the "QDR" (Quad Data Rate) speed. The connections are point to point and full duplex, unidirectional.

See [Chapter 9: Fiber Attachment - 2.5 Gb/s, 5.0 Gb/s, & 10 Gb/s](#) for specification details of SDR (2.5 Gb/s), DDR (5.0 Gb/s) and QDR (10 Gb/s) operation over optical fiber.

3.5 LINK PHYSICAL LAYER

The Link Physical layer provides the interface between the packet byte stream of Link Layer defined in *InfiniBand Architecture Specification, Volume 1* and the serial bit stream(s) of the physical media. The packet byte stream will be byte striped across the available physical lanes. The byte stream on each physical lane is encoded using either 8b/10b encoding, for SDR, DDR, and QDR rates, or 64b/66b encoding, for FDR and EDR

rates. In addition to encode and decode, the link physical layer includes link training and initialization logic, clock tolerance compensation logic, receive error detecting logic, link heartbeat logic, and transmitter and receiver testing functionality. This layer is described in [Chapter 5: Link/Phy Interface](#).

3.5.1 SPEED NEGOTIATION

The original release of the InfiniBand Architecture supported only 2.5Gb/s signaling. However, provisions were made for potential bit rate (speed) increases in a later versions of this specification. The 1.2 release specified negotiation for the use of SDR (2.5Gb/s), DDR (5.0 Gb/s) and QDR (10.0Gb/s) speeds. The 1.3 release specifies negotiation for for the use of FDR (14 Gb/s) and EDR (25 Gb/s) speeds.

The link speed negotiation mechanism is part of the Link Initialization and Training process. This process provides for two nodes to determine the speed of the interface that will allow for operation at the maximum frequency architecturally supported that is achievable based on endnode capability and interconnect signal integrity.

3.5.2 WIDTH NEGOTIATION

The link width negotiation mechanism is also part of the Link Initialization and Training. This process provides for two nodes to determine the width of the interface that will allow for the maximum bandwidth that is achievable based on endnode capability and interconnect.

This specification defines four interface widths as follows:

- 1) 1x
 - Electrical: 2 differential pair, 1 per direction for a total of 4 wires
 - Optical: 1 transmit/1 receive per direction for a total of 2 fibers
- 2) 4x
 - Electrical: 8 differential pair, 4 per direction for a total of 16 wires
 - Optical: 4 transmit / 4 receive per direction for a total of 8 fibers
- 3) 8x
 - Electrical: 16 differential pair, 8 per direction for a total of 32 wires
 - Optical: 8 transmit / 8 receive per direction for a total of 16 fibers
- 4) 12x
 - Electrical: 24 differential pair, 12 per direction, for a total of 48 wires
 - Optical: 12 transmit / 12 receive per direction for a total of 24 fibers

3.6 MODULE MECHANICAL

Implementation Note

This section describes features which were deprecated in Rel. 1.3. They may not be compatible with current InfiniBand architecture features and functions.

An InfiniBand Module conforms to any of the defined form factors in [Volume 2-DEPR: Chapter 1: Mechanical Specification](#). A module consists of the following:

- a) The module carrier - the basic metal structure
- b) The module ejector & latch - a handle and latch
- c) The module cover - easily removable cover to protect the board(s)
- d) The board(s) which implement the module's functionality

This version of the specification defines two module heights (tall and standard) and two module widths. Double width modules occupy two IB defined chassis slots.

As double wide standard and both forms of tall modules (single and double wide) allow for the presence of more than one physical connector, a nomenclature is needed to refer to the connector positions.

Every module contains a connector positioned for the primary slot - primary being defined as the left-most slot the module covers. As shown in [Volume 2-DEPR: Chapter 1: Mechanical Specification](#), a tall module allow for two connector locations: one in the lower portion and one in the upper portion; lower and upper in this context is when the module is in the vertical orientation. For consistency between standard and tall module designations, the lower (for tall) or only (for standard) location is termed "Primary" and designated "C" by convention. The upper location (for tall) is designated "O" (for optional).

Architectural Note

The most obvious designation for "Primary" would be "P". However, "P" is used in this specification to designate "Ports". The designation of "C" indicates "Primary Connector".

For the case of a standard module, the only connector present is designated C1.

For the case of standard wide modules, there may be a second connector positioned for the adjacent slot - adjacent being defined as the slot to the right of the left-most slot the module covers. The two connectors, both in the "C" location, are designated C1 and C2.

For the case of a tall module, as noted previously, there is a primary connector designated C1 and may optionally have a connector in the upper location designated O1.

For the case of a tall wide module, in addition to the low there may optionally be connectors in the upper locations for both the primary and adjacent positions. The two upper connectors, both in the upper "O" locations are designated O1 and O2.

Within any of the physical connector locations, up to three (3) link ports is defined. The pin designation for these ports is defined in [Volume 2-DEPR: Chapter 2: Backplane Connector Specification](#).

[Figure 6](#) depicts these designations.

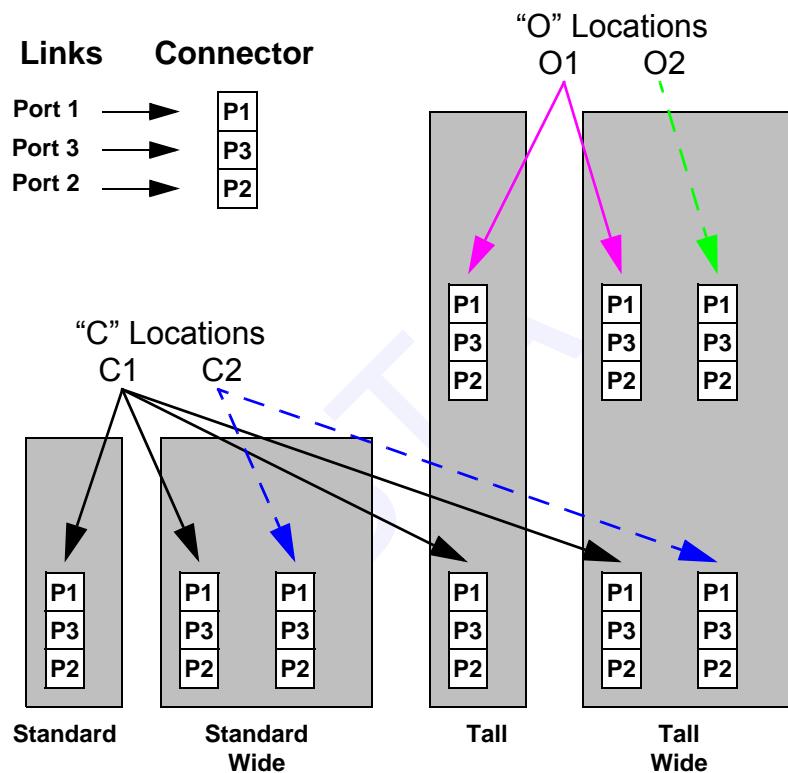


Figure 6 Module Connector Location Designation

3.7 CHASSIS SLOT MECHANICAL

Implementation Note

This section describes features which were deprecated in Rel. 1.3. They may not be compatible with current InfiniBand architecture features and functions.

A slot is the volumetric envelope with a specified backplane connector into which one of the defined InfiniBand Modules plug. This specification provides for the enablement of 3U and 6U chassis implementations with modules oriented vertically or horizontally.

Volume 2-DEPR: Chapter 1: Mechanical Specification provides details for system and chassis designers to implement slots which will accommodate the module form factor options. Only single width slots are physically defined; double width modules occupy two single width slots.

As only single width slots are defined, the connector location designation only indicate whether a connector is in the lower position (for standard and tall) designated "C" or in the upper position (for tall only) designated "O". The additional designation is for the slot number.

As for the module, any of the physical connector locations present may contain up to three (3) link ports. The pin designation for these ports is defined in Volume 2-DEPR: Chapter 2: Backplane Connector Specification.

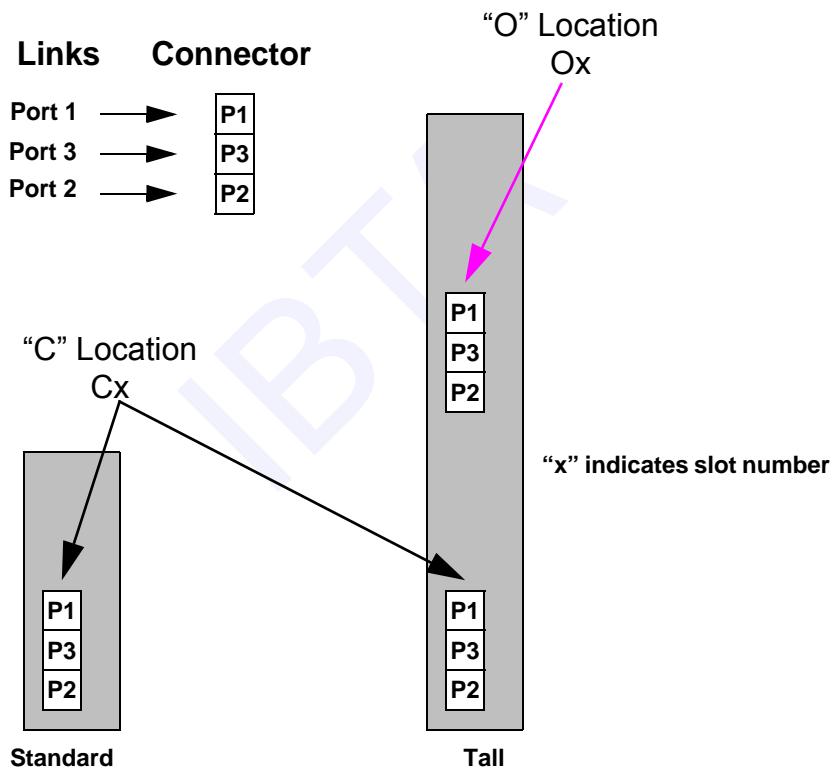


Figure 7 Chassis Connector Location Designation

3.8 POWER

Implementation Note

This section describes features which were deprecated in Rel. 1.3. They may not be compatible with current InfiniBand architecture features and functions.

InfiniBand modules are provided two forms of power:

- Bulk Power to perform the major intended functions of the module. This is supplied at nominally 12V and must be converted on the module by DC-DC converter(s) to produce the voltages required by the module's electronics.
- Auxiliary Power to perform management functions to the module even when the Bulk Power is not available to the module. This is supplied at nominally 5V and may be regulated as necessary by the module.

The combination of connector contact staggering and on-module power control circuitry allow for InfiniBand module to be hot plugged, both inserted and removed, without physical damage. Additionally, features are specified to allow for a "Graceful Removal" operation whereby software is notified of a pending removal such that appropriate actions can take place so as to not disrupt operations. LED indicators on the module indicate when an appropriate state has been achieved so that operationally safe removal can be performed.

Please see [Volume 2-DEPR: Chapter 4: Power / Hot Plug](#) for details on power functionality and [Volume 2-DEPR: Chapter 3: Low Speed Electrical Signaling](#) for details on the electrical parameters of the Bulk and Auxiliary Power groups.

3.9 HARDWARE MANAGEMENT

Implementation Note

This section describes features which were deprecated in Rel. 1.3. They may not be compatible with current InfiniBand architecture features and functions.

Hardware Management describes the functions that manage, control, and monitor physical components of InfiniBand modules and the Chassis in which they reside. Additionally, xCAs and switches that are packaged in some form factor other than those defined by this specification but optionally provide the hardware management functionality defined are described. The functions defined include:

- 1) Communication mechanisms for optionally present Baseboard Manager software running on one or more nodes attached to the InfiniBand fabric,
- 2) Communication mechanisms for an optionally present Chassis Management Entity (CME) which is local managing the physical elements of a chassis,
- 3) Graceful Hot Removal mechanisms,

- 4) Standard visual indicators (LEDs) to assist the user in Hot Add and Hot Removal operations,
- 5) Module Vital Product Data (VPD) accessible to both a Baseboard Manager and a CME,
- 6) Access to module optional environmental variables.

Baseboard Manager software access these facilities using defined datagrams of the Baseboard class on the InfiniBand fabric. The Chassis Management Entity, typically with firmware, access these facilities through the InfiniBand Management Link (IB-ML) interface defined on the standard InfiniBand backplane connector to a module.

Please see [Volume 2-DEPR: Chapter 5: Hardware Management](#) for details on this functionality.

3.10 OPERATING SYSTEM (OS) POWER MANAGEMENT

Implementation Note

This section describes features which were deprecated in Rel. 1.3. They may not be compatible with current InfiniBand architecture features and functions.

Operating System Power Management defines a set of states and facilities that allow for an operating system (or a prescribed agent) to control the power consumption of InfiniBand modules and switches that provide this support. Power management of devices or media not directly attached to the InfiniBand fabric as an addressable node are outside the scope this specification.

The defined states are controlled from the InfiniBand fabric connection of the module. This includes the ability to “power down” to a state only drawing auxiliary power and “power up” from activity on the InfiniBand link. Further, in the presence of a fully operational link, states are defined to allow modules varied levels of power consumption.

Please see [Volume 2-DEPR: Chapter 6: OS Power Management](#) for details on this functionality.

CHAPTER 4: PORT SIGNAL DEFINITIONS

4.1 SIGNAL NAMING CONVENTIONS

The symbol substitution and type notation used for the signal naming within this specification are defined in [Table 2](#) and [Table 3](#).

Table 2 Symbol Substitution

Symbol Convention	Substitution
w in Signal Name	Differential pair distinction - IB Signaling Group Replaced by 'p' for positive rail of differential pair Replaced by 'n' for negative rail of differential pair
x in Signal Name	Replaced by respective board or backplane Port x may range from 1 to n where n is the number of connectors on a board edge or backplane slot
s in Signal Name	For Boards, replaced by 't' for I/O Plate connection, or 'b' for backplane connection. For Backplanes, replaced by slot number. Slots are numbered from 1 to n from left to right, or bottom to top for horizontal mounting

Table 3 Type Notation

Signal Type	Definition
InDiff	Differential input signal - IB signaling levels
OutDiff	Differential output signal - IB signaling levels
ShRet	Shield Return - Shield Ground for both the IB board and the IB backplane
Rx	Optical Receiver Input
Tx	Optical Transmitter Output
PwrU	Pull Up derived from 12V Bulk power input (VB_In)
PwrD	Pull Down to Bulk power return (VB_Ret)
APwrU	Pull Up to Auxilliary power input (VA_In)
APwrD	Pull Down to Auxiliary power return (VA_Ret)
InOut	Open Drain - Wired OR Input/Output pull-up referenced to Auxiliary power (VA_In)
In	Input - logic is powered from Auxiliary power (VA_In)

Table 3 Type Notation (Continued)

Signal Type	Definition	
Out	Output - derived from Auxiliary power (VA_In)	4
PwrIn	12 Volt Input from backplane/system power supply	5
PwrRet	12 Volt Return to backplane/system power supply	6
APwrIn	5 Volt Input from backplane/system power supply	8
APwrRet	5 Volt Return to backplane/system power supply	9

Signal names ending with "_L" indicates the "asserted" or "true" condition is a low voltage.

4.2 PORT SIGNAL SUMMARY

This section summarizes the signals that comprise the physical ports defined by this specification.

4.2.1 BACKPLANE PORT**Implementation Note**

This section describes features which were deprecated in Rel. 1.3. They may not be compatible with current InfiniBand architecture features and functions.

[Table 4](#) summarizes the signals of backplane port.

Table 4 Backplane Port Signal Summary (Sheet 1 of 2)

Interface		Signal Name	Signal Type	Mate Order ^a	Break Order ^b	Number of Contacts	Description		
Signaling Group - High Speed Differential									
12x	8x	4x	1x	IBsxIw(0)	InDiff	High Speed	High Speed	2 1 pair	IB Symbol Input Differential Signaling
				IBsxOw(0)	OutDiff	High Speed	High Speed	2 1 pair	IB Symbol Output Differential Signaling
				IBsxIw(3:1)	InDiff	High Speed	High Speed	6 3 pair	IB Symbol Input Differential Signaling
				IBsxOw(3:1)	OutDiff	High Speed	High Speed	6 3 pair	IB Symbol Output Differential Signaling
				IBsxIw(7:4)	InDiff	High Speed	High Speed	8 4 pair	IB Symbol Input Differential Signaling
				IBsxOw(7:4)	OutDiff	High Speed	High Speed	8 4 pair	IB Symbol Output Differential Signaling
				IBsxIw(11:8)	InDiff	High Speed	High Speed	8 4 pair	IB Symbol Input Differential Signaling
				IBsxOw(11:8)	OutDiff	High Speed	High Speed	8 4 pair	IB Symbol Output Differential Signaling
		IB_Sh_Ret	ShRet	High Speed	High Speed	24	Shield Return tied to Logic Ground		
Hardware Management Group									
		IMxClk	InOut	3	2	1	IB Management Link Clock		
		IMxDat	InOut	3	2	1	IB Management Link Data		
		IMxInt_L	Out	3	2	1	IB Management Link Interrupt		
		IMxPRst	InOut	4	1	1	Presence/Reset		
		IMxPReq_L	Out	3	2	1	Power Request		

Table 4 Backplane Port Signal Summary (Sheet 2 of 2)

Interface	Signal Name	Signal Type	Mate Order ^a	Break Order ^b	Number of Contacts	Description
Bulk Power Group						
	VB_In^c	PwrIn	2	3	4	12 Volt Bulk Power Input
	VB_Ret	PwrRet	1	4	4	12 Volt Bulk Power Return
	VBxEn_L	PwrU / PwrD	4	1	1	Bulk Power Enable Not Asserted - Disable Asserted - Enable
	VBxCap	APwrU / APwrD	3	2	1	Bulk Power Capability Not Asserted - 25W; Asserted - 50W
	VBxPFW_L	In	3	2	1	Power Fail Warning
Auxilliary Power Group						
	VA_In^c	APwrIn	2	3	1	5 Volt Auxiliary Power Input
	VA_Ret^c	APwrRet	1	4	1	Logic Ground

a. See the *InfiniBand Architecture Specification, Volume 2-DEPR, Section 4.6.1 on page 132* for detailed description

b. See the *InfiniBand Architecture Specification, Volume 2-DEPR, Section 4.6.2 on page 134* and *Volume 2-DEPR, Section 4.6.3 on page 136* for detailed description

c. This signal may or may not be specific to a port. If it is not specific to a port (i.e., it is common among multiple ports), the “_” is used in the signal name; if it is specific to a port, then the “_” shown should be replaced with the port number in the same manner as “x” is used for other signals as described in [Table 2 Symbol Substitution on page 71](#). For documentation clarity, the “_” is used through the remainder of this specification.

As shown in [Figure 2 Backplane Port - Single Physical Link on page 61](#), the Signaling Group signals make up the IB Link for the width supported by a backplane port. The ports are physically located on the defined module form factors and the backplanes into which they plug as specified in [Volume 2-DEPR, Chapter 1: Mechanical Specification](#). Some of the module form factors defined allow for multiple backplane ports to be present. However, one is always required.

C4-1: In order to establish an IB Link, all modules shall have at least one backplane port containing the Signaling Group signals defined in [Table 4](#) located at Primary Port (1) as defined by [Volume 2-DEPR, Figure 63 Module Bulk Power Ports \(Logical\) on page 121](#) and [Volume 2-DEPR, Section 4.5, “Chassis Power Rules,” on page 131](#).

C4-2: In order to establish an IB Link, all chassis that accept InfiniBand modules shall have at least a backplane port containing the Signaling Group signals defined in [Table 4](#) located at Primary Port (1) as defined by [Volume 2-DEPR, Figure 63 Module Bulk Power Ports \(Logical\) on page 121](#) and [Volume 2-DEPR, Section 4.5, “Chassis Power Rules,” on page 131](#).

Primary Port (1) requirements for the Hardware Management Group are found in [Volume 2-DEPR, Chapter 5: Hardware Management](#). Requirements for the Bulk Power Group and Auxiliary Power Group are found in [Volume 2-DEPR, Chapter 4: Power / Hot Plug](#).

The backplane connector is specified in [Volume 2-DEPR, Chapter 2: Backplane Connector Specification](#).

4.2.2 CABLE PORT USING MICROGIGACN CONNECTOR

[Table 5](#) summarizes the signals of a cable port. Cable port requirements are found in [Chapter 7: Electrical Connectors for Modules and Cables](#).

Table 5 Cable Port Signal Summary

Interface		Signal Name	Signal Type	Number of Pins	Description		
Signaling Group - High Speed Differential							
12x	8x	4x	1x	IBsxIw(0)	InDiff	2 1 pair	IB Symbol Input Differential Signaling
				IBsxOw(0)	OutDiff	2 1 pair	IB Symbol Output Differential Signaling
				IBsxIw(3:1)	InDiff	6 3 pair	IB Symbol Input Differential Signaling
				IBsxOw(3:1)	OutDiff	6 3 pair	IB Symbol Output Differential Signaling
				IBsxIw(7:4)	InDiff	8 4 pair	IB Symbol Input Differential Signaling
				IBsxOw(7:4)	OutDiff	8 4 pair	IB Symbol Output Differential Signaling
				IBsxIw(11:8)	InDiff	8 4 pair	IB Symbol Input Differential Signaling
				IBsxOw(11:8)	OutDiff	8 4 pair	IB Symbol Output Differential Signaling
		IB_Sh_Ret	ShRet		Inner Shield Return tied to Logic Ground		

4.2.3 FIBER OPTIC PORT

[Table 6](#) summarizes the signals of a fiber optic port. Fiber optic port requirements are found in [Chapter 9: Fiber Attachment - 2.5 Gb/s, 5.0 Gb/s, & 10 Gb/s](#).

Table 6 Fiber Optic Port Signal Summary

Interface		Signal Name ^a	Signal Type	Number of Fibers	Description
Signaling Group - High Speed Optical					
12x	8x	IBsxIp(0)	Rx	1	IB Symbol Input
		IBsxOp(0)	Tx	1	IB Symbol Output
		IBsxIp(3:1)	Rx	3	IB Symbol Input
		IBsxOp(3:1)	Tx	3	IB Symbol Output
		IBsxIp(7:4)	Rx	4	IB Symbol Input
		IBsxOp(7:4)	Tx	4	IB Symbol Output
		IBsxIp(11:8)	Rx	4	IB Symbol Input
		IBsxOp(11:8)	Tx	4	IB Symbol Output

a. As each lane is made of a single fiber per direction versus a differential pair per direction for electrical interfaces, the "w" portion of the signal name only will have one polarity present. By convention, the "p" (positive) polarity is used.

4.2.4 ACTIVE CABLE PORT USING MICROGIGACN CONNECTOR

[Table 7](#) summarizes the signals of an active cable port using the MicroGigaCN connector. Active Cable port requirements are found in [Section 7.3.2.3 on page 362](#), [Section](#)

[7.7.3.3 on page 405](#), and [Section 7.7.3.6 on page 410](#).

Table 7 Active Cable Port Signal Summary

Interface				Signal Name	Signal Type	Number of Pins	Description			
Signaling Group - High Speed Differential										
12x	8x	4x	1x	IBsxlw(0)	InDiff	2 1 pair	IB Symbol Input Differential Signaling			
				IBsxOw(0)	OutDiff	2 1 pair	IB Symbol Output Differential Signaling			
				IBsxlw(3:1)	InDiff	6 3 pair	IB Symbol Input Differential Signaling			
				IBsxOw(3:1)	OutDiff	6 3 pair	IB Symbol Output Differential Signaling			
				IBsxlw(7:4)	InDiff	8 4 pair	IB Symbol Input Differential Signaling			
				IBsxOw(7:4)	OutDiff	8 4 pair	IB Symbol Output Differential Signaling			
				IBsxlw(11:8)	InDiff	8 4 pair	IB Symbol Input Differential Signaling			
				IBsxOw(11:8)	OutDiff	8 4 pair	IB Symbol Output Differential Signaling			
				IB_Sh_Ret	ShRet		Inner Shield Return tied to Logic Ground			
Active Cable Power Group										
12x	8x	4x		Sense-3.3V	In	1	Detection of Active Cable operating at 3.3V			
12x	8x	4x		Sense-12V	In	1	Detection of Active Cable operating at 12V			
12x	8x			Vcc	PwrIn	1	3.3Volt or 12Volt Bulk Power Input (for 4x). Return is through IB_Sh_Ret Logic Ground			
12x	8x			Vcc	PwrIn	2	3.3V or 12V Bulk Power Input (for 8x & 12x). Return is through IB_Sh_Ret Logic Ground			

4.2.5 CABLE PORT USING QSFP/QSFP+ CONNECTOR INTERFACE

Signals for passive and active cables using the QSFP/QSFP+ connector receptacles and plugs are defined in [Table 97, “Contact Assignment for 4x QSFP+ Interface,” on page 380](#) in [Section 7.5, “4X QSFP+ Interface connectors,” on page 373](#).

4.2.6 CABLE PORT USING CXP/CXP+ CONNECTOR INTERFACE

Signals for passive and active cables using the CXP/CXP+ connector receptacles and plugs are defined in [Table 109, “Contact Assignments for 12x Pluggable-CXP Interface.” on page 419](#) in [Section 7.8, “CXP Interface,” on page 413](#).

4.3 SIGNALING GROUP

This group is comprised of the high speed signals that make up the InfiniBand link. The link consists of two uni-directional interfaces, one operating as an input (or receiver) and one operating as an output (or driver). These are **IBsxIw** and **IBsxOw**, respectively.

4.3.1 HIGH SPEED ELECTRICAL

Based on the width of interface supported, the number of signals within the group varies: 1x consists of 4 conductors (1 differential pair in, 1 differential pair out); 4x consists of 16 conductors (4 differential pair in, 4 differential pair out); 8x consists of 32 conductors (8 differential pair in, 8 differential pair out); and 12x consists of 48 conductors (12 differential pair in, 12 differential pair out).

These signals conform to the electrical signaling specifications as defined in [Chapter 6: High Speed Electrical Interfaces](#) and [Chapter 7: Electrical Connectors for Modules and Cables](#).

4.3.2 HIGH SPEED OPTICAL

Based on the width of interface supported, the number of signals within the group varies: 1x consists of 2 fibers (1 in, 1 out); 4x consists of 8 fibers (4 in, 4 out); 8x consists of 16 fibers (8 in, 8 out); and 12x consists of 24 fibers (12 in, 12 out).

These signals conform to the electrical signaling as defined in [Chapter 9: Fiber Attachment - 2.5 Gb/s, 5.0 Gb/s, & 10 Gb/s](#).

4.4 HARDWARE MANAGEMENT GROUP

Implementation Note

This section describes features which were deprecated in Rel. 1.3. They may not be compatible with current InfiniBand architecture features and functions.

The Hardware Management Group provides a serial management interface which supports communication between the module and an optional intelligent environmental controller (hereafter called the Chassis Management Element or “CME”) associated with the IB backplane. It also includes additional signals to allow for low level interaction between the chassis backplanes and the module.

The InfiniBand Management Link, also known as IB-ML, is made up by the **IMxDat** and **IMxCIk** signals and allow access to a number of architected facilities that provide identification, status, and control of hardware management features.

The **IMxInt_L** signal originates from the module and allows the module to provide an interrupt notification of a condition that requires backplane attention. IB-ML operations are used in response to this signal to determine the event type.

The **IMxPRst** signal is a bidirectional signal that allows the backplane to detect module presence or force module reset

The **IMxPReq_L** signal originates from the module and allows the module to request that bulk power be supplied.

These signals conform to the electrical signaling as defined in [Volume 2-DEPR, Chapter 3: Low Speed Electrical Signaling](#).

4.5 BULK POWER GROUP

Implementation Note

This section describes features which were deprecated in Rel. 1.3. They may not be compatible with current InfiniBand architecture features and functions.

The Bulk Power Group provides the majority of the IB module's operational power. In particular, the IB Signaling Group and the IB module application (i.e. Fibre Channel, Gigabit Ethernet, etc.) are powered only by the Bulk Power Group. Power is delivered at nominally 12V.

The **VB_In** and **VB_Ret** connections drive an on-board power sequencer that controls power to the module's local DC-DC power converter(s). The IB standard requires an on module power sequencer capable of:

- Sensing full insertion of the module
- Checking power required by a board and the power provided by the backplane
- Enabling board power-up on board/backplane power compatibility
- Power-up override by a Chassis Management Element (CME) associated with a managed backplane (See [Volume 2-DEPR, Chapter 5: Hardware Management](#) for information on Chassis Management Elements as it pertains to this specification.)

These capabilities are provided by the **VBxEn_L** and **VBxCap** pins.

The **VBxPFW_L** signal originates from the system or chassis power supply and provides an indication that the module's bulk power input may be about to go out of specified tolerance.

These signals conform to the electrical signaling as defined in [Volume 2-DEPR, Chapter 3: Low Speed Electrical Signaling](#). Further details of the usage of the Bulk Power group can be found in [Volume 2-DEPR, Chapter 4: Power / Hot Plug](#) of this specification.

4.6 AUXILIARY POWER GROUP

Implementation Note

This section describes features which were deprecated in Rel. 1.3. They may not be compatible with current InfiniBand architecture features and functions.

The Auxiliary Power Group provides a low power, yet always available supply from which certain management functions can be available in the absence of Bulk Power. Functions include module information access through IB-ML (See [Volume 2-DEPR, Chapter 5: Hardware Management](#)), beacon sequence detection for in-band Power Management (See [Volume 2-DEPR, Chapter 6: OS Power Management](#), [Chapter 5: Link/Phy Interface](#)), and detection or generation of Wake Request Events (See [Volume 2-DEPR, Chapter 6: OS Power Management](#)).

Power is delivered at nominally 5V.

The **VA_In** connection provides the power. The return for this supply is **VA_Ret** which may be tied to logic ground.

These signals conform to the electrical signaling as defined in [Volume 2-DEPR, Chapter 3: Low Speed Electrical Signaling](#). Further details of the usage of the Auxiliary Power group can be found in [Volume 2-DEPR, Chapter 4: Power / Hot Plug](#) of this specification.

4.7 ACTIVE CABLE POWER GROUP

The Active Cable Power Group provides power for active cables using MicroGigaCN Connectors, as specified in [Section 7.3.3.2, “Active Cable power requirements,” on page 367](#).

Two **Sense** pins are used to distinguish Active Cables operating at either 3.3V or 12V from passive (non-powered) cables, which use the same physical connector. The **Vcc** pins provide power for active components in the cable, with return through signal ground. Power is delivered nominally at either 12V or 3.3V, as determined by needs of the active cable assembly.

CHAPTER 5: LINK/PHY INTERFACE

5.1 INTRODUCTION

The Link Physical layer provides an interface between the packet byte stream of upper layers and the serial bit stream(s) of the physical media. The physical media may be implemented as 1, 4, 8, or 12 physical lanes. The packet byte stream will be byte striped across the available physical lanes. (See [Figure 8](#)) The byte stream on each physical lane is encoded using either 8b/10b or 64b/66b coding. In addition to encode and decode, the link physical layer includes link training and initialization logic, clock tolerance compensation logic, and receive error detecting logic.

The Transmit Data Flow is responsible for:

- Insertion of control sequence information
- 8b/10b or 64b/66b encoding and scrambling
- Forward Error Correction, when enabled

The Receive Data Flow is responsible for:

- Deletion of control sequence information
- Error detection and handling
- 8b/10b or 64b/66b decoding and descrambling
- Forward Error Correction decoding and correction, when enabled

[Figure 8](#) shows a block diagram of the link physical layer for 8b/10b encoding.

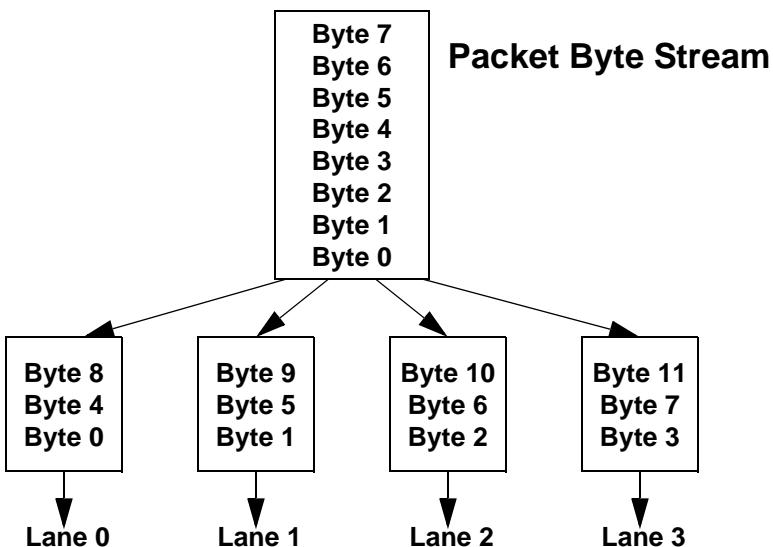


Figure 8 8b/10b Example 4x Byte Striping Diagram

[Figure 9](#) shows a block diagram of the link physical layer for 64b/66b encoding.

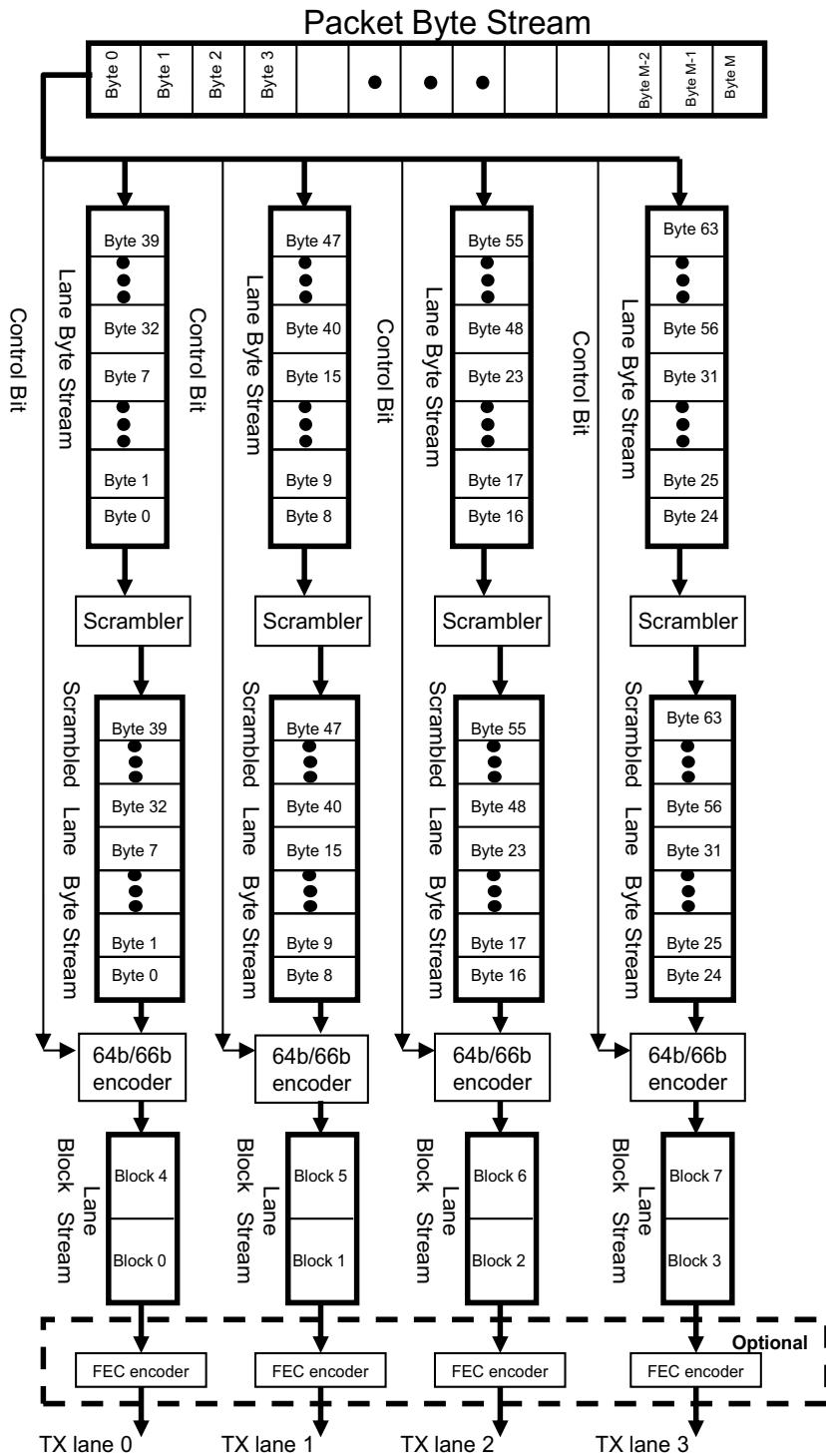


Figure 9 64b/66b Example 4x Byte Striping Diagram

Figure 10 shows a block diagram of the link physical layer for 64b/66b encoding and Reed-Solomon forward error correction enabled, for a 4x link width.

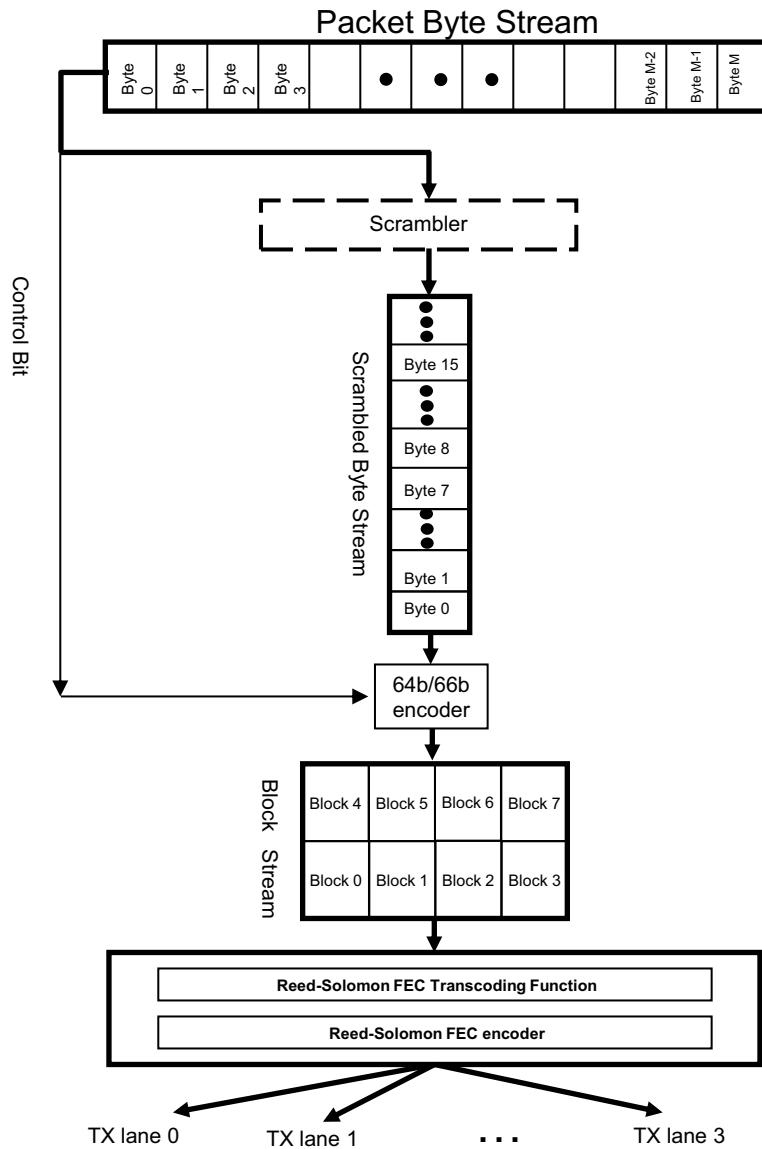


Figure 10 64b/66b Example 4x Byte Striping Diagram - with RS-FEC Enabled

5.2 SYMBOL ENCODING (8B/10B CODING)

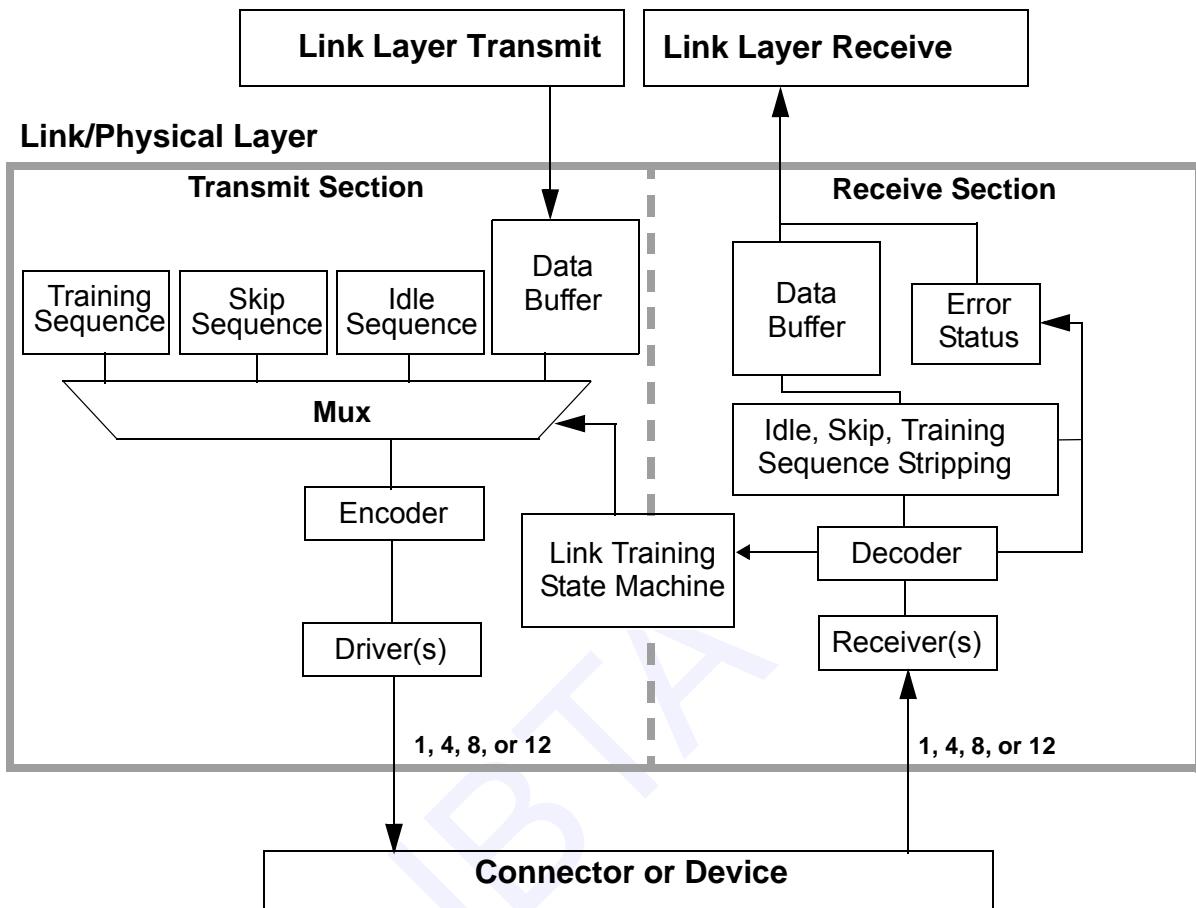


Figure 11 Link/Physical Interface Block Diagram

The InfiniBand physical lane encoding for SDR, DDR and QDR rates uses the 8b/10b code which is used by Fibre Channel, Gigabit Ethernet (IEEE 802.3), FICON, and ServerNet. The 8b/10b code provides DC balance, limited run lengths, byte (symbol) synchronization, and the ability to distinguish between data characters and control characters.

C5-1: All ports shall use the 8b/10b code as defined in [Section 5.2. "Symbol Encoding \(8b/10b coding\)," on page 84](#) for operation at SDR, DDR, and QDR signaling rates.

5.2.1 NOTATION CONVENTIONS

The 8b/10b transmission code uses letter notation for describing the bits of an unencoded information byte and a single control variable. Each bit of the unencoded information byte contains either a binary zero or a binary one. A control variable, Z, has either the value D or the value K. When the control variable associated with an unencoded information byte contains the value D, the associated encoded code-group is referred to

as a data code-group. When the control variable associated with an unencoded information byte contains the value K, the associated encoded code-group is referred to as a special code-group.

The bit notation of A,B,C,D,E,F,G,H for an unencoded information byte is used in the description of the 8b/10b transmission code. The bits A,B,C,D,E,F,G,H are translated to bits a,b,c,d,e,i,f,g,h,j of 10-bit transmission code-groups. [Figure 12](#) illustrates the byte and bit nomenclature as a byte stream is encoded (decoded) and serialized (de-serialized). 8b/10b code-group bit assignments are illustrated in [Table 8](#) and [Table 9](#). Each valid code-group has been given a name using the following convention:

- 1) /Dx.y/ for the 256 valid data code-groups, and
- 2) /Kx.y/ for the special control code-groups where x is the decimal value of bits EDCBA, and y is the decimal value of bits HGF.

Examples of this are D10.2 or K28.5.

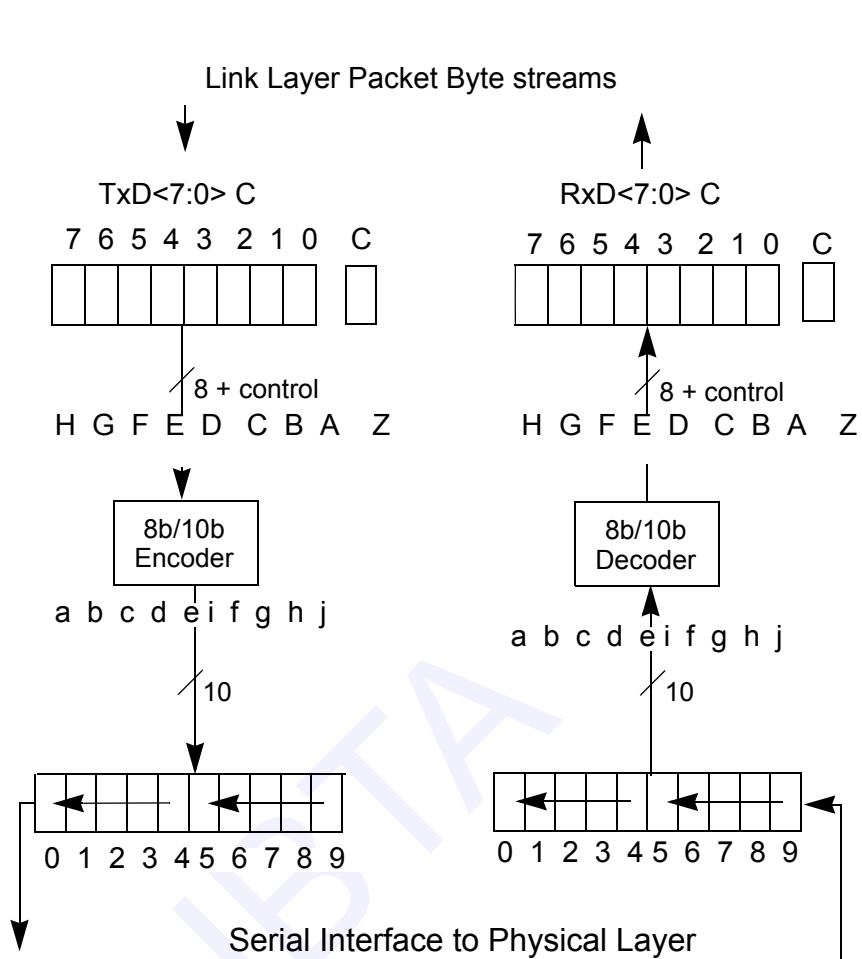


Figure 12 Transmit & Receive Data Ordering

5.2.2 VALID AND INVALID CODE-GROUPS

[Table 8](#) defines the valid data code-groups (D code-groups) of the 8b/10b transmission code. [Table 9](#) defines the valid special code-groups (K code-groups) of the code. The tables are used both for generating valid code-groups (encoding) and for checking the validity of received code-groups (decoding).

In the tables, each byte entry has two columns that represent two code-groups which are not necessarily different. The two columns correspond to the valid code-group based on the current value of the running disparity (Current RD - or Current RD +). Running disparity is a binary parameter with either the value negative (-) or the value positive (+). See [Section 5.2.3, “Running disparity rules,” on page 93](#) for the definition of and rules for disparity.

Table 8 Valid Data Code Groups (Sheet 1 of 6)

Data Byte Name	Data Byte Value	Bits HGF EDCBA	Current RD - abcdei fghj	Current RD + abcdei fghj
D0.0	00	000 00000	100111 0100	011000 1011
D1.0	01	000 00001	011101 0100	100010 1011
D2.0	02	000 00010	101101 0100	010010 1011
D3.0	03	000 00011	110001 1011	110001 0100
D4.0	04	000 00100	110101 0100	001010 1011
D5.0	05	000 00101	101001 1011	101001 0100
D6.0	06	000 00110	011001 1011	011001 0100
D7.0	07	000 00111	111000 1011	000111 0100
D8.0	08	000 01000	111001 0100	000110 1011
D9.0	09	000 01001	100101 1011	100101 0100
D10.0	0A	000 01010	010101 1011	010101 0100
D11.0	0B	000 01011	110100 1011	110100 0100
D12.0	0C	000 01100	001101 1011	001101 0100
D13.0	0D	000 01101	101100 1011	101100 0100
D14.0	0E	000 01110	011100 1011	011100 0100
D15.0	0F	000 01111	010111 0100	101000 1011
D16.0	10	000 10000	011011 0100	100100 1011
D17.0	11	000 10001	100011 1011	100011 0100
D18.0	12	000 10010	010011 1011	010011 0100
D19.0	13	000 10011	110010 1011	110010 0100
D20.0	14	000 10100	001011 1011	001011 0100
D21.0	15	000 10101	101010 1011	101010 0100
D22.0	16	000 10110	011010 1011	011010 0100
D23.0	17	000 10111	111010 0100	000101 1011
D24.0	18	000 11000	110011 0100	001100 1011
D25.0	19	000 11001	100110 1011	100110 0100
D26.0	1A	000 11010	010110 1011	010110 0100
D27.0	1B	000 11011	110110 0100	001001 1011
D28.0	1C	000 11100	001110 1011	001110 0100
D29.0	1D	000 11101	101110 0100	010001 1011
D30.0	1E	000 11110	011110 0100	100001 1011
D31.0	1F	000 11111	101011 0100	010100 1011
D0.1	20	001 00000	100111 1001	011000 1001
D1.1	21	001 00001	011101 1001	100010 1001
D2.1	22	001 00010	101101 1001	010010 1001
D3.1	23	001 00011	110001 1001	110001 1001
D4.1	24	001 00100	110101 1001	001010 1001
D5.1	25	001 00101	101001 1001	101001 1001
D6.1	26	001 00110	011001 1001	011001 1001
D7.1	27	001 00111	111000 1001	000111 1001
D8.1	28	001 01000	111001 1001	000110 1001

Table 8 Valid Data Code Groups (Sheet 2 of 6)

Data Byte Name	Data Byte Value	Bits HGF EDCBA	Current RD - abcdei fghj	Current RD + abcdei fghj
D9.1	29	001 01001	100101 1001	100101 1001
D10.1	2A	001 01010	010101 1001	010101 1001
D11.1	2B	001 01011	110100 1001	110100 1001
D12.1	2C	001 01100	001101 1001	001101 1001
D13.1	2D	001 01101	101100 1001	101100 1001
D14.1	2E	001 01110	011100 1001	011100 1001
D15.1	2F	001 01111	010111 1001	101000 1001
D16.1	30	001 10000	011011 1001	100100 1001
D17.1	31	001 10001	100011 1001	100011 1001
D18.1	32	001 10010	010011 1001	010011 1001
D19.1	33	001 10011	110010 1001	110010 1001
D20.1	34	001 10100	001011 1001	001011 1001
D21.1	35	001 10101	101010 1001	101010 1001
D22.1	36	001 10110	011010 1001	011010 1001
D23.1	37	001 10111	111010 1001	000101 1001
D24.1	38	001 11000	110011 1001	001100 1001
D25.1	39	001 11001	100110 1001	100110 1001
D26.1	3A	001 11010	010110 1001	010110 1001
D27.1	3B	001 11011	110110 1001	001001 1001
D28.1	3C	001 11100	001110 1001	001110 1001
D29.1	3D	001 11101	101110 1001	010001 1001
D30.1	3E	001 11110	011110 1001	100001 1001
D31.1	3F	001 11111	101011 1001	010100 1001
D0.2	40	010 00000	100111 0101	011000 0101
D1.2	41	010 00001	011101 0101	100010 0101
D2.2	42	010 00010	101101 0101	010010 0101
D3.2	43	010 00011	110001 0101	110001 0101
D4.2	44	010 00100	110101 0101	001010 0101
D5.2	45	010 00101	101001 0101	101001 0101
D6.2	46	010 00110	011001 0101	011001 0101
D7.2	47	010 00111	111000 0101	000111 0101
D8.2	48	010 01000	111001 0101	000110 0101
D9.2	49	010 01001	100101 0101	100101 0101
D10.2	4A	010 01010	010101 0101	010101 0101
D11.2	4B	010 01011	110100 0101	110100 0101
D12.2	4C	010 01100	001101 0101	001101 0101
D13.2	4D	010 01101	101100 0101	101100 0101
D14.2	4E	010 01110	011100 0101	011100 0101
D15.2	4F	010 01111	010111 0101	101000 0101
D16.2	50	010 10000	011011 0101	100100 0101
D17.2	51	010 10001	100011 0101	100011 0101
D18.2	52	010 10010	010011 0101	010011 0101
D19.2	53	010 10011	110010 0101	110010 0101

Table 8 Valid Data Code Groups (Sheet 3 of 6)

Data Byte Name	Data Byte Value	Bits HGF EDCBA	Current RD - abcdei fghj	Current RD + abcdei fghj
D20.2	54	010 10100	001011 0101	001011 0101
D21.2	55	010 10101	101010 0101	101010 0101
D22.2	56	010 10110	011010 0101	011010 0101
D23.2	57	010 10111	111010 0101	000101 0101
D24.2	58	010 11000	110011 0101	001100 0101
D25.2	59	010 11001	100110 0101	100110 0101
D26.2	5A	010 11010	010110 0101	010110 0101
D27.2	5B	010 11011	110110 0101	001001 0101
D28.2	5C	010 11100	001110 0101	001110 0101
D29.2	5D	010 11101	101110 0101	010001 0101
D30.2	5E	010 11110	011110 0101	100001 0101
D31.2	5F	010 11111	101011 0101	010100 0101
D0.3	60	011 00000	100111 0011	011000 1100
D1.3	61	011 00001	011101 0011	100010 1100
D2.3	62	011 00010	101101 0011	010010 1100
D3.3	63	011 00011	110001 1100	110001 0011
D4.3	64	011 00100	110101 0011	001010 1100
D5.3	65	011 00101	101001 1100	101001 0011
D6.3	66	011 00110	011001 1100	011001 0011
D7.3	67	011 00111	111000 1100	000111 0011
D8.3	68	011 01000	111001 0011	000110 1100
D9.3	69	011 01001	100101 1100	100101 0011
D10.3	6A	011 01010	010101 1100	010101 0011
D11.3	6B	011 01011	110100 1100	110100 0011
D12.3	6C	011 01100	001101 1100	001101 0011
D13.3	6D	011 01101	101100 1100	101100 0011
D14.3	6E	011 01110	011100 1100	011100 0011
D15.3	6F	011 01111	010111 0011	101000 1100
D16.3	70	011 10000	011011 0011	100100 1100
D17.3	71	011 10001	100011 1100	100011 0011
D18.3	72	011 10010	010011 1100	010011 0011
D19.3	73	011 10011	110010 1100	110010 0011
D20.3	74	011 10100	001011 1100	001011 0011
D21.3	75	011 10101	101010 1100	101010 0011
D22.3	76	011 10110	011010 1100	011010 0011
D23.3	77	011 10111	111010 0011	000101 1100
D24.3	78	011 11000	110011 0011	001100 1100
D25.3	79	011 11001	100110 1100	100110 0011
D26.3	7A	011 11010	010110 1100	010110 0011
D27.3	7B	011 11011	110110 0011	001001 1100
D28.3	7C	011 11100	001110 1100	001110 0011
D29.3	7D	011 11101	101110 0011	010001 1100
D30.3	7E	011 11110	011110 0011	100001 1100

Table 8 Valid Data Code Groups (Sheet 4 of 6)

Data Byte Name	Data Byte Value	Bits HGF EDCBA	Current RD - abcdei fghj	Current RD + abcdei fghj
D31.3	7F	011 11111	101011 0011	010100 1100
D0.4	80	100 00000	100111 0010	011000 1101
D1.4	81	100 00001	011101 0010	100010 1101
D2.4	82	100 00010	101101 0010	010010 1101
D3.4	83	100 00011	110001 1101	110001 0010
D4.4	84	100 00100	110101 0010	001010 1101
D5.4	85	100 00101	101001 1101	101001 0010
D6.4	86	100 00110	011001 1101	011001 0010
D7.4	87	100 00111	111000 1101	000111 0010
D8.4	88	100 01000	111001 0010	000110 1101
D9.4	89	100 01001	100101 1101	100101 0010
D10.4	8A	100 01010	010101 1101	010101 0010
D11.4	8B	100 01011	110100 1101	110100 0010
D12.4	8C	100 01100	001101 1101	001101 0010
D13.4	8D	100 01101	101100 1101	101100 0010
D14.4	8E	100 01110	011100 1101	011100 0010
D15.4	8F	100 01111	010111 0010	101000 1101
D16.4	90	100 10000	011011 0010	100100 1101
D17.4	91	100 10001	100011 1101	100011 0010
D18.4	92	100 10010	010011 1101	010011 0010
D19.4	93	100 10011	110010 1101	110010 0010
D20.4	94	100 10100	001011 1101	001011 0010
D21.4	95	100 10101	101010 1101	101010 0010
D22.4	96	100 10110	011010 1101	011010 0010
D23.4	97	100 10111	111010 0010	000101 1101
D24.4	98	100 11000	110011 0010	001100 1101
D25.4	99	100 11001	100110 1101	100110 0010
D26.4	9A	100 11010	010110 1101	010110 0010
D27.4	9B	100 11011	110110 0010	001001 1101
D28.4	9C	100 11100	001110 1101	001110 0010
D29.4	9D	100 11101	101110 0010	010001 1101
D30.4	9E	100 11110	011110 0010	100001 1101
D31.4	9F	100 11111	101011 0010	010100 1101
D0.5	A0	101 00000	100111 1010	011000 1010
D1.5	A1	101 00001	011101 1010	100010 1010
D2.5	A2	101 00010	101101 1010	010010 1010
D3.5	A3	101 00011	110001 1010	110001 1010
D4.5	A4	101 00100	110101 1010	001010 1010
D5.5	A5	101 00101	101001 1010	101001 1010
D6.5	A6	101 00110	011001 1010	011001 1010
D7.5	A7	101 00111	111000 1010	000111 1010
D8.5	A8	101 01000	111001 1010	000110 1010
D9.5	A9	101 01001	100101 1010	100101 1010

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

Table 8 Valid Data Code Groups (Sheet 5 of 6)

Data Byte Name	Data Byte Value	Bits HGF EDCBA	Current RD - abcdei fghj	Current RD + abcdei fghj
D10.5	AA	101 01010	010101 1010	010101 1010
D11.5	AB	101 01011	110100 1010	110100 1010
D12.5	AC	101 01100	001101 1010	001101 1010
D13.5	AD	101 01101	101100 1010	101100 1010
D14.5	AE	101 01110	011100 1010	011100 1010
D15.5	AF	101 01111	010111 1010	101000 1010
D16.5	B0	101 10000	011011 1010	100100 1010
D17.5	B1	101 10001	100011 1010	100011 1010
D18.5	B2	101 10010	010011 1010	010011 1010
D19.5	B3	101 10011	110010 1010	110010 1010
D20.5	B4	101 10100	001011 1010	001011 1010
D21.5	B5	101 10101	101010 1010	101010 1010
D22.5	B6	101 10110	011010 1010	011010 1010
D23.5	B7	101 10111	111010 1010	000101 1010
D24.5	B8	101 11000	110011 1010	001100 1010
D25.5	B9	101 11001	100110 1010	100110 1010
D26.5	BA	101 11010	010110 1010	010110 1010
D27.5	BB	101 11011	110110 1010	001001 1010
D28.5	BC	101 11100	001110 1010	001110 1010
D29.5	BD	101 11101	101110 1010	010001 1010
D30.5	BE	101 11110	011110 1010	100001 1010
D31.5	BF	101 11111	101011 1010	010100 1010
D0.6	C0	110 00000	100111 0110	011000 0110
D1.6	C1	110 00001	011101 0110	100010 0110
D2.6	C2	110 00010	101101 0110	010010 0110
D3.6	C3	110 00011	110001 0110	110001 0110
D4.6	C4	110 00100	110101 0110	001010 0110
D5.6	C5	110 00101	101001 0110	101001 0110
D6.6	C6	110 00110	011001 0110	011001 0110
D7.6	C7	110 00111	111000 0110	000111 0110
D8.6	C8	110 01000	111001 0110	000110 0110
D9.6	C9	110 01001	100101 0110	100101 0110
D10.6	CA	110 01010	010101 0110	010101 0110
D11.6	CB	110 01011	110100 0110	110100 0110
D12.6	CC	110 01100	001101 0110	001101 0110
D13.6	CD	110 01101	101100 0110	101100 0110
D14.6	CE	110 01110	011100 0110	011100 0110
D15.6	CF	110 01111	010111 0110	101000 0110
D16.6	D0	110 10000	011011 0110	100100 0110
D17.6	D1	110 10001	100011 0110	100011 0110
D18.6	D2	110 10010	010011 0110	010011 0110
D19.6	D3	110 10011	110010 0110	110010 0110
D20.6	D4	110 10100	001011 0110	001011 0110

Table 8 Valid Data Code Groups (Sheet 6 of 6)

Data Byte Name	Data Byte Value	Bits HGF EDCBA	Current RD - abcdei fghj	Current RD + abcdei fghj
D21.6	D5	110 10101	101010 0110	101010 0110
D22.6	D6	110 10110	011010 0110	011010 0110
D23.6	D7	110 10111	111010 0110	000101 0110
D24.6	D8	110 11000	110011 0110	001100 0110
D25.6	D9	110 11001	100110 0110	100110 0110
D26.6	DA	110 11010	010110 0110	010110 0110
D27.6	DB	110 11011	110110 0110	001001 0110
D28.6	DC	110 11100	001110 0110	001110 0110
D29.6	DD	110 11101	101110 0110	010001 0110
D30.6	DE	110 11110	011110 0110	100001 0110
D31.6	DF	110 11111	101011 0110	010100 0110
D0.7	E0	111 00000	100111 0001	011000 1110
D1.7	E1	111 00001	011101 0001	100010 1110
D2.7	E2	111 00010	101101 0001	010010 1110
D3.7	E3	111 00011	110001 1110	110001 0001
D4.7	E4	111 00100	110101 0001	001010 1110
D5.7	E5	111 00101	101001 1110	101001 0001
D6.7	E6	111 00110	011001 1110	011001 0001
D7.7	E7	111 00111	111000 1110	000111 0001
D8.7	E8	111 01000	111001 0001	000110 1110
D9.7	E9	111 01001	100101 1110	100101 0001
D10.7	EA	111 01010	010101 1110	010101 0001
D11.7	EB	111 01011	110100 1110	110100 1000
D12.7	EC	111 01100	001101 1110	001101 0001
D13.7	ED	111 01101	101100 1110	101100 1000
D14.7	EE	111 01110	011100 1110	011100 1000
D15.7	EF	111 01111	010111 0001	101000 1110
D16.7	F0	111 10000	011011 0001	100100 1110
D17.7	F1	111 10001	100011 0111	100011 0001
D18.7	F2	111 10010	010011 0111	010011 0001
D19.7	F3	111 10011	110010 1110	110010 0001
D20.7	F4	111 10100	001011 0111	001011 0001
D21.7	F5	111 10101	101010 1110	101010 0001
D22.7	F6	111 10110	011010 1110	011010 0001
D23.7	F7	111 10111	111010 0001	000101 1110
D24.7	F8	111 11000	110011 0001	001100 1110
D25.7	F9	111 11001	100110 1110	100110 0001
D26.7	FA	111 11010	010110 1110	010110 0001
D27.7	FB	111 11011	110110 0001	001001 1110
D28.7	FC	111 11100	001110 1110	001110 0001
D29.7	FD	111 11101	101110 0001	010001 1110
D30.7	FE	111 11110	011110 0001	100001 1110
D31.7	FF	111 11111	101011 0001	010100 1110

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

Table 9 Valid Special Code Groups

Data Byte Name	Special Byte Value	Bits HGF EDCBA	Current RD - abcdei fghj	Current RD + abcdei fghj
K28.0	1C	000 11100	001111 0100	110000 1011
K28.1	3C	001 11100	<u>001111</u> 0001	<u>110000</u> 0110
K28.2	5C	010 11100	001111 0101	110000 1010
K28.3	7C	011 11100	001111 0011	110000 1100
K28.4	9C	100 11100	001111 0010	110000 1101
K28.5	BC	101 11100	<u>001111</u> 1010	<u>110000</u> 0101
K28.6	DC	110 11100	001111 0110	110000 1001
K28.7	FC	111 11100	<u>001111</u> 1000	<u>110000</u> 0111
K23.7	F7	111 10111	111010 1000	000101 0111
K27.7	FB	111 11011	110110 1000	001001 0111
K29.7	FD	111 11101	101110 1000	010001 0111
K30.7	FE	111 11110	011110 1000	100001 0111

Note: Underlined special code-groups contain the comma bit patterns 0011111 or 1100000. This pattern is not found in any valid Data code group or combination of Data code-groups, and is used for symbol synchronization.

Note: Codes that are not contained in the above two tables are considered invalid and shall generate a code violation error.

5.2.3 RUNNING DISPARITY RULES

Running disparity shall be calculated using the following rules. Calculations on code groups that have been transmitted will yield a transmitter's running disparity. Similarly, calculations on code groups that have been received will yield a receiver's running disparity.

Running disparity for a code-group is calculated on the basis of sub-blocks, where the first six bits (abcdei) ([Table 10](#) and [Table 12](#)) form one six-bit sub-block, and the second four bits (fghj) ([Table 11](#) and [Table 13](#)) form the other four-bit sub-block. Running disparity at the beginning of the six-bit sub-block is the running disparity at the end of the previous code-group. Running disparity at the beginning of the four-bit sub-block is the running disparity at the end of the six-bit sub-block. Running disparity at the end of the code-group is the running disparity at the end of the four-bit sub-block.

Running disparity for the sub-blocks is calculated as follows:

- 1) Running disparity at the end of any sub-block is positive if the sub-block contains more ones than zeros. It is also positive at the end of the six-bit sub-block if the six-

bit sub-block is 000111. It is likewise positive at the end of the four-bit sub-block if the four-bit sub-block is 0011;

- 2) Running disparity at the end of any sub-block is negative if the sub-block contains more zeros than ones. It is also negative at the end of the six-bit sub-block if the six-bit sub-block is 111000. It is also negative at the end of the four-bit sub-block if the four-bit sub-block is 1100;
- 3) Otherwise, running disparity at the end of the sub-block is the same as at the beginning of the sub-block.

Note: All sub-blocks with equal numbers of zeros and ones are disparity neutral. In order to limit the run length of 0's or 1's between sub-blocks, the 8b/10b transmission code rules specify that sub-blocks encoded as 000111 or 0011 are generated only when the running disparity at the beginning of the sub-block is positive; thus, running disparity at the end of these sub-blocks is also positive. Likewise, sub-blocks containing 111000 or 1100 are generated only when the running disparity at the beginning of the sub-block is negative; thus, running disparity at the end of these sub-blocks is also negative.

Table 10 5b/6b Coding for Data Characters

Data Byte Name	Unencoded Bits	Current RD -	Current RD +
	EDCBA	abcdei	abcdei
D0	00000	100111	011000
D1	00001	011101	100010
D2	00010	101101	010010
D3	00011	110001	110001
D4	00100	110101	001010
D5	00101	101001	101001
D6	00110	011001	011001
D7	00111	111000	000111
D8	01000	111001	000110
D9	01001	100101	100101
D10	01010	010101	010101
D11	01011	110100	110100
D12	01100	001101	001101
D13	01101	101100	101100
D14	01110	011100	011100
D15	01111	010111	101000
D16	10000	011011	100100
D17	10001	100011	100011
D18	10010	010011	010011
D19	10011	110010	110010

Table 10 5b/6b Coding for Data Characters

Data Byte Name	Unencoded Bits EDCBA	Current RD - abcdei	Current RD + abcdei
D20	10100	001011	001011
D21	10101	101010	101010
D22	10110	011010	011010
D23	10111	111010	000101
D24	11000	110011	001100
D25	11001	100110	100110
D26	11010	010110	010110
D27	11011	110110	001001
D28	11100	001110	001110
D29	11101	101110	010001
D30	11110	011110	100001
D31	11111	101011	010100

Table 11 3b/4b Coding for Data Characters

Data Byte Name	Unencoded Bits HGF	Current RD - fghj	Current RD + fghj
--.0	000	1011	0100
--.1	001	1001	1001
--.2	010	0101	0101
--.3	011	1100	0011
--.4	100	1101	0010
--.5	101	1010	1010
--.6	110	0110	0110
--.7	111	1110/0111	0001/1000

Table 12 5b/6b Coding for Special Characters

Data Byte Name	Unencoded Bits	Current RD -	Current RD +
	EDCBA	abcdei	abcdei
K28	11100	001111	110000
K23	10111	111010	000101
K27	11011	110110	001001
K29	11101	101110	010001
K30	11110	011110	100001

Table 13 3b/4b Coding for Special Characters

Data Byte Name	Unencoded Bits	Current RD -	Current RD +
	HGF	fghj	fghj
--.0	000	1011	0100
--.1	001	0110	1001
--.2	010	1010	0101
--.3	011	1100	0011
--.4	100	1101	0010
--.5	101	0101	1010
--.6	110	1001	0110
--.7	111	0111	1000

5.2.4 GENERATING CODE-GROUPS

The byte to be encoded and current value of the transmitter's running disparity shall be used to select the code-group from its [Table 8](#) or [Table 9](#). For each code-group transmitted, a new value of the running disparity is calculated. This new value is used as the transmitter's current running disparity for the next byte to be encoded and transmitted.

5.2.5 CHECKING THE VALIDITY OF RECEIVED CODE-GROUPS

The following rules shall be used to determine the validity of received code groups:

- 1) The columns in [Table 8](#) and [Table 9](#) corresponding to the current value of the receiver's running disparity shall be searched for the received code-group.
- 2) If the received code-group is found in the proper column according to the current running disparity, then the code-group shall be considered valid and the associated data byte determined (decoded) for data code-groups.
- 3) If the received code-group is not found in that column, then the code-group shall be considered invalid.

- 4) Independent of the code-group's validity, the received code-group shall be used to calculate a new value of running disparity. The new value shall be used as the receiver's current running disparity for the next received code-group.

Detection of an invalid code-group does not necessarily indicate that the code-group in which the invalid code-group was detected is in error. Invalid code-groups may result from a prior error which altered the running disparity of the bit stream but which did not result in a detectable error at the code-group in which the error occurred.

5.3 BLOCK ENCODING (64B/66B)

The InfiniBand physical lane encoding for FDR and EDR rates uses a 64b/66b code, similar to the encoding used by 10 Gigabit Ethernet (IEEE Std 802.3 Clause 49). The 64b/66b code provides block synchronization and the ability to distinguish between data and control blocks. The two-bit block header provides guaranteed limit on the run-length. Each 64-bit block is scrambled using a self-synchronizing scrambler. The scrambling provides good likelihood of a high transition density. The full structure and organization of block encoding and scrambling function ports that operate without Reed-Solomon forward error correction (RS-FEC) enabled is illustrated in [Figure 9 on page 82](#). The structure and organization for ports with RS-FEC enabled is illustrated in [Figure 10 on page 83](#).

The first two bits of a 66-bit block (bits 0 and 1) are the synchronization header (sync header). The sync header is 01 for data blocks and 10 for control blocks, as shown in [Table 14](#). All other values of the sync header (00 and 11) are illegal. The remaining 64 bits (bits 2-65) of the block are the block payload taken from the scrambled Lane Stream. The sync header bits are the only bits in the block that will always contain a transition; this attribute is used to acquire block synchronization.

The sync header is transmitted first, followed by byte 0 of the block. Byte 7 of the block is transmitted last.

For each byte of the block payload, the Least Significant Bit is transmitted first. For example, 1Eh will be transmitted from left to right in the following order: 01111000.

Data blocks contain 8 bytes of data following the sync header. Control blocks contain one block type byte following the sync header and an additional 7 bytes, as defined by the control block type, as shown in [Table 14](#).

Table 14 64b/66b Block Format

Block	Sync Header	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
Data	01	Data 0	Data 1	Data 2	Data 3	Data 4	Data 5	Data 6	Data 7
Control	10	Block Type	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6

5.3.1 SCRAMBLING

When RS-FEC is not enabled, the scrambler function scrambles each Packet Lane Stream independently to create the Scrambled Lane Stream as illustrated in [Figure 9 on page 82](#).

When RS-FEC is enabled, the scrambler function scrambles the Packet Byte Stream to create the Scrambled Byte Stream, as illustrated in [Figure 10 on page 83](#).

All of the data and control blocks excluding the two bit sync header are scrambled.

The payload of the block shall be scrambled using a self-synchronizing scrambler. The scrambler shall produce the same result as the implementation shown in [Figure 13](#). The scrambling polynomial is:

$$G(x) = 1 + x^{39} + x^{58}$$

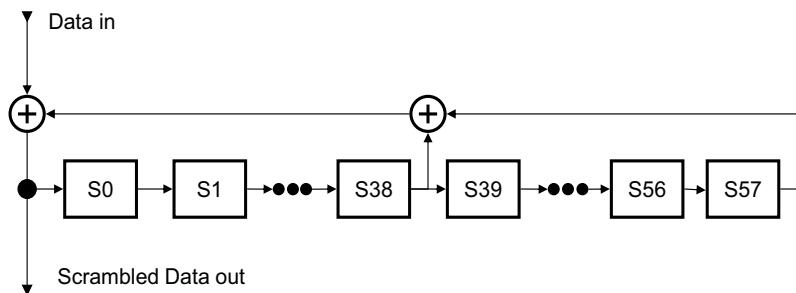


Figure 13 Scrambler

Each lane independently scrambles and de-scrambles the block payload. The de-scrambler shall produce the same result as the implementation shown in [Figure 14](#).

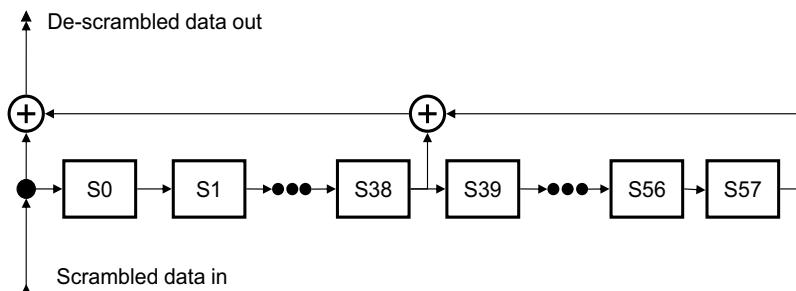


Figure 14 De-scrambler

There is no need to initialize the scrambler. The scrambler runs continuously on all payload bits.

The self-synchronizing scrambler and descrambler is designed so that it is entirely dependent on the scrambled data. Regardless of the initial state of the descrambler, the descrambler state is synchronized after the first 58 bits are received, and correctly descrambles data thereafter. The scrambler byte stream format is shown in [Figure 15](#),

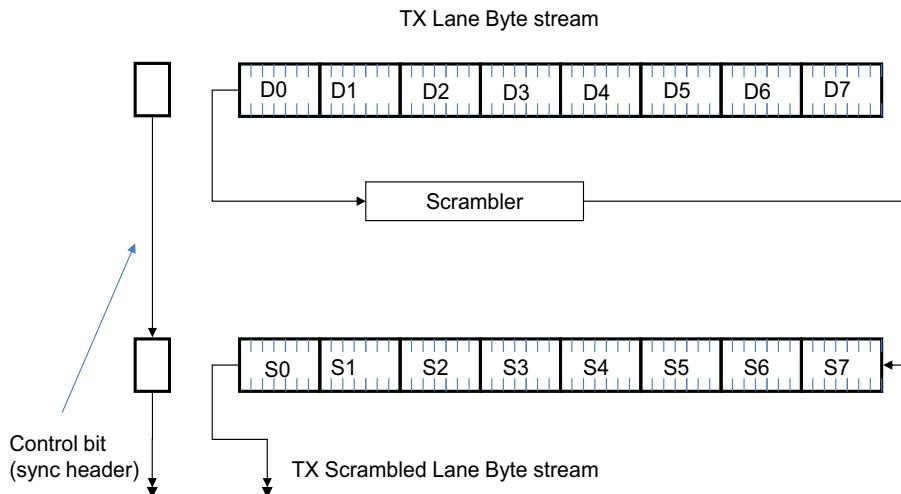


Figure 15 Scrambler Byte Stream Format

The receiver shall descramble all 8 bytes of the block payload it receives. The descrambler byte stream format is shown in [Figure 16](#),

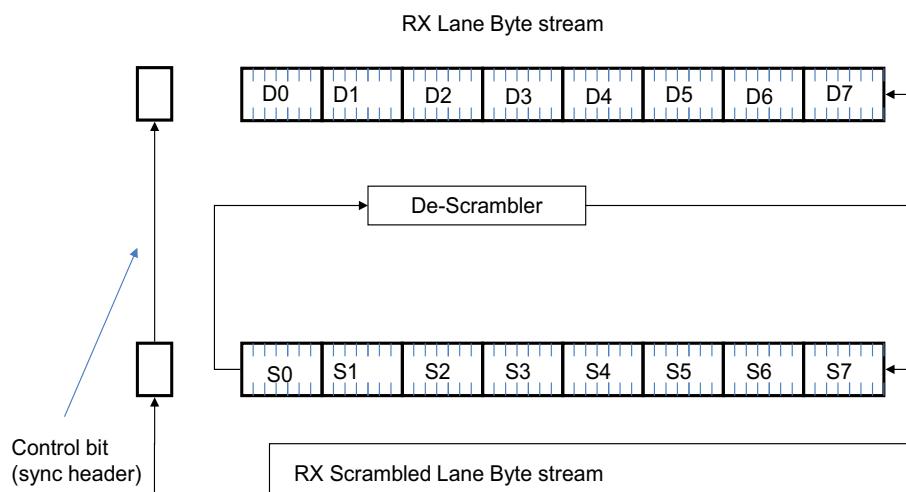


Figure 16 De-Scrambler Byte Stream Format

When RS-FEC is enabled, a single scrambler function and de-scrambler function will scramble and de-scramble all of the lanes of a port. See [Figure 17 on page 100](#).

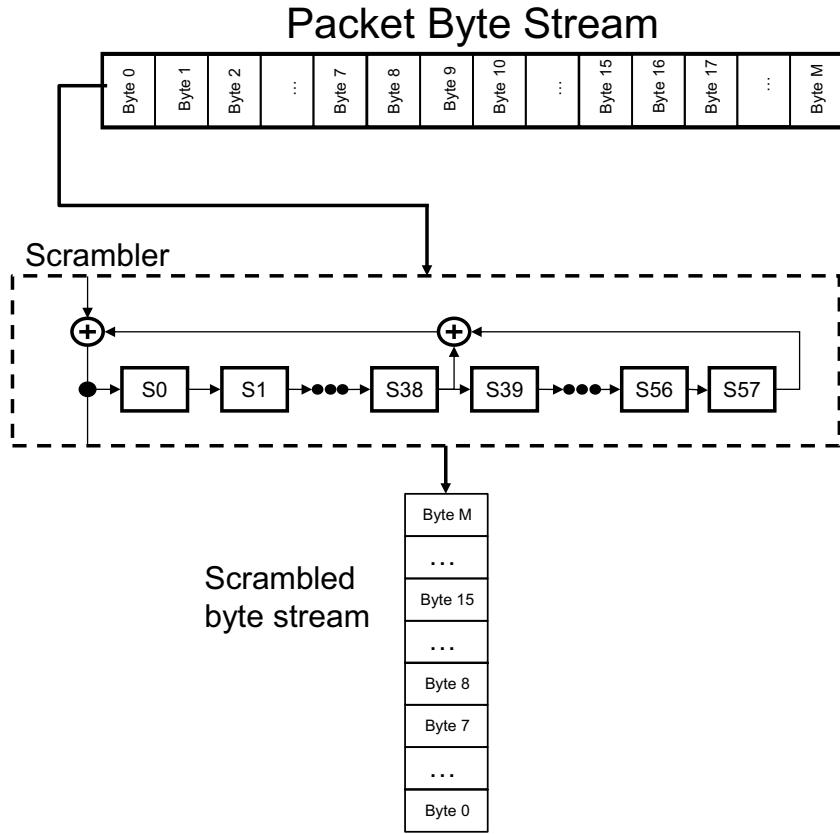


Figure 17 Scrambling order when RS-FEC is enabled

5.3.2 CHECKING THE VALIDITY OF A RECEIVED BLOCK

The following rules shall determine the validity of a received block:

- 1) Sync header must be either 01 for data blocks or 10 for control blocks. Any other value (00 or 11) is a sync header error and shall increment the sync header error counter **SyncHeaderErrorCounter** and the **UnknownBlockCounter**. This error is considered to be a minor error - see [Section 5.9.2, “Minor Link Physical Errors Events,” on page 236](#).
- 2) Control Blocks (sync header 10) must have a Block Type value defined in [Table 27](#). Any other Block Type is a block type error and shall increment the **UnknownBlockCounter**. This error is considered to be a minor error - see [Section 5.9.2, “Minor Link Physical Errors Events,” on page 236](#).
- 3) Control Blocks with Block Type Idle must consist of 7 bytes of idle data (00h) following the control block type byte (1Eh). Any Idle block with any byte other than 00h

in any byte after the control block type byte is considered a block error and shall increment the ***UnknownBlockCounter***. This error is considered to be minor error - see [Section 5.9.2, “Minor Link Physical Errors Events,” on page 236.](#)

- 4) Any Control Block that does not meet all the control block requirements as defined in [5.5.3 on page 159](#) is considered a block error and shall increment the ***UnknownBlockCounter***. This error is considered to be a minor error - see [Section 5.9.2, “Minor Link Physical Errors Events,” on page 236.](#)

5.3.3 BLOCK SYNCHRONIZATION

The receiver lane must acquire block lock before it can start processing the incoming bit stream.

A receiver will acquire block lock by detecting 64 consecutive blocks with a transition (10 or 01) in the sync header bits (positions 0 and 1 in the 66 bit block). If the receiver is unable to acquire block lock in the tested block position, the receiver shall try another possible position out of the 66 possible bit positions.

The receiver will change the status from “block lock” (`block_lock=true`) to “loss of block lock” (`block_lock=false`) by detecting 65 invalid sync header within a 1024 block window. An invalid sync header is a 11 or 00.

Acquiring or loss of block lock is implementation specific. For an informative example block lock FSM implementation, see [Section 5.3.3.1, “Block Lock Finite State Machine,” on page 101.](#)

5.3.3.1 BLOCK LOCK FINITE STATE MACHINE

This section describes an informative implementation example of the block lock FSM (Finite State Machine). The implementation of the block lock FSM is left up to the implementer as long as it meets the conditions described in [Section 5.3.3, “Block synchronization,” on page 101](#) and [Figure 18 on page 102.](#)

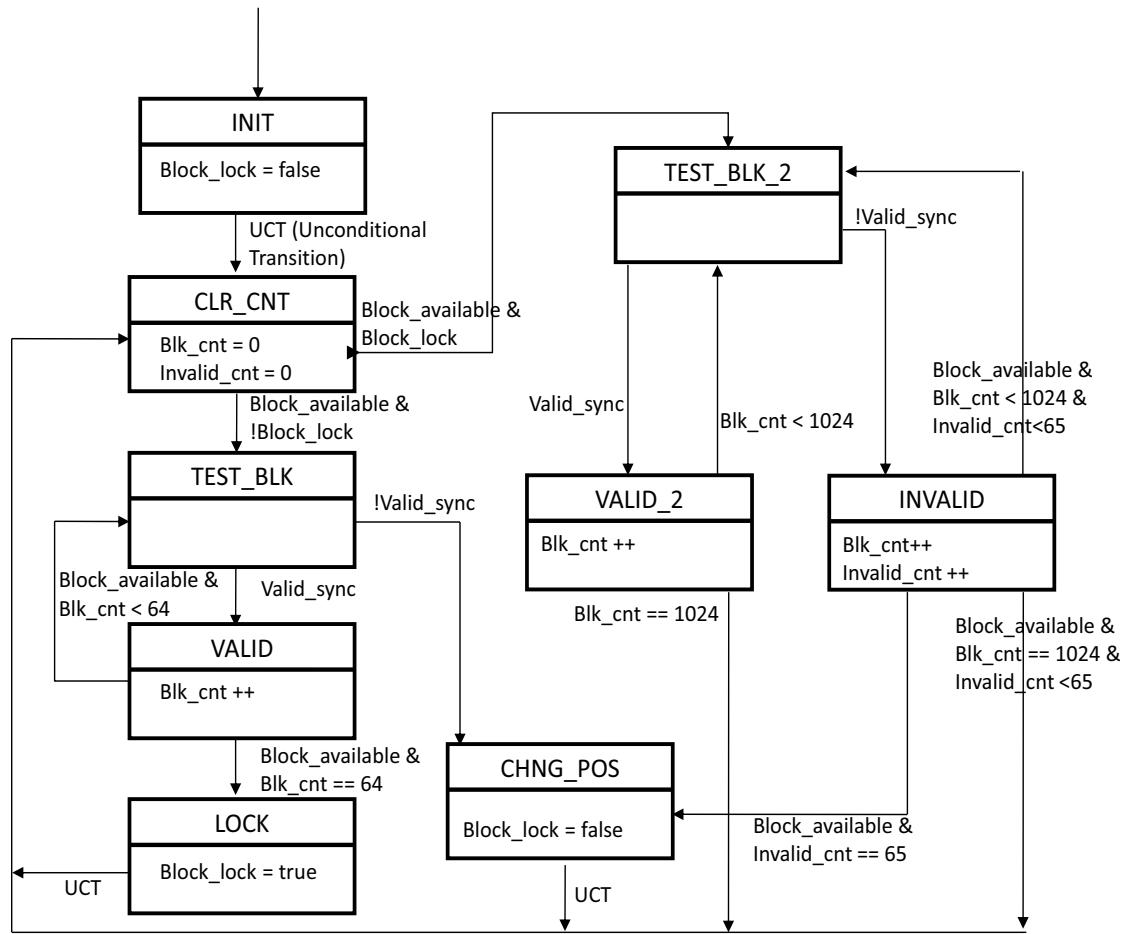


Figure 18 Block lock Finite StateMachine

5.3.3.1.1 INIT STATE

The INIT state is entered after transition from 8b/10b encoding to 64b/66b encoding. Block_lock is set to false.

- 1) Block_lock is set to false
- 2) The next state is CLR_CNT

5.3.3.1.2 CLR_CNT STATE

In the Clear Counters state the Blk_cnt and the Invalid_cnt are reset to 0.

- 1) Blk_cnt is set to 0
- 2) Invalid_cnt is set to 0
- 3) If a 66-bit block is available for evaluation and block_lock is false, the next state is TEST_BLK

-
- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
- 4) If a 66-bit block is available for evaluation and block_lock is true, the next state is TEST_BLK_2

5.3.3.1.3 TEST_BLK STATE

In the Test Block state the two sync header bits in positions 0 and 1 of the 66 bits are evaluated. If the two bits are not identical (i.e. 10 or 01) then Valid_sync is set to true, else Valid_sync is set to false.

- 1) If Valid_sync is true, the next state is VALID
- 2) If Valid_sync is false, the next state is CHNG_POS

5.3.3.1.4 TEST_BLK_2 STATE

In the Test Block 2 state the two sync header bits in positions 0 and 1 of the 66 bits are evaluated. If the two bits are not identical (i.e. 10 or 01) then Valid_sync is set to true, else Valid_sync is set to false.

- 1) If Valid_sync is true, the next state is VALID_2
- 2) If Valid_sync is false, the next state is INVALID

5.3.3.1.5 VALID STATE

In the VALID state the number of consecutive valid blocks is checked.

- 1) Blk_cnt is incremented by one.
- 2) If a new 66 bit block is available for evaluation and Blk_cnt is less than 64, the next state is TEST_BLK
- 3) If a new 66 bit block is available for evaluation and Blk_cnt is equal to 64, the next state is LOCK

5.3.3.1.6 INVALID STATE

In the Invalid state the number of invalid blocks out of the last 1024 tested blocks is checked.

- 1) Invalid_cnt and Blk_cnt are each incremented by one.
- 2) If a new 66 bit block is available for evaluation, and Invalid_cnt is equal to 65, the next state is CHNG_POS.
- 3) If a new 66 bit block is available for evaluation, and Invalid_cnt is less than 65, and Blk_cnt is less than 1024, the next state is TEST_BLK_2.
- 4) If a new 66 bit block is available for evaluation, and Invalid_cnt is less than 65, and Blk_cnt is less than 1024, the next state is CLR_CNT.

5.3.3.1.7 VALID_2 STATE

In the VALID_2 state the number of consecutive valid blocks is checked.

- 1) Blk_cnt is incremented by one.
- 2) If a new 66 bit block is available and Blk_cnt is less than 1024, the next state is TEST_BLK_2.

- 3) If a new 66 bit block is available and Blk_cnt is equal to 1024 the next state is CLR_CNT.

5.3.3.1.8 CHNG_POS STATE

In the CHNG_POS (Change Position) state, Block_lock is set to false, and the next candidate sync header position to be tested is selected by shifting the 66 bits block by one bit.

- 1) Block_lock is set to false
- 2) Next state is CLR_CNT.

5.3.3.1.9 LOCK STATE

In the LOCK state the Block_lock is set to true.

- 1) Next state is CLR_CNT.

5.3.4 ERROR DETECTION PER LANE

In 8b/10b encoding, a bit flip is highly likely to cause a running disparity or decode error, allowing error detection per lane at the Phy layer. The 64b/66b encoding does not allow associating a bit flip error to a single lane. Therefore, error detection per lane is added to the 64b/66b encoding to allow isolation of link errors to a particular physical lane.

For this purpose, Error Detection Per Lane (EDPL) is used in the SKIP ordered-sets. EDPL is the calculated CRC-8 on the entire bit stream (including the sync headers) from the first transmitted non-SKIP Block to the last transmitted non-SKIP block. EDPL is calculated prior to the scrambler. The SKIP blocks are excluded from EDPL calculation. The seed used for calculating the EDPL is FFh, and is reset when transmitting SKIP blocks.

The transmitter lane places the 1's complement of the calculated EDPL on the last byte (8) of the transmitted SKIP blocks, as shown in [Figure 20](#). CRC-8 is performed by applying the polynomial shown below and in [Figure 19](#) on the bit stream in the order it is transmitted - from bit 0 through bit 65.

$$g(x) = x^8 + x^2 + x + 1$$

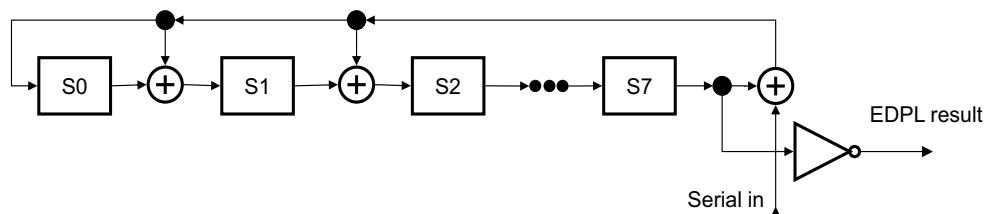


Figure 19 EDPL calculator

A receiver lane calculates EDPL by performing CRC-8, as shown in [Figure 19](#), on all the received bit stream (including the sync header) after the de-scrambler. The receiver compares the calculated EDPL with the received EDPL in the received SKIP block, as shown in [Figure 20](#). If the two are not equal the receiver increments the EDPL error counter. A receiver may optionally compare the EDPL value in all received SKIP blocks. If the EDPL value is not identical in all the received blocks the receiver may optionally increment the EDPL counter; a receiver that does not perform this optional check shall use the first received SKIP block for the EDPL check.

The EDPL CRC-8 calculation is not used for packet-level error detection, and is not used in the algorithm for replacing EGP with EBP packet delimiters, as described in [Section 5.5.3.4, “End of Bad Packet Delimiter \(EBP\).” on page 161](#).

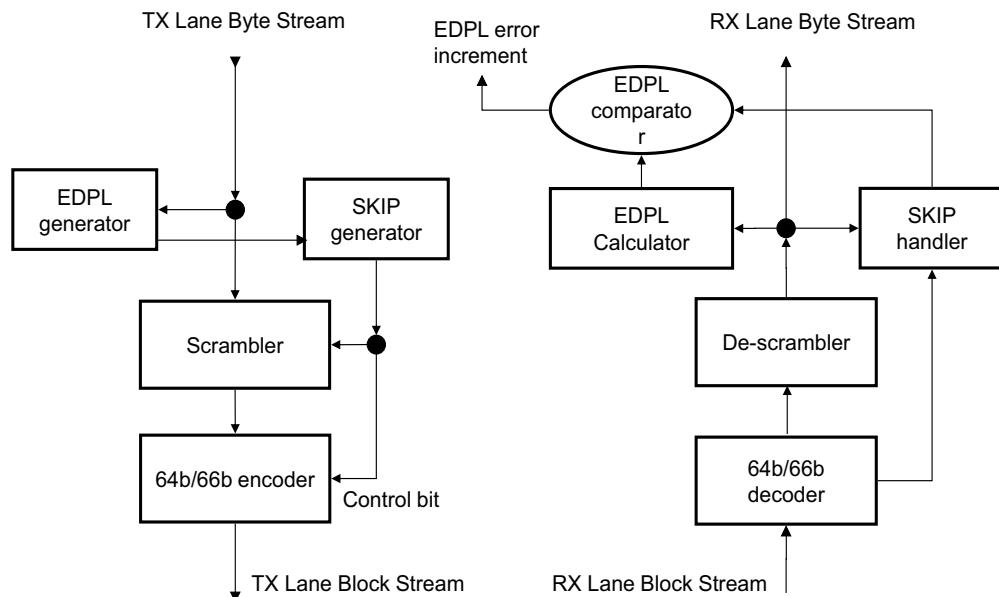


Figure 20 EDPL block diagram

5.3.4.1 EDPL EXAMPLE

This EDPL calculation example is for a single block. The EDPL initial value for this example is FFh.

The Byte 7 down to Byte 0 values are: 17h, 0Eh, 00h, 5Ch, 37h, 12h, 70h and 78, respectively. Sync header is 01.

The calculated EDPL after applying the CRC-8 polynomial is E1h.

5.4 FORWARD ERROR CORRECTION

FEC provides additional protection against signal integrity interference and increases BER performance. Moreover, FEC provides additional margin for channel, manufacturing, and environmental variations.

Forward Error Correction (FEC) is only applicable to 64b/66b encoding.

This specification supports two types of Forward Error Corrections: a Fire-Code FEC and a Reed-Solomon Code FEC (RS-FEC). Two variations for RS-FEC are defined - an RS-FEC(528,514) and an RS-FEC(271, 257), also referred to as Low-Latency FEC, which provides lower latency with more capability to correct errors, at a cost of somewhat more overhead.

- The Fire-Code FEC support is mandatory for FDR data rate and prohibited for EDR data rate.
- Support for the Reed-Solomon Code FEC, RS-FEC(528,514), is mandatory for EDR data rate and optional for FDR data rate.
- Support for the Low-Latency FEC, RS-FEC(271,257), is optional for both FDR and EDR data rates.

Activation of any of the supported FEC codes is optional and may be negotiated between the ports during the link training. The ports negotiate the FEC activation and the FEC type by advertising a support for the optional FEC capabilities and requesting to enable an FEC encoding. An EDR capable port shall allow to disable the FEC request by setting **SM.PortInfo.CapabilityMask(IsFECDisableSupported)** and disable the FEC request when **SM.PortInfo(FECDisable)** is set.

The main difference between the two main FEC types, beside the code type, is that the Fire-Code FEC encodes each physical lane independently, while the two RS-FEC types encode all the lanes of a port together.

Architectural Note

EDR compliant links are designed to achieve the BER target with RS-FEC enabled. However, in some systems, the desired BER can be achieved without FEC protection. In addition, in some systems a centralized control of the FEC type may be desired.

In order to enable the Subnet Manager to control and monitor the FEC protection on the link, an EDR port advertises the supported and enabled FEC types and active FEC type and allows the SM to set the enabled FEC types on the link.

During the tuning, a port can evaluate the link quality, determine whether FEC protection is required and negotiate a FEC type that will meet the BER requirements.

Therefore it is recommended to set the enabled FEC types to FECModeSupported. A Subnet Manager should not restrict the enabled FEC types or disable the FEC protection on a link unless it has the information that the link can achieve the BER requirement with that FEC setting.

Further details on SM management of FEC usage are described in [Section 5.6, “Management Datagram Control and Status Interface,” on page 165](#).

5.4.1 FIRE-CODE FEC

The Fire-Code FEC function encodes the Lane Block Stream to create the FEC Lane Stream as illustrated in [Figure 9](#).

Schematic descriptions of the transmitter and receiver Fire-Code FEC functions are shown in [Figure 21](#). The Fire-Code FEC transmitter scrambles and codes the 64-bit payload of the 66 bit-block and transmits it. The Fire-Code FEC transmitter uses one redundant bit to replace the two sync header bits. The Fire-Code FEC transmitter adds 32 bits to each FEC block for parity checking. The Fire-Code FEC receiver de-scrambles the re-

ceived data, acquires FEC block lock, decodes the FEC block and reconstructs the 64b/66b blocks.

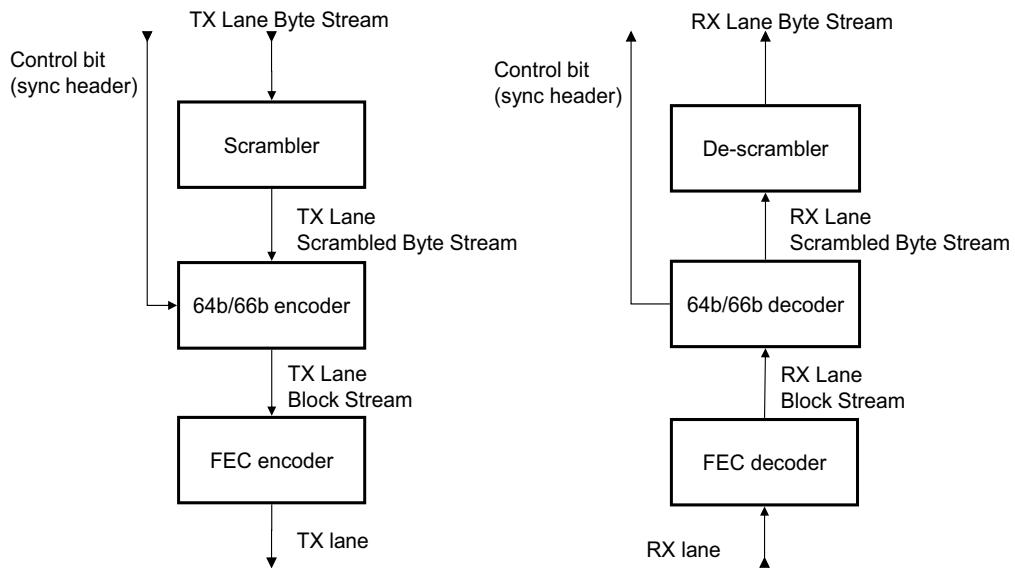


Figure 21 Fire-Code FEC layer block diagram

The Fire-Code FEC block includes 32 rows of 65 bits each; and 32 additional parity bits.

The Fire-Code FEC code to be used is a shortened cyclic code (2112, 2080) for error checking and forward error correction. The Fire-Code FEC block length is 2112 bits. The code encodes 2080 bits of payload (or information symbols) and adds 32 bits of overhead (or parity symbols). The code is systematic; that is the information symbols are not disturbed in anyway in the encoder and the parity symbols are added separately to the end of each block.

The (2112, 2080) code is constructed by shortening the cyclic code (42987, 42955). The shortened cyclic code (2112, 2080) is guaranteed to correct an error burst of up to 11 bits per block.

See Blahut [B23] and Lin and Costello [B48] for additional information on cyclic codes and shortened cyclic codes for correcting burst errors.

5.4.1.1 FIRE-CODE FEC BLOCK FORMAT

The Fire-Code FEC block is composed of 2112 bits. The Fire-Code FEC block is composed of 32 rows, each row is 65 bits long. The 65 bit row is composed of 64 bits taken from the 64b/66b block payload and one bit (T bit) of transcoding overhead. The transformation from the 2 bit header to the transcoding bit is done by deleting the first bit of the sync header:

Sync header of 10 is transformed to T=0 as the transcoding bit; and, Sync header of 01 is transformed to T=1 as the transcoding bit.

The Transcoding bit is then further modified by XOR with bit 8 of the 64-bit payload.

The 64-bit payload words are defined as bits 0:63 from left to right.

The FEC block is terminated by a 32-bit parity field.

Table 15 Fire-Code FEC Format

T0	64 bits payload word 0	T1	64 bits payload word 1	T2	64 bits payload word 2	T3	64 bits payload word 3
T4	64 bits payload word 4	T5	64 bits payload word 5	T6	64 bits payload word 6	T7	64 bits payload word 7
T8	64 bits payload word 8	T9	64 bits payload word 9	T10	64 bits payload word 10	T11	64 bits payload word 11
T12	64 bits payload word 12	T13	64 bits payload word 13	T14	64 bits payload word 14	T15	64 bits payload word 15
T16	64 bits payload word 16	T17	64 bits payload word 17	T18	64 bits payload word 18	T19	64 bits payload word 19
T20	64 bits payload word 20	T21	64 bits payload word 21	T22	64 bits payload word 22	T23	64 bits payload word 23
T24	64 bits payload word 24	T25	64 bits payload word 25	T26	64 bits payload word 26	T27	64 bits payload word 27
T28	64 bits payload word 28	T29	64 bits payload word 29	T30	64 bits payload word 30	T31	64 bits payload word 31
32 bits Parity							

The receiver reconstructs the two-bit sync header using the transcoding bit by adding the missing bit. This bit is always the inversion of the transcoding bit. The transcoding bit is reconstructed by XORing the received transcode bit with bit 8 of the received 64-bit payload. The receiver uses the FEC error correction property to acquire FEC block synchronization, as shown in [Figure 25 on page 112](#).

5.4.1.2 FIRE-CODE FEC ENCODER

A schematic description of the Fire-Code FEC encoder is shown in [Figure 23 Fire-Code FEC Tx block diagram on page 111](#).

The Fire-Code FEC encoder encodes 32 blocks of 66 bits into one 2112-bit FEC block.

The two-bit sync header is converted into one transcode bit. To ensure better DC balance, the transcode bit is XORed with the data bit 8 of 64-bit payload.

The 32 rows of 65 bits (total of 2080 bits) are fed into the (2112, 2080) encoder. The encoder produces a 32-bit parity check. The parity check bits are appended to the end of the FEC block, as shown in [Table 15, “Fire-Code FEC Format,” on page 109](#).

The Fire-Code FEC block is further scrambled using the PN-2112 pseudo-noise sequence as shown in [Figure 22](#), and is transmitted on the transmitter lane.

5.4.1.2.1 FEC (2112,2080) ENCODER

The (2112, 2080) code is a shortened cyclic code that can be encoded by the generated polynomial $g(x)$ below.

$$g(x) = x^{32} + x^{23} + x^{21} + x^{11} + x^2 + 1$$

The initial value of the parity polynomial is 00000000h (32 zeros).

The code word $c(x)$ can be calculated as described in the equation below, where $m(x)$ is the polynomial representation of the information bits.

$$p(x) = x^{32}m(x) \bmod g(x)$$

$$c(x) = p(x) + x^{32}m(x)$$

5.4.1.2.2 PN-2112 PSEUDO-NOISE SEQUENCE GENERATOR

The PN-2112 pseudo-noise sequence is performed using the polynomial $r(x)$ shown in equation below and in [Figure 22](#). The initial value of the scrambler is S57=1, Si-1=Si XOR 1, which produces the 101010... binary sequence.

$$r(x) = 1 + x^{39} + x^{58}$$

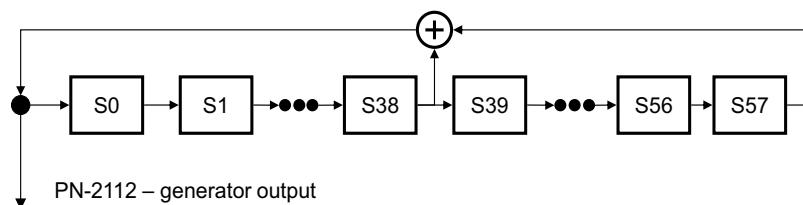


Figure 22 PN-2112 Scrambler

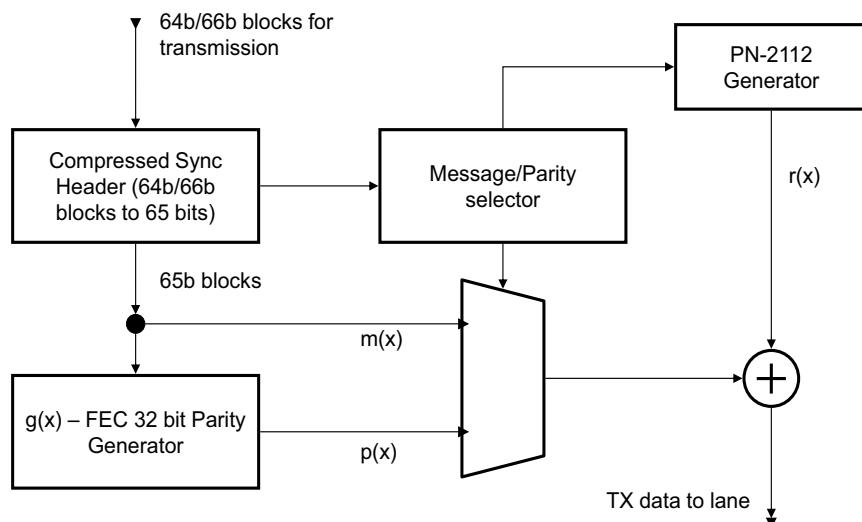


Figure 23 Fire-Code FEC Tx block diagram

5.4.1.3 FIRE-CODE FEC DECODER

The Fire-Code FEC decoder acquires FEC block synchronization, decodes the FEC blocks, corrects the correctable errors, reports correctable and uncorrectable errors, and passes 64b/66b blocks to the receive path. A schematic description of the Fire-Code FEC Decoder is shown in [Figure 24 Fire-Code FEC Rx block diagram on page 111](#).

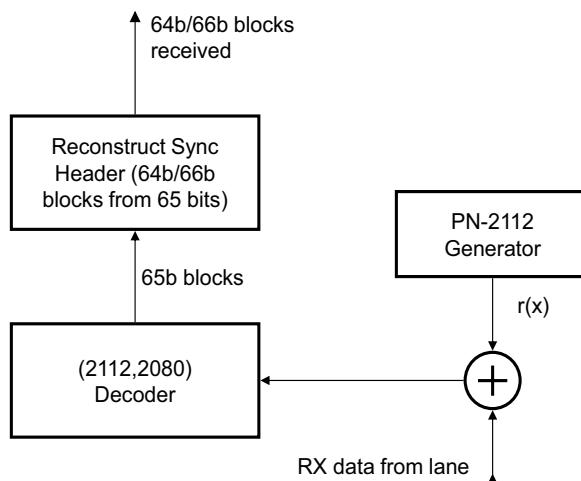


Figure 24 Fire-Code FEC Rx block

5.4.1.3.1 FIRE-CODE FEC ERROR COUNTERS

The Fire-Code FEC function increments one of the two counters - **FECCorrectableBlockCounter** and **FECUncorrectableBlockCounter** - when an error is detected on a received FEC block. The two counters are per-lane counters.

The Fire-Code FEC correctable block is a block that has an invalid parity and has been corrected by the Fire-Code FEC Decoder. The **FECCorrectableBlockCounter** is incremented once for each correctable block.

The Fire-Code FEC uncorrectable block is a block that has an invalid parity and has not been corrected by the Fire-Code FEC Decoder. The **FECUncorrectableBlockCounter** is incremented once for each uncorrectable block.

5.4.1.3.2 FIRE-CODE FEC BLOCK SYNCHRONIZATION

The Fire-Code FEC block synchronization is achieved by de-scrambling using the PN-2112 Generator, as shown in [Figure 22 on page 110](#), testing all possible block positions (2112), and using the parity check as indication of a good block.

Block synchronization is acquired by receiving 4 consecutive good blocks. Block synchronization is lost by receiving 8 consecutive bad blocks. [Figure 25](#) shows an example state machine for assuring FEC block lock.

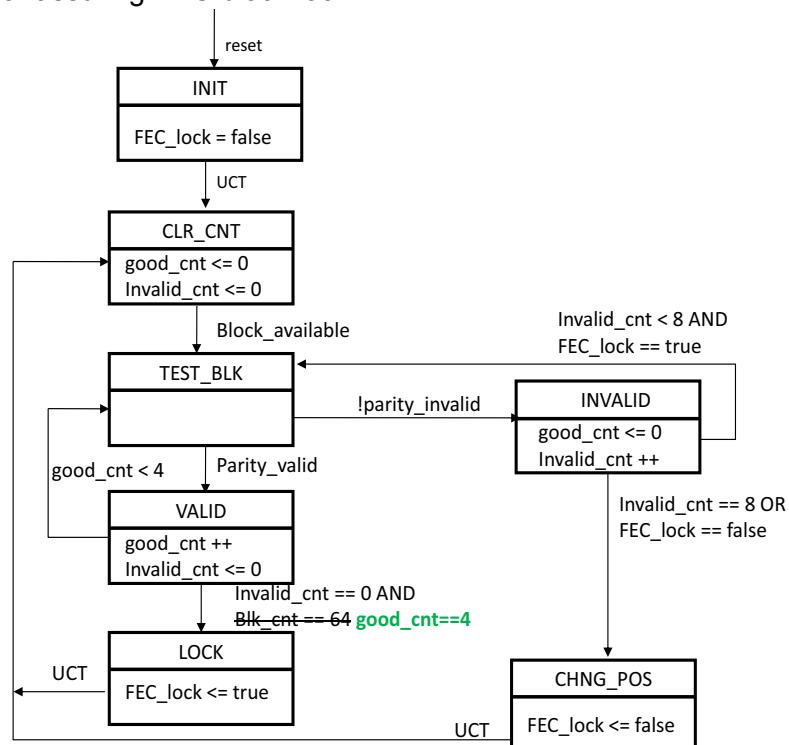


Figure 25 Fire-Code FEC block lock FSM

5.4.1.4 FIRE-CODE FEC TRANSMIT/RECEIVE BIT ORDERING

The transmitted/received bit stream for the Fire-Code FEC blocks is as follows:

Block 0 (65 bits) is transmitted/received first, followed by block 1

The 32 bits parity check are transmitted/received last.

The transcoding bit is transmitted/received first in the 65 bit block, followed by byte 0.

Byte 7 is transmitted/received last in the 65 bits block.

5.4.1.5 FIRE-CODE FEC EXAMPLE

The below example applies the Fire-Code FEC function on a Scrambled Lane Stream. The Scrambled Lane Stream of 2112 bits input to the FEC function is shown in [Table 16](#). This example shows an output of the 64b/66b encoder of Idle data. For the table below the contents are transmitted from left to right within each row and from top to bottom between the rows. The first transmitted bit is on the top left hand corner.

Table 16 Scrambled Lane Stream

sync [0:1]	64 bit payload hex [0:63]	sync [0:1]	64 bit payload hex [0:63]	sync [0:1]	64 bit payload hex [0:63]	sync [0:1]	64 bit payload hex [0:63]
10	40eale77eed301ec	10	ad5a3bf86d9acf5c	10	de55cb85df0f7ca0	10	e6ccff8e8212b1c6
10	d63bc6c309000638	10	70e3b0ce30e0497d	10	dc8df31ec3ab4491	10	66fb9139c81cd37b
10	b57477d4f05e3602	10	8cf495012947a31	10	e7777cf0c6d06280	10	44529cf4b4900528
10	85ce1d27750ad61b	10	456d5c71743f5c69	10	c1bf62e5dc5464b5	10	dc6011be7ea1ed54
10	1cf92c450042a75f	10	cc4b940eaf3140db	10	77bb612a7abf401f	10	c22d341e90545d98
10	ce6daf1f248bbd6d	10	dd22d0b3f9551ed6	10	574686c3f9e93898	10	2e52628f4a1282ce
10	f20c86d71944aab1	10	55133c9333808a2c	10	1aa825d8b817db4d	10	637959989f3021eb
10	976806641b26aae9	10	6a37d4531b7ed5f2	10	53c3e96d3b12fb46	10	528c7eb8481bc969

The output of the (2112, 2080) encoder is shown below in [Table 17](#).

Table 17 FEC (2112, 2080)

T bit [0]	64 bit payload hex [0:63]	T bit [0]	64 bit payload hex [0:63]	T bit [0]	64 bit payload hex [0:63]	T bit [0]	64 bit payload hex [0:63]
1	40eale77eed301ec	0	ad5a3bf86d9acf5c	0	de55cb85df0f7ca0	1	e6ccff8e8212b1c6
0	d63bc6c309000638	1	70e3b0ce30e0497d	1	dc8df31ec3ab4491	1	66fb9139c81cd37b
0	b57477d4f05e3602	1	8cf495012947a31	0	e7777cf0c6d06280	0	44529cf4b4900528
1	85ce1d27750ad61b	0	456d5c71743f5c69	1	c1bf62e5dc5464b5	0	dc6011be7ea1ed54
1	1cf92c450042a75f	0	cc4b940eaf3140db	1	77bb612a7abf401f	0	c22d341e90545d98
0	ce6daf1f248bbd6d	0	dd22d0b3f9551ed6	0	574686c3f9e93898	0	2e52628f4a1282ce
0	f20c86d71944aab1	0	55133c9333808a2c	1	1aa825d8b817db4d	0	637959989f3021eb
0	976806641b26aae9	0	6a37d4531b7ed5f2	1	53c3e96d3b12fb46	1	528c7eb8481bc969
Parity hex [0:31] d96e7685							

The 2112-bit FEC block is further scrambled using PN-2112 as described in [Section 5.4.1.2.2, “PN-2112 pseudo-noise sequence generator,” on page 110](#). The

output of the PN-2112 is in [Table 18. “Example of PN2112 Scrambler Output.” on page 114.](#)

Table 18 Example of PN2112 Scrambler Output

| 64 bit stream
hex [0:63] |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| 5f8af0c4083cd5b6 | 2b57dbab4e33e17d | b1354680bbe0bac1 | 4193315242cb81b6 |
| cc1ba1c9f7b7fe64 | 90838ec46d969470 | a913b019c27f5689 | 7633f46ec762b6d9 |
| d1e410905587d0e4 | f9b66a42540af04a | 9909b64535a725b8 | 5005107c48b4a6aa |
| f9d684ce4396f7a9 | 1b26e0a025c5d0fd | a4f2c62bc4611217 | 3638dc7504ea755e |
| 13fe232e3cdd2a84 | 5c5118ed10f6ffd8 | 5077fba23970c87d | 52ec1279d355fc57 |
| 48263899cc6652da | f746ec8b31bd6b40 | 006f5809784c86a7 | 989b9bd1aab70f0f |
| 57d99a87b9a9cc74 | 09ffb2754f318f33 | ca8fce7654fb1e57 | 03a9c3acc87e6cdd |
| b2574be1e93fcc9a | 26c4fde242df5ca6 | c645fd2bf2d3d525 | 5b25e6d7f9d78153 |
| bd49683cd87b293a | | | |

5.4.2 REED-SOLOMON CODE FEC

The Reed-Solomon Code FEC (RS-FEC) support is mandatory for EDR data rates, and optional for FDR data rates.

A functional block diagram of the Reed-Solomon Code FEC (RS-FEC) is shown in [Figure 26 on page 115.](#)

The Reed-Solomon Code used by this specification uses a symbol size of 10 bits. The specification supports two different RS-FEC codeword sizes - RS(528,514) and RS(271, 257).

Supporting the RS(528, 514)-FEC code is mandatory for EDR ports, whereas supporting the RS(271, 257) code is optional. Both codes can correct at least seven error symbols in arbitrary positions within the codeword.

For details of the codes, see [Section 5.4.2.1.3, “RS-FEC Encoding Function,” on page 123.](#)

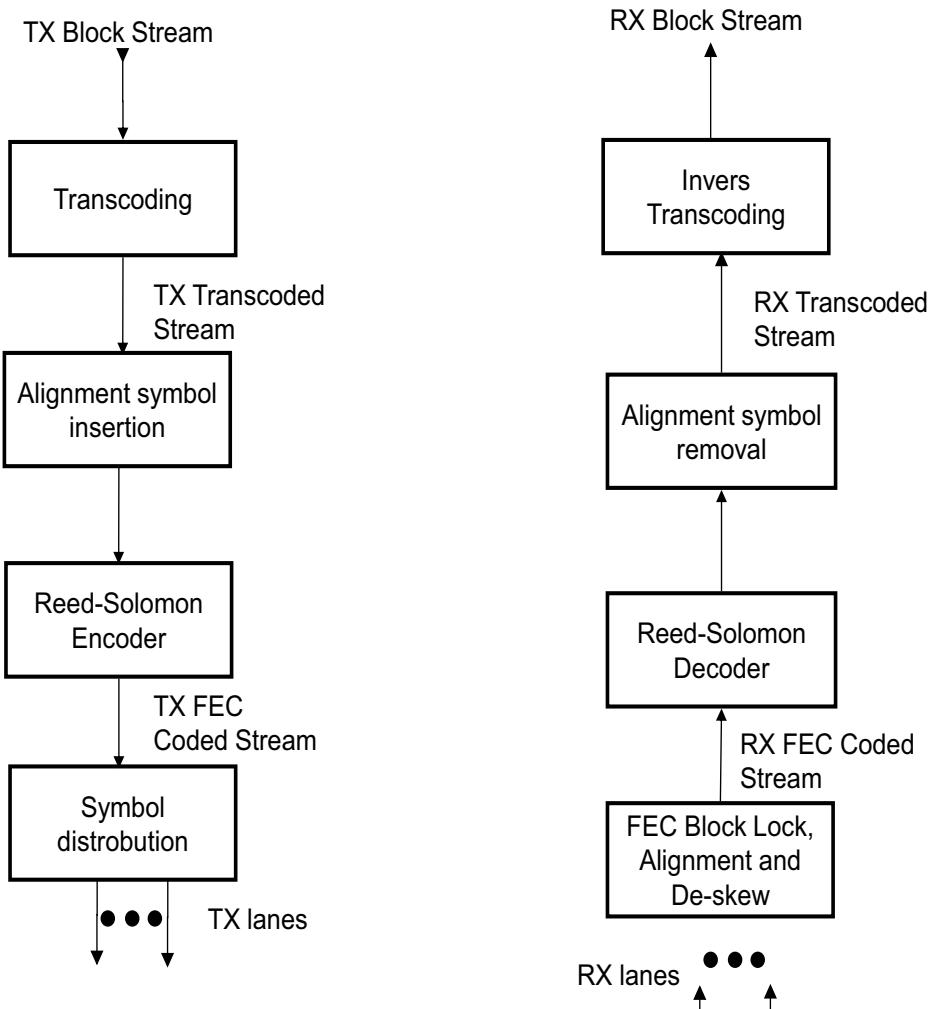


Figure 26 RS-FEC Functional Block Diagram

5.4.2.1 RS-FEC TRANSMIT FUNCTION

The RS-FEC transmit function **shall** receive a transmit block stream, **shall** transcode it into 257-bit transcoding blocks (see [Section 5.4.2.1.1. “Transcoding Function.” on page 116](#)), **shall** add an alignment sequence when needed (see [Section 5.4.2.1.2. “Alignment Sequence Insertion Function.” on page 118](#)), **shall** perform Reed-Solomon encoding to create a RS-FEC codeword (see [Section 5.4.2.1.3. “RS-FEC Encoding Function.” on page 123](#)), and **shall** distribute the FEC codeword symbols across the physical lanes to create a TX lane stream (see [Section 5.4.2.1.5. “Symbol Distribution Function.” on page 128](#)).

5.4.2.1.1 TRANSCODING FUNCTION

The 66b/64b to 256b/257b transcoding function **shall** convert four 66-bit blocks into one 257-bit block.

The first bit of the transcoding block defines the transcoding block type, whether a data transcoding block or a control transcoding block. The transcoding function **shall** indicate a data transcoding block by setting the first bit in the transcoding block to 1. The transcoding function **shall** indicate a control transcoding block by setting the first bit in the transcoding block to 0.

For transcoding examples, see [Figure 27 on page 118](#).

When all four 66-bit blocks are data blocks (sync header = 01), the transcoding function **shall**:

- 1) Remove all the sync header bits
- 2) Indicate the transcoding block as a data transcoding block.
- 3) Place all the 64-bit blocks in order in the following 256 bits of the transcoding block

When at least one of the 66-bit blocks is a control block (sync header = 10), the transcoding function **shall**:

- 1) Remove all the sync header bits
- 2) Indicate the transcoding block as a control transcoding block
- 3) Remove the second nibble (based on transmission order) of the first control block
- 4) Add four bits following the transcoding block type, indicating which of the four blocks are control blocks. A value of 1 indicates a data block, and a value of 0 indicates a control block.
- 5) Place the remaining 252 bits constructed of the four 64-bit blocks payload without the second nibble of the first control block

The transcoding function **shall** scramble the first five bits of the transcoding block.

The transcoding function **shall** comply with the following detailed description:

Let the four 66-bit blocks subject for transcoding be $tx_block[i]<65:0>$ where $i=0\text{-}3$.

Let the output of the transcoding block be $tx_transcoded<256:0>$.

If for all $i=0\text{-}3$ $tx_block[i]<0>=0$ and $tx_block[i]<1>=1$ (all four are data blocks), then the transcoding block shall be constructed as follows:

- 1) $tx_transcoded<0> = 1$ (data transcoding block)
- 2) $tx_transcoded<(64j+64):(1+64j)> = tx_block[j]<65:2>$ for $j=0\text{-}3$

If for any $i=0\text{-}to\text{-}3$ $\text{tx_block}[i]<0>=1$ and $\text{tx_block}[i]=0$ (at least one block is a control block), then the transcoding block **shall** be constructed as follows:

- 1) Let c be the smallest value of i for which $\text{tx_block}[i]<0>=0$ and $\text{tx_block}[i]<1>=1$ (the first control block out of the group of four blocks).
- 2) $\text{tx_transcoded}<0> = 0$ (control block)
- 3) $\text{tx_transcoded}<j+1> = \text{tx_block}[j]<1>$ for $j=0\text{-}to\text{-}3$ (set bit 1 for data and 0 for control)
- 4) Let $\text{tx_payload_before_removal}<255:0>$ be $\text{tx_payload_before_removal}<(63+64j):(64j)> = \text{tx_block}[j]<65:2>$ for $j=0\text{-}to\text{-}3$
- 5) $\text{tx_transcoded}<(64c+8):5>=\text{tx_payload_before_removal}<(64c+3):0>$ (remove the second nibble of the block type). $\text{tx_transcoded}<256:(64c+9)>=\text{tx_payload_before_removal}<255:(64c+8)>$

For examples of transcoding blocks before scrambling, see [Figure 27 on page 118](#).

After constructing of the 257 transcoding block, the transcoding function **shall** scramble the first five bits of the transcoding block by XORing them with bits 12:8 of the transcoding block.

The transcoding function scrambler **shall** comply with the following detailed description:

Let $\text{tx_scrambled}<256:0>$ be the output of the transcoding function scrambler.

- 1) Set $\text{tx_scrambled}<4:0>$ to the bit-wise exclusive-OR (XOR) of $\text{tx_transcoded}<4:0>$ and $\text{tx_transcoded}<12:8>$
- 2) $\text{tx_scrambled}<256:5> = \text{tx_transcoded}<256:5>$

The output of the transcoding function shall be transmitted from the LSB bit to the MSB bit such that $\text{tx_scrambled}<0>$ is transmitted first and $\text{tx_scrambled}<256>$ is transmitted last.

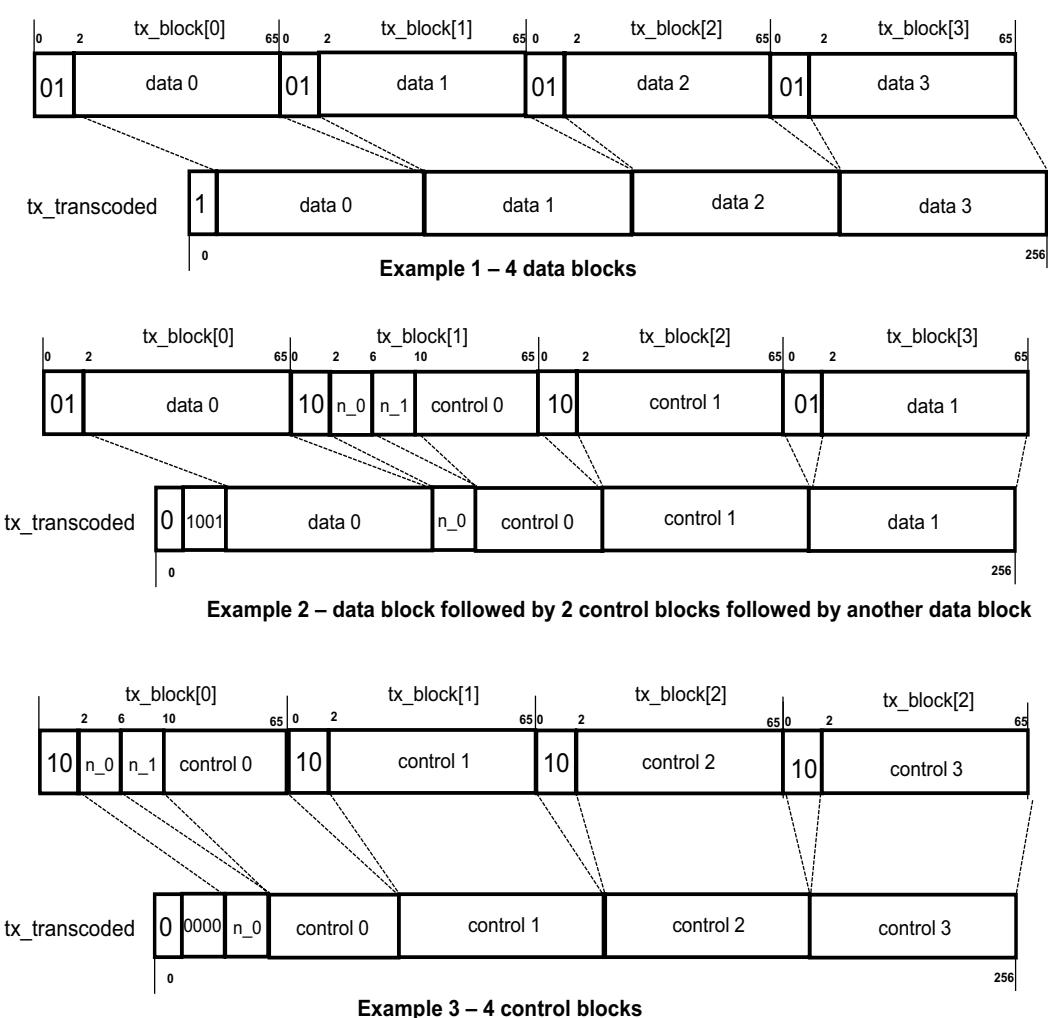


Figure 27 Transcoding Examples

5.4.2.1.2 ALIGNMENT SEQUENCE INSERTION FUNCTION

In order to acquire RS-FEC codeword lock and achieve lane-to-lane de-skew, the RS-FEC transmitter may insert alignment sequences at the start of some FEC codewords. The alignment sequence insertion is only performed during link training and is bypassed when FEC codeword lock and lane-to-lane de-skew are achieved. See [Section 5.8.4.6, “Configuration States - Enhanced Signaling,” on page 203](#) and [Section 5.8.4.8, “Link Error Recovery States,” on page 222](#),

When Alignment Sequence Insertion is enabled as indicated by the align_enable variable, for link width of x1 or x4, the Alignment Insertion function **shall** insert one transcoding block containing an alignment sequence at the beginning of every 16th RS-FEC codeword. For link width of x8, the Alignment Sequence Insertion function **shall** in-

sert two transcoding blocks on every 32th RS-FEC codeword. For link width of x12, the Alignment Sequence Insertion function **shall** insert three transcoding blocks on every 48th RS-FEC codeword.

The result of distributing the transcoding block across the lanes as performed by the Symbol Distribution function (see [Section 5.4.2.1.5, "Symbol Distribution Function," on page 128](#)) **shall** be that the first five symbols (50 bits) on all lanes are the alignment sequence, the 6th symbol **shall** be the lane ID symbol. The remaining bits in the transcoding block are padding bits.

5.4.2.1.2.2 ALIGNMENT SEQUENCE AND LANE ID

As the alignment sequence is never scrambled, it shall ensure DC balance by transmitting the same number of 1s and 0s.

In order to avoid a non-alignment sequence, in which the transcoding block is identified as alignment sequence block and vice versa, the format of the alignment sequence shall be such that an attempt to perform inverse transcoding function (see [Section 5.4.2.2.4, "Inverse Transcoding Function," on page 137](#)) will result in an invalid transcoding block, as it is marked as a control block, but all four blocks are marked as data blocks. When a block encoded as alignment sequence is received unexpectedly (RxCMD other than EnDeSkew), the block shall be discarded and shall not be passed to the Rx Block stream.

The five (5) symbols of the alignment sequence are:

Symbol 0 = 0x2AC
Symbol 1 = 0x284
Symbol 2 = 0x255
Symbol 3 = 0x2B6
Symbol 4 = 0x336

Symbol 0 is transmitted first, and symbol 4 is transmitted last. Each symbol is transmitted from the LSB bit to the MSB bit.

The bit-wise value of the alignment sequence is as follows, transmitted from left to right:

0011010101 0010000101 1010101001 0110110110 0110110011

The 6th symbol is the lane ID symbol. The lane ID is constructed from 5 bits of the lane number and 5 bits of the inverse of the lane ID number: {lane number, inverse lane number} transmitted LSB to MSB.

An example of the lane ID symbol for lane number 1

Lane ID symbol for lane 1 = 0x03E (lane number = 00001, inverse lane number = 11110)

The bit-wise value of the lane 1 ID is as follows, transmitted from left to right:

0111110000

5.4.2.1.2.3 ALIGNMENT SEQUENCE TRANSCODING BLOCK

For a x1 link width:

The transcoding block shall contain an alignment sequence and a lane ID symbol, followed by 197 bits of padding.

The transcoding block is constructed such that:

Symbol 0 = alignment sequence symbol 0

Symbol 1 = alignment sequence symbol 1

Symbol 2 = alignment sequence symbol 2

Symbol 3 = alignment sequence symbol 3

Symbol 4 = alignment sequence symbol 4

Symbol 5 = Lane ID 0 symbol, followed by 19 Pad 0 = 0x26C symbols

The last 7 bits are PAD bits 0x32.

For a x4 link width:

The transcoding block shall contain four (4) alignment sequences, four lane ID symbols, and 17 bits of padding.

The transcoding block is constructed such that:

Symbols 3:0 = alignment sequence symbol 0

Symbols 7:4 = alignment sequence symbol 1

Symbols 11:8 = alignment sequence symbol 2

Symbols 15:12 = alignment sequence symbol 3

Symbols 19:16 = alignment sequence symbol 4

Symbol 20 = Lane ID 0 symbol

Symbol 21 = Lane ID 1 symbol

Symbol 22 = Lane ID 2 symbol

Symbol 23 = Lane ID 3 symbol

Symbol 24 = Pad 0 = 0x26C

The last 7 bits are PAD bits 0x32. See [Figure 28 on page 121](#). An example of the AM transcoding block after the symbol distribution function is in [Figure 29 on page 121](#).

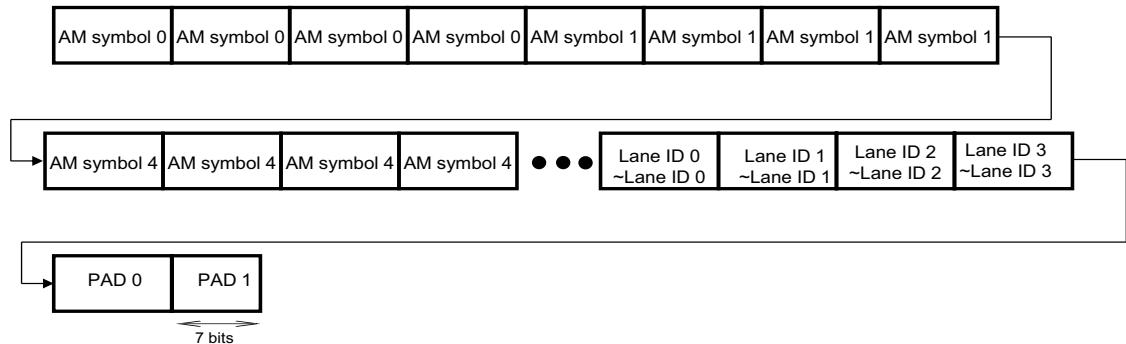


Figure 28 Alignment Transcoding Block for x4 Link Width

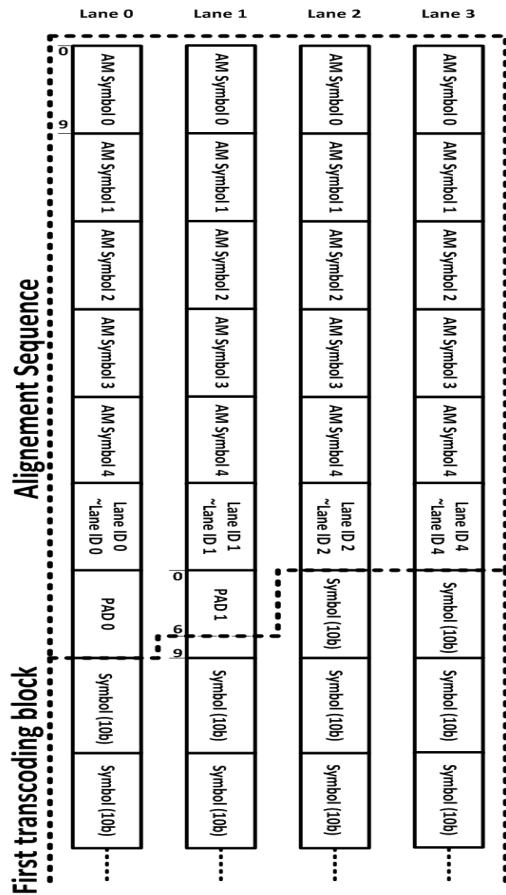


Figure 29 Alignment Sequence After Lane Distribution for x4 Link Width

For a x8 link width:

The two transcoding blocks shall contain eight (8) alignment sequences, eight Lane ID symbols, and 34 bits of padding. The two transcoding blocks contain 8 AM symbol 0, followed by 8 AM symbol 1, followed by 8 AM symbol 2, followed by 8 AM symbol 3, followed by 8 AM symbol 4, followed by Lane ID 0-to-7, followed by 3 PAD 0 symbols. The last 4 bits are PAD bits 0x5. See [Figure 30 on page 122](#).

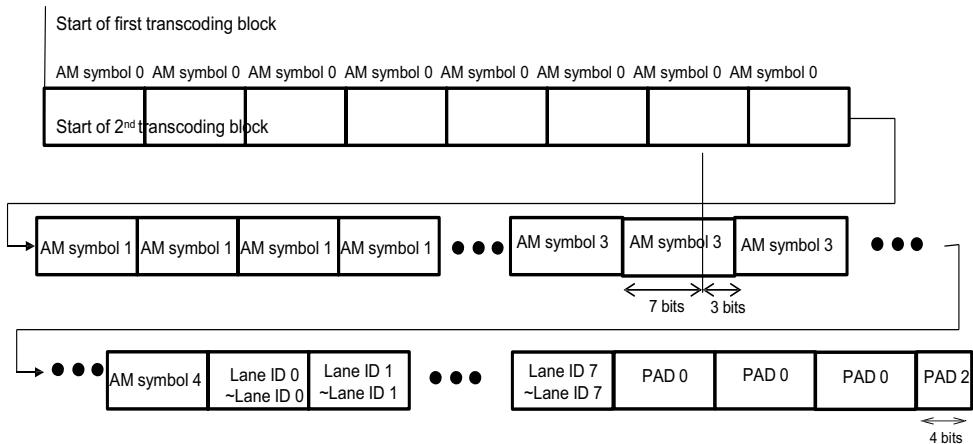


Figure 30 Alignment Transcoding Block for x8 Link Width

For a x12 link width:

The three transcoding blocks shall contain twelve (12) alignment sequences, twelve Lane ID symbols, and 51 bits of padding. The three transcoding blocks contain 12 AM symbol 0, followed by 12 AM symbol 1, followed by 12 AM symbol 2, followed by 12 AM symbol 3, followed by 12 AM symbol 4, followed by Lane ID 0-to-11, followed by 5 PAD 0 symbol. The last bit is PAD bit 0x0. See [Figure 31 on page 122](#).

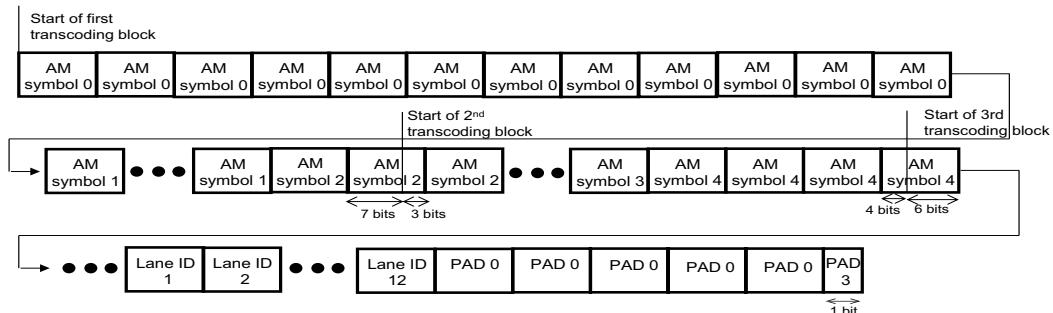


Figure 31 Alignment Transcoding Block for x12 Link Width

5.4.2.1.3 RS-FEC ENCODING FUNCTION

The RS-FEC encoder function employs a Reed-Solomon code operating over the Galois Field GF(2¹⁰), where the symbol size is 10 bits. The encoder processes k message symbols to generate 2t parity symbols, which are then appended to the message to produce a FEC codeword of n=k+2t symbols. For the purposes of this clause, a particular Reed-Solomon code is denoted RS(n, k).

Any port supporting EDR data rate and higher **shall** implement the mandatory RS(528, 514). Any port supporting EDR data rate may implement the optional RS(271, 257). When using the RS(528,514), each k-symbol message corresponds to twenty (20) 257-bit transcoding blocks produced by the transcoding function. When using the RS(271,257), each k-symbol message corresponds to ten (10) 257-bit transcoding blocks produced by the transcoding function. Each code is based on the generating polynomial given by the following equation:

$$g(x) = \prod_{j=0}^{2t-1} (x - \alpha^j) = g_{2t}x^{2t} + g_{2t-1}x^{2t-1} + \dots + g_1x + g_0$$

In the equation above, α is a primitive element of the finite field defined by the polynomial $x^{10} + x^3 + 1$.

The following equation defines the message polynomial m(x), whose coefficients are the message symbols m_{k-1} to m₀:

$$m(x) = m_{k-1}x^{n-1} + m_{k-2}x^{n-2} + \dots + m_1x^{2t+1} + m_0x^{2t}$$

Each message symbol m_i is the bit vector $(m_{i,9}, m_{i,8}, \dots, m_{i,1}, m_{i,0})$, which is identified with the element $m_{i,9}\alpha^9 + m_{i,8}\alpha^8 + \dots + m_{i,1}\alpha + m_{i,0}$ of the finite field. The message symbols are comprised of the symbols output of the transcoding function and of the alignment insertion function when applicable. The first symbol of the first transcoding block is mapped to symbol m_{k-1} when the first bit of the transcoding block is mapped to bit 0 of symbol m_{k-1} . The last symbol of the last transcoding block (20th for RS(528,514) and 10th for RS(271,257)) is mapped to symbol m_0 when the last bit of the transcoding block is mapped to bit 9 of symbol m_0 .

The following equation defines the parity polynomial p(x), whose coefficients are the parity symbols p_{2t-1} to p_0 :

$$p(x) = p_{2t-1}x^{2t-1} + p_{2t-2}x^{2t-2} + \dots + p_1x + p_0$$

The parity polynomial is the remainder from the division of $m(x)$ by $g(x)$. This may be computed using the shift register implementation illustrated in [Figure 32 on page 124](#). The outputs of the delay elements are initialized to zero prior to the computation of the parity for a given message. After the last message symbol, m_0 , is processed by the encoder, the outputs of the delay elements are the parity symbols for that message.

The codeword polynomial $c(x)$ is then the sum of $m(x)$ and $p(x)$, where the coefficient of the highest power of x , $c_{n-1} = m_{k-1}$, is transmitted first, and the coefficient of the lowest power of x , $c_0 = p_0$, is transmitted last. The first bit transmitted from each symbol is bit 0.

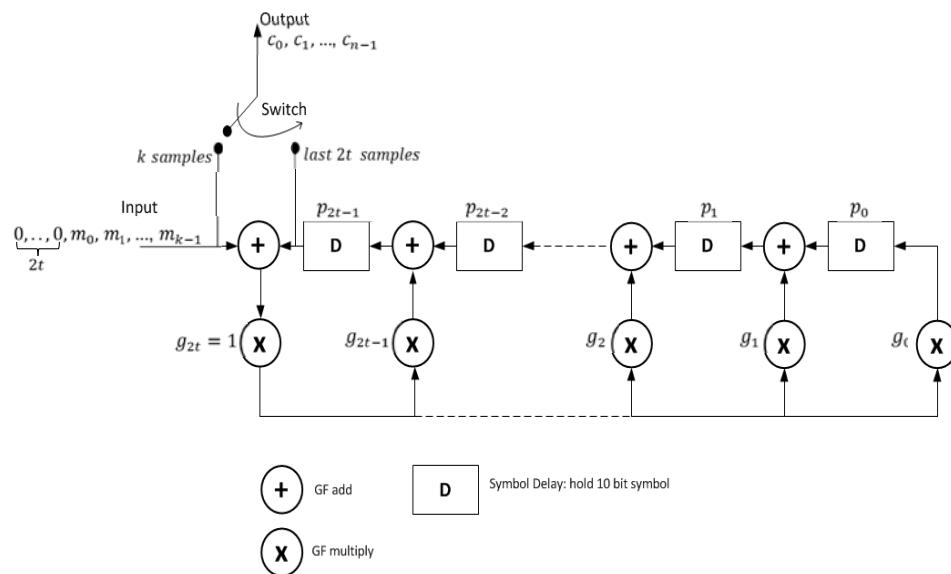


Figure 32 Reed-Solomon Encoder Functional Model

The coefficients of the generator polynomial for each code are presented in [Table 19](#).

Table 19 Coefficients of the Generator Polynomial (Decimal)

i	RS(528,514)	RS(271,257)
0	432	432
1	290	290
2	945	945
3	265	265
4	592	592
5	391	391
6	614	614
7	900	900
8	925	925
9	656	656
10	32	32
11	701	701
12	6	6
13	904	904
14	1	1

5.4.2.1.4 TRANSCODER AND RS-FEC ENCODER EXAMPLE

The following is an example of the output of the transcoding function and the two different RS-FEC encoders:

Table 20 Transcoder Function Input 80 66-Bit Blocks

Synch <0:1>	64-Bit Payload hex<2:65>						
10	ad5a3bf86d9acf5c	10	de55cb85df0f7ca0	10	e6ccff8e8212b1c6	10	d63bc6c309000638
10	70e3b0ce30e0497d	10	dc8df31ec3ab4491	10	66fb9139c81cd37b	10	b57477d4f05e3602
10	8cf495012947a31	10	e7777cf0c6d06280	10	44529cf4b4900528	10	85ce1d27750ad61b
10	456d5c71743f5c69	10	c1bf62e5dc5464b5	10	dc6011be7ea1ed54	10	1cf92c450042a75f
10	cc4b940eaf3140db	10	77bb612a7abf401f	10	c22d341e90545d98	10	ce6daf1f248bbd6d
10	dd22d0b3f9551ed6	10	574686c3f9e93898	10	2e52628f4a1282ce	10	f20c86d71944aab1
10	55133c9333808a2c	10	1aa825d8b817db4d	10	637959989f3021eb	10	976806641b26aae9
10	6a37d4531b7ed5f2	10	53c3e96d3b12fb46	10	528c7eb8481bc969	10	ab8f9980d5a54559
10	9a4d2abfd465cc33	10	94fe646ef5af02d	10	9a65ae5fc88c03a	10	5ef08673168def9b
10	220c871a953fffc6	10	ce0bb95ac263e6c1	10	4f6a917d1a676571	10	5890918c7b687d75
10	44d2b3e43096f836	10	84cdd4fc48b79608	10	b3e4503e3c824a8c	10	fd6d0b1a39687929
10	1730167c08302a69	10	4c15ff56de92b1ad	10	d0c2f0d4ff0dee95	10	e1422ee2e8b92125
10	ed5acaf86592fce	10	de799be0b903c880	10	2714ffbf40bc09f6	10	c3be97c3c285009f
10	1020faf19f606631	10	93007cabbb3f8c9d	10	ef6955f7f43df5d0	10	4dbd0616afe60e1f
10	3a1e49b7c7f7bb5d	10	901d828746ceec61	10	71ed3c097158c224	10	11adb3d81e13d263
10	a350d1a343b2394b	10	eab30ca27b5b34e3	10	90359ef711ed53d9	10	9b446763c8627ea8
10	6e891c0f4842b823	10	c4d786a25727a7fc	10	094fe7da31fb60cd	10	9f9a004de5e70767
10	054bdd77b7cb4e7b	10	c598cb710558af67	10	fc386d1f99d3a925	10	4928e0b43e781893
10	5a44dd3eb8b2ad6c	10	94462af4f583d770	10	8061ba9381f51f55	10	476d4eded7c90fcc
10	1efc25aa6a7e0b4c	10	93dd968c06a56809	10	9768e9d1ba74d3b6	10	014e9dc9f13670bb

The output of the two RS-FEC encoders for the above table is as follows:

Table 21 RS-FEC Encoder Output for RS(528,514)

Header <0:4>	Payload hex<5:64>	Payload hex<65:128>	Payload hex<129:192>	Payload hex<193:256>
00101	a5a3bf86d9acf5c	de55cb85df0f7ca0	e6ccff8e8212b1c6	d63bc6c309000638
11110	7e3b0ce30e0497d	dc8df31ec3ab4491	66fb9139c81cd37b	b57477d4f05e3602
01111	8fd495012947a31	e7777cf0c6d06280	44529cf4b4900528	85ce1d27750ad61b
00110	46d5c71743f5c69	c1bf62e5dc5464b5	dc6011be7ea1ed54	1cf92c450042a75f
00100	c4b940eaf3140db	77bb612a7abf401f	c22d341e90545d98	ce6daf1f248bbd6d
10010	d22d0b3f9551ed6	574686c3f9e93898	2e52628f4a1282ce	f20c86d71944aab1
10001	5133c9333808a2c	1aa825d8b817db4d	637959989f3021eb	976806641b26aae9
00011	637d4531b7ed5f2	53c3e96d3b12fb46	528c7eb8481bc969	ab8f9980d5a54559
10100	94d2abfd46cc33	94fe646efe5af02d	9a65ae5fc8803a	5ef08673168def9b
00000	20c871a953ffffc6	ce0bb95ac263e6c1	4f6a917d1a676571	5890918c7b687d75
01101	4d2b3e43096f836	84cdd4fc48b79608	b3e4503e3c824a8c	fd6d0b1a39687929
10011	130167c08302a69	4c15ff56de92b1ad	d0c2f0d4ff0dee95	e1422ee2e8b92125
00101	e5acaf86592fce	de799be0b903c880	2714ffb40bc09f6	c3be97c3c285009f
10010	120faf19f606631	93007cabbb3f8c9d	ef6955f7f43df5d0	4dbd0616afe60e1f
10001	31e49b7c7f7bb5d	901d828746ceec61	71ed3c097158c224	11adb3d81e13d263
00101	a50d1a343b2394b	eab30ca27b5b34e3	90359ef711ed53d9	9b446763c8627ea8
01000	6891c0f4842b823	c4d786a25727a7fc	094fe7da31fb60cd	9f9a004de5e70767
00100	04bdd77b7cb4e7b	c598cb710558af67	fc386d1f99d3a925	4928e0b43e781893
10100	544dd3eb8b2ad6c	94462af4f583d770	8061ba9381f51f55	476d4ded7c90fcc
11111	1fc25aa6a7e0b4c	93dd968c06a56809	9768e9d1ba74d3b6	014e9dc9f13670bb
Parity hex<0:64>	Parity hex<65:127>	Parity hex<128:139>		
ed0e78f1734bc808	a38c0c417bd68f36	825		

Table 22 RS-FEC Encoder Output for RS(271,257)

Header <0:4>	Payload hex<5:64>	Payload hex<65:128>	Payload hex<129:192>	Payload hex<193:256>
00101	a5a3bf86d9acf5c	de55cb85df0f7ca0	e6ccff8e8212b1c6	d63bc6c309000638
11110	7e3b0ce30e0497d	dc8df31ec3ab4491	66fb9139c81cd37b	b57477d4f05e3602
01111	8fd495012947a31	e7777cf0c6d06280	44529cf4b4900528	85ce1d27750ad61b
00110	46d5c71743f5c69	c1bf62e5dc5464b5	dc6011be7ea1ed54	1cf92c450042a75f
00100	c4b940eaf3140db	77bb612a7abf401f	c22d341e90545d98	ce6daf1f248bbd6d
10010	d22d0b3f9551ed6	574686c3f9e93898	2e52628f4a1282ce	f20c86d71944aab1
10001	5133c9333808a2c	1aa825d8b817db4d	637959989f3021eb	976806641b26aae9
00011	637d4531b7ed5f2	53c3e96d3b12fb46	528c7eb8481bc969	ab8f9980d5a54559
10100	94d2abfda65cc33	94fe646efe5af02d	9a65ae5fc88c03a	5ef08673168def9b
00000	20c871a953ffffc6	ce0bb95ac263e6c1	4f6a917d1a676571	5890918c7b687d75
Parity hex<0:64>	Parity hex<65:127>	Parity hex<128:139>		

5.4.2.1.5 SYMBOL DISTRIBUTION FUNCTION

The symbol distribution function **shall** distribute the FEC codeword across the link lanes on a symbol boundary so that for an N lane link the first 10-bit symbol shall be transmitted on lane 0, the second 10-bit symbol shall be transmitted on lane 1, the Nth 10-bit symbol shall be transmitted on lane N-1, and the N+1th 10-bit symbol shall be transmitted on lane 0.

For the RS(528,514) code for a x4 link width, 132 symbols are transmitted on each lane. For a x8 link width, 66 symbols are transmitted on each lane, and for a x12 link width, 44 symbols are transmitted on each lane. See [Figure 33 on page 129](#).

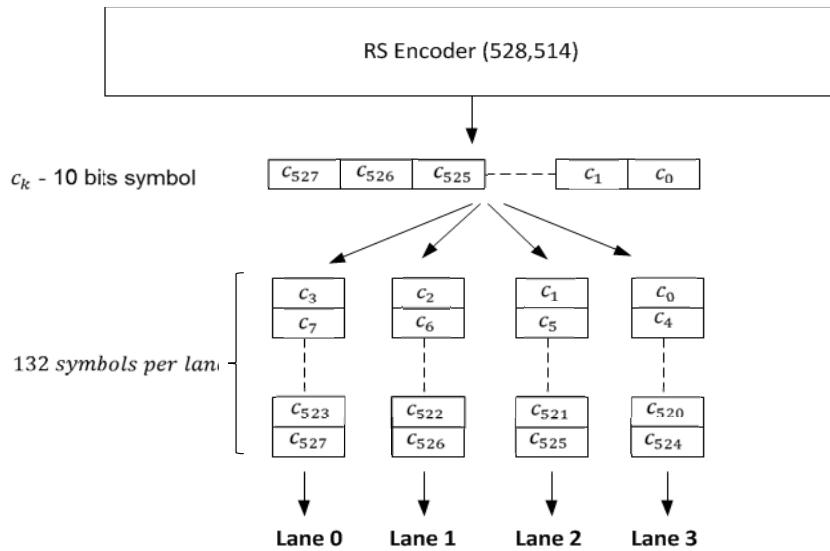


Figure 33 Symbol Distribution RS(528,514) for a 4x Link

For the RS(271,257) code, for any link width greater than x1, 271 is not a multiple of link width, therefore the first symbol of any codeword is not necessarily transmitted on lane 0. For example, for a 4x link width the first symbol of the first codeword is transmitted on lane 0, the last symbol of the first codeword is then transmitted on lane 2; the first symbol of the second codeword is transmitted on lane 3, and the last symbol of the second codeword is then transmitted on lane 1; the first symbol of the third codeword is transmitted on lane 2, and the last symbol of the third codeword is then transmitted on lane 0; the first symbol of the fourth codeword is transmitted on lane 1, and the last symbol of the fourth codeword is then transmitted on lane 3. This is repeated every four codewords. For 8x and 12x link widths, respectively, the cycles are 8 and 12 codewords. See [Figure 34 on page 130](#).

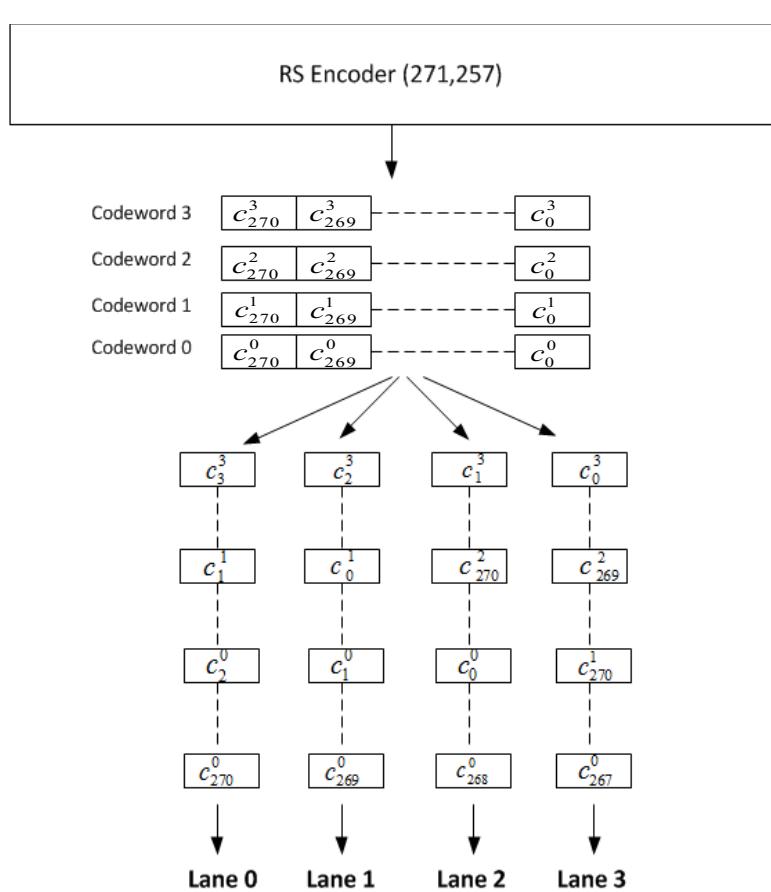


Figure 34 Symbol Distribution RS(271,257) for a 4x Link

All the transmit bit ordering is illustrated in [Figure 35 on page 131](#).

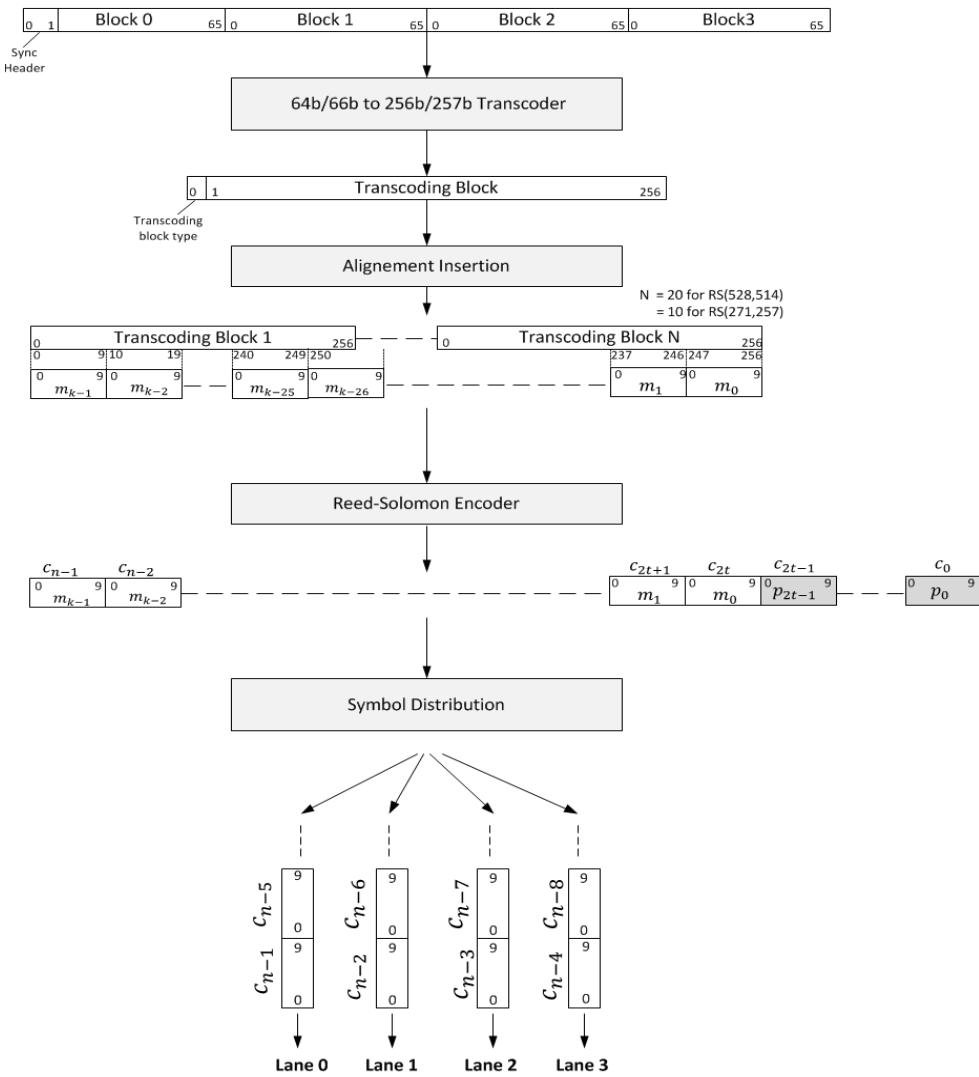


Figure 35 Transmit Bit Ordering

5.4.2.2 RS-FEC RECEIVE FUNCTION

The RS-FEC receive function **shall** perform lane-to-lane de-skew and acquire FEC codeword lock when necessary, extract the message symbols from the codeword, correct them as necessary and discard the parity symbols, remove alignment sequences, and perform inverse transcoding. The message symbols correspond to 20 transcoded blocks when using the RS(528,514) code and 10 transcoded blocks when using RS(271,257) code.

The RS-FEC decoder **shall** be capable of correcting any combination of up to $t=7$ symbol errors in a codeword.

5.4.2.2.1 FEC LOCK, ALIGNMENT AND DE-SKEW FUNCTION

During the initial link training, the RS-FEC receiver needs to acquire FEC codeword lock and to align and de-skew the lanes. When alignment is enabled as indicated by the LinkPhy state machine (see [Figure 56 on page 196](#)), the Alignment function **shall** look for the unique alignment sequence on all lanes. The alignment function shall tolerate up to three nibbles of error when comparing the received stream to the alignment sequence (**the five alignment sequence symbols - 50 bits**) due to the lack of FEC correction. After detecting an alignment sequence candidate, the alignment function **shall** look for a second alignment sequence at the expected offset. For x1 link width, the expected offset is 16 RS-FEC codewords, for x4,x8,x12 link width, the expected offset per lane equals the length of four RS-FEC codewords (respectively, 16, 32, and 48 divided by 4, 8, and 12). After detecting a second alignment sequence at the expected offset, the alignment lock function **shall** move to alignment lock state. The detailed description of the alignment function is described in [Section 5.4.2.2.1.4, “Alignment Lock State Machine,” on page 132](#).

Once the alignment function has acquired alignment lock on all lanes, the de-skew function **shall** de-skew the lanes as needed so that each lane is based on the alignment sequences received on all lanes.

After the lanes have been de-skewed, the FEC lock function shall lock on the FEC codeword. As long as the FEC lock function reports FEC lock, the alignment and de-skew functions are disabled and shall not attempt to re-acquire alignment or de-skew. For details of the FEC lock function, see [Section 5.4.2.2.1.5, “RS-FEC Lock State Machine,” on page 134](#).

5.4.2.2.1.4 ALIGNMENT LOCK STATE MACHINE

Variables:

- **candidate_available** - This boolean variable is set to true when the next 50-bit candidate for alignment sequence is made available. The method in which an implementation chooses the next candidate position is implementation specific and is not covered by this specification.
- **valid_al_symbol** - This boolean variable is set to true if the 50-bit alignment sequence candidate meets the following conditions: the alignment lock function **shall** divide the first 48 bits (in the order received) into twelve 4-bit nibbles; the alignment lock function **shall** bit-wise compare each nibble to the expected value of the alignment sequence as defined in [Section 5.4.2.1.2.2, “Alignment Sequence and Lane ID,” on page 119](#); if nine nibbles or more match the expected value, the alignment lock function **shall** set the valid_al_symbol to true; otherwise, it **shall** set it to false.
- **next_count_done** - This is a variable set to true when the next_cnt counter reaches its terminate value. The next_cnt counter counts the interval between the position of two consecutive alignment sequence positions. For x1 link width, the counter terminate value shall be 16, while for other link widths (x4, x8, and x12) it shall be 4.
- **align_enable** - Boolean variable set to true by the LinkPhy state machine when alignment and de-skew are enabled.

- **fec_fail** - Boolean variable set to true by the FEC lock state machine when the FEC lock loses lock due to three consecutive uncorrectable FEC codewords.

INIT state:

This state is entered when alignment is enabled by the linkphy state machine. Set lane_lock to false.

Next state is GET_CANDIDATE state.

GET_CANDIDATE state:

In this state, the alignment state machine waits for the next alignment sequence candidate.

Next state is TEST_ALIGN state if candidate available is true.

TEST_ALIGN state:

In this state, the alignment state machine compares the alignment sequence candidate.

Next state is VALID state if valid_al_symbol is true.

Next state is GET_CANDIDATE if valid_al_symbol is false.

VALID state:

In this state, the alignment state machine waits for the next alignment sequence position. The alignment state machine starts the next_cnt when entering this state.

Next state is TEST_ALIGN2 state when next_cnt_done.

TEST_ALIGN2 state:

In this state, the alignment state machine compares the alignment sequence candidate.

Next state is LOCK state if valid_al_symbol is true.

Next state is GET_CANDIDATE if valid_al_symbol is false.

LOCK state:

In this state, the alignment state machine compares sets the lane_lock to true.

Next state is INIT state if align_enable is true.

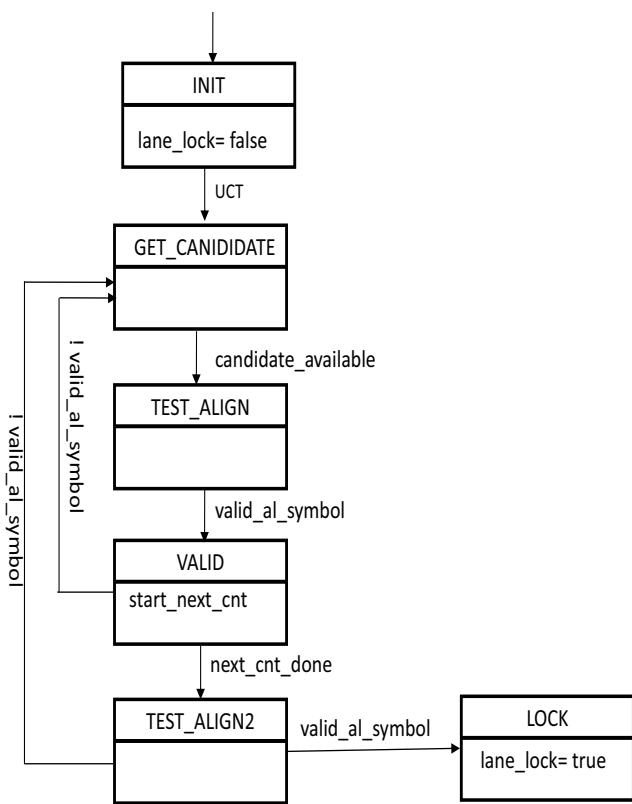


Figure 36 Lane Alignment Lock FSM

5.4.2.2.1.5 RS-FEC Lock STATE MACHINE

Variables:

- **all_locked** - Boolean variable set to true when all the lanes of the link are locked, as indicated by the lane_lock variable. all_locked is set to false when lane_lock is false for any lane of the link.
- **bad_cnt** - Counter that counts the number of uncorrectable FEC codewords within the measured FEC codewords window.
- **codeword_cnt** - Counter that counts the number of FEC codewords within the measured FEC codewords window.

INIT state:

This state is entered when alignment is enabled by the linkphy state machine or when all_locked is false. Set fec_fail to false to enable the alignment state machine to align the lanes.

Next state is DESKEW state if all_locked is true.

DESKEW state:

In this state, the FEC codeword lock state machine shall de-skew the lanes and
clears the codeword_cnt counter and the bad_cnt counter.

Next state is TEST_FEC unconditionally.

TEST_FEC state:

In this state, the FEC lock state machine checks the FEC codeword. If the FEC code-
word is uncorrectable, the FEC codeword lock state machine **shall** set the
valid_fec_codeword to false. Otherwise, it **shall** set it to true.

Next state is CODEWORD_CNT if valid_fec_codeword is true.

Next state is INVALID if valid_fec_codeword is false.

CODEWORD_CNT state:

In this state, the FEC codeword lock state machine increments the FEC codeword
window's codeword_cnt.

Next state is TEST_FEC if codeword_cnt < 256.

Next state is RESET_WINDOW if codeword_cnt = 256.

RESET_WINDOW state:

In this state, the FEC codeword lock state machine resets the codeword_cnt counter
and the bad_cnt counter.

Next state is TEST_FEC unconditionally.

INVALID state:

In this state, the FEC codeword lock state machine shall increment the bad_cnt
counter.

Next state is CODEWORD_CNT if bad_cnt < 2.

Next state is RESET_ALIGN if bad_cnt = 2.

RESET_ALIGN state:

In this state, the FEC codeword lock state machine shall set the fec_fail to true. By
setting the fec_fail to true, the alignment state machine will move to INIT state, setting
the lane_lock to false for all lanes and restarting the FEC codeword lock state ma-
chine.

Next state is INIT.

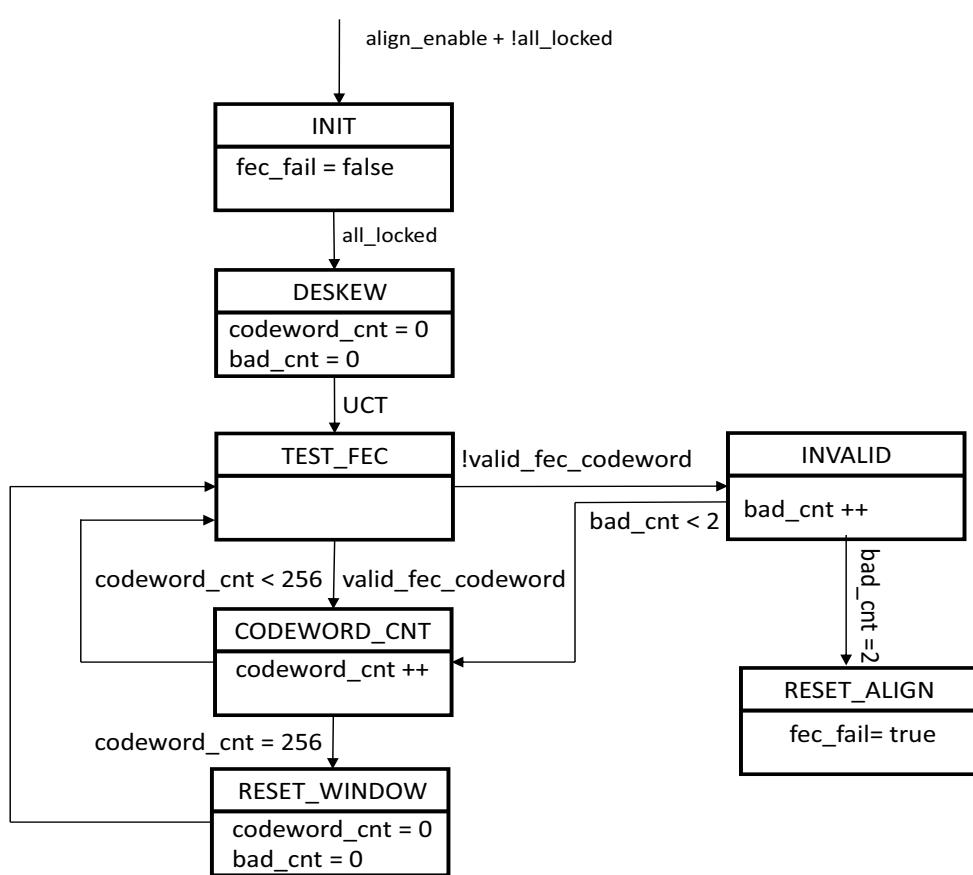


Figure 37 RS-FEC Codeword Lock FSM

5.4.2.2.2 RS-FEC DECODER FUNCTION

The Reed-Solomon decoder extracts the message symbols from the codeword, corrects them as necessary, and discards the parity symbols. The message symbols correspond to 20 transcoding blocks when using RS(528,514) code and to 10 transcoding blocks when using RS(271, 257) code.

The RS-FEC decoder function **shall** be capable of correcting any combination of up to $t=7$ symbol errors in a codeword. The RS-FEC decoder function **shall** also be capable of indicating when an erroneous codeword was not corrected. The probability that the decoder fails to indicate a codeword with $t+1$ errors as uncorrected is not expected to exceed 10^{-6} . This limit is also expected to apply for $t+2$ errors, $t+3$ errors, and so on.

The RS-FEC decoder may optionally nullify an uncorrected FEC codeword or not pass it to higher layers to reduce the Mean Time To False Packet Acceptance (MTTFPA). The means for the RS-FEC decoder to achieve this is implementation specific and is outside the scope of this specification.

The RS-FEC decoder function **shall** also count the number of corrected codewords, uncorrected codewords, symbols corrected per lane, and total corrected symbols across all lanes. The RS-FEC function increments one of the two per-port counters - **PortFECCorrectableBlockCounter** and **PortFECUncorrectableBlockCounter** - when an error is detected on a received FEC block. The RS-FEC function increments two counters - **FEC-CorrectedSymbolCounterLane< n >** and **PortFECCorrectedSymbolCounter** - when a symbol is corrected. These counters **shall** be 32 bits long.

The RS-FEC decoder function **shall** only increment the FEC symbol error counters for correctable codewords. The FEC receive function will increment lane i FEC symbol error counter for an N lane link width if the correctable symbol modulo N equals i.

The RS-FEC decoder may have an option to perform error detection without error correction in order to reduce the delay that is introduced by the RS-FEC error correction. When RS-FEC is enabled with or without error correction, the port is required to verify the FEC block lock and count the RS-FEC layer counters. When the error correction is bypassed, the FECPortCorrectedSymbolCounter shall increment when an erroneous symbol is detected and the FECLaneCorrectedSymbolCounter shall increment when an erroneous symbol is detected on the corresponding lane.

5.4.2.2.3 ALIGNMENT REMOVAL FUNCTION

When alignment is enabled, as indicated by the align_enable variable, the Alignment Removal function **shall** remove the first transcoding block of every 16th FEC codeword for x1 and x4 link width, the first two transcoding blocks of every 32nd FEC codeword for x8 link width, and the first three transcoding blocks of every 48th FEC codeword for x12 link width.

The Alignment Removal shall pass all the other transcoding blocks to the inverse transcoding function.

5.4.2.2.4 INVERSE TRANSCODING FUNCTION

The Inverse Transcoding function **shall** receive a 257-bit transcoding block and create four 66-bit blocks.

The Inverse Transcoding function **shall** first de-scramble the first five bits of the transcoding block.

For data only transcoding blocks (bit 0 = 1), the Inverse Transcoding function **shall** add a 10 sync header to each 64 bits.

For control transcoding blocks (bit 0 = 0), the Inverse Transcoding function **shall** add a 01 sync header for each data block and 10 sync header for each control block, as indicated by bits 4:1. The Inverse Transcoding function shall reconstruct the four bits removed by the transcoding function from the first control block using a lookup on the remaining four bits of that block.

Let rx_scrambled<256:0> be the 257 transcoded block received from the Alignment Removal function (bit 0 being the first transcoding bit and bit 256 the last transcoding bit).
1
2

Let rx_block[i] for i=0-to-3 be the output of the Inverse Transcoding function.
3
4

The Inverse Transcoding function **shall** de-scramble the first five bits as follows:
5
6

Set bits rx_transcoded<4:0> to the bit-wise exclusive OR (XOR) of rx_scrambled<4:0> with rx_scrambled<12:8>.
7
8

Set rx_transcoded<256:5> = rx_scrambled<256:5>.
9
10

For rx_transcoded<0> = 1 (data transcoding block), the Inverse Transcoding function **shall**:
11
12

- 1) Set rx_block[i]<65:2> = rx_transcoded<(64i+64):(i+1)> for i=0-to-3.
14
- 2) Set rx_block[i]<1:0> = 10 for i = 0-to-3.
15

For rx_transcoded<0> = 0 (control transcoding block), the Inverse Transcoding function **shall**:
16
17

- 1) Let c be the smallest i=0-to-3 for which rx_transcoded<1+i> = 0. That is the first control block in the transcoding block.
19
20
- 2) For i= 0-to-3: rx_block[i]<0> = rx_transcoded<i+1>. rx_block[i]<1> = ~rx_transcoded<i+1> (the inverse value of de_scrambled<j+1>).
21
22
- 3) For c > i >= 0: rx_block[i]<65:2> = rx_transcoded<(64i+68):(i+5)>.
23
24
- 4) For c < i <= 3: rx_block[i]<65:2> = rx_transcoded<(64i+64):(i+1)>.
25
26
- 5) rx_block[c]<65:10> = rx_transcoded<(64c+64):(64c+9)> rx_block[c]<5:2> = rx_transcoded<(64c+8):(64c+5)>.
27
28
- 6) Let de_scrambled_nibble<3:0> be the de-scrambled value of rx_block[c]<5:2>. Note that when c=0, c-1 denotes to the rx_block[3] of the previously inverse transcoded block: de_scrambled_nibble<j> = rx_block[c]<(j+2)>^rx_block[c-1]<j+10>^rx_block[c-1]<j+29> for j=0 to j=3.
29
30
31
32
- 7) The control block c type can be identified uniquely using the 4 bits de_scrambled_nibble<3:0> using the control block type in [Table 27 on page 159](#). Based on that values, the Inverse Transcoding function can reconstruct the LSB nibble of the block type before scrambling. Let lsb_before_scrambling<3:0> be the value reconstructed based on the aforementioned table. If no match is found in the table, the Inverse Transcoding function will set the rx_block[c]<1:0> = 11 - invalid block type.
33
34
35
36
37
- 8) Set rx_block[c]<6+j> = lsb_before_scrambling<j>^rx_block[c-1]<j+14>^rx_block[c-1]<j+33> for j=0 to j=3.
38
39
40

For any transcoding block other than the alignment transcoding block, if the block type is control (rx_transcoded<0>=0) and all four blocks are marked as data blocks (rx-
41
42

_transcoded<4:1> = 1111), the inverse transcoding block shall discard the transcoding block and shall not pass it to the Rx Block stream.

5.5 CONTROL SYMBOLS/BLOCKS AND ORDERED-SETS

The InfiniBand link uses the Control Symbols/Blocks and Ordered-Sets of control and data symbols to implement:

- 1) packet delimiters,
- 2) ordered-sets,
- 3) packet padding, and
- 4) clock tolerance compensation.

C5-6: This compliance statement is obsolete and has been replaced by [C5-6.2.1](#).

C5-6.2.1: All ports shall use the control symbols/Blocks and the SKIP, TS1, and TS2 ordered-sets specified in [Section 5.5, “Control Symbols/Blocks and Ordered-Sets,” on page 139](#) for Link/Phy control and communication.

o5-6.2.1: All ports claiming compliance with InfiniBand Rel. 1.2 shall also use the TS3, HRTBT, and TS-T ordered-sets specified in [Section 5.5. “Control Symbols/Blocks and Ordered-Sets,” on page 139](#) for Link/Phy control and communication.

o5-6.2.2: All ports claiming compliance with InfiniBand Rel. 1.3 shall also support the Rev. 1 TS3 as specified in [Section 5.5.2.4, “Training Sequence Three Ordered-Set \(TS3\),” on page 143](#)

5.5.1 CONTROL SYMBOLS/BLOCKS

5.5.1.1 CONTROL SYMBOLS

For 8b/10b encoding the IB control symbols have been chosen from the 8b/10b special code-groups and shall be used as defined in [Table 23](#) below. For 64b/66b encoding, the InfiniBand™ control symbols are a one byte symbol, and shall be used as indicated in [Table 23](#). The control code-groups are non-data symbols that are uniquely identifiable.

Of the twelve available special code-groups, seven are used as control symbols, one is provided for vendor-specific use, and four are reserved.

Table 23 Link Control Symbols

Symbol	8b/10b Encoding	64b/66b Byte ^a	Description
COM	K28.5	NA	Comma, character boundary alignment symbol.
SDP	K27.7	78	Start of Data Packet Delimiter
SLP	K28.2	55	Start of Link Packet Delimiter
EGP	K29.7	B4, FF	End of Good Packet Delimiter
EBP	K30.7	AA, E1	End of Bad Packet Delimiter
PAD	K23.7	00	Packet padding symbol
SKP	K28.0	4B	Skip symbol
	K28.1 K28.7	NA	Reserved control symbols. These symbols have “comma” characteristics.
	K28.3 K28.4	NA	Reserved control symbols.
	K28.6	4B	Vendor-specific control symbol.

a. Byte codes are in hexadecimal.

5.5.1.1.1 COMMA CONTROL SYMBOL (COM)

The comma control symbol (K28.5) is used by the physical lane receiver logic to identify symbol boundaries. Comma symbols are required to synchronize the receive logic when the links are being trained. The comma symbol is also used as the start of ordered-set delimiter.

5.5.1.1.2 START OF DATA PACKET DELIMITER (SDP)

The Start of Data Packet Delimiter symbol (K27.7 for 8b/10b encoding and 78h for 64b/66b encoding) is transmitted to identify the start of a data packet. Packet formatting rules specify which physical lanes may be used by the “SDP” control symbol. (See [Section 5.7 on page 175](#))

5.5.1.1.3 START OF LINK PACKET DELIMITER (SLP)

The Start of Link Packet Delimiter symbol (K28.2 for 8b/10b encoding and 55h for 64b/66b encoding) is transmitted to identify the start of a link control packet. Packet formatting rules specify which physical lanes may be used by the “SLP” control symbol. (See [Section 5.7 on page 175](#))

5.5.1.1.4 END OF GOOD PACKET DELIMITER (EGP)

The End of Good Packet Delimiter symbol (K29.7 for 8b/10b encoding and either B4h or FFh for 64b/66b encoding) is used to mark the end of each packet as it is transmitted by the originating port. Packet length rules restrict which physical lanes may be used to transmit the “EGP” control symbol. (See [Section 5.7 on page 175](#))

5.5.1.1.5 END OF BAD PACKET DELIMITER (EBP)

The End of Bad Packet Delimiter symbol (K30.7 for 8b/10b encoding and either AAh or E1h for 64b/66b encoding) is used to mark the end of a bad packet forwarded by a switch or router node. When an error (e.g.: decode error, CRC error, etc.) is detected in a data packet it is marked bad by replacing the original “EGP” symbol with a “EBP” packet symbol. Receiving end nodes are required to recognize either EGP or EBP as the end of packet delimiter. Any data packet terminated with “EBP” symbol must be treated as if it had a CRC error.

5.5.1.1.6 PADDING SYMBOL (PAD)

For 8b/10b encoding the Padding symbol (K23.7) is used on the 8x and 12x physical link to align the physical lanes. Alignment is required at the end of any packet that does not end (EGP or EBP) in physical lane eleven (11) on a 12x physical link, or physical lane seven (7) on an 8x link. (See [Section 5.7.5 on page 182](#) and [Section 5.7.6 on page 184](#)) Pad symbols are also used by the retiming repeaters to forward error conditions (See [Section 5.12 on page 241](#)). For 64b/66b encoding the PAD symbol (00h) is used to PAD a EGP3 or EBP3 block to align to the eight (8) byte block. (See [Section 5.5.3.3 on page 160](#) and [Section 5.5.3.4 on page 161](#))

5.5.1.1.7 SKIP SYMBOL (SKP)

For 8b/10b encoding the “Skip” symbol (K28.0) is used as part of the SKIP ordered-set which is used for clock tolerance compensation. (See [Section 5.11 on page 239](#))

5.5.1.2 RESERVED CONTROL SYMBOLS

Four of the special code-groups listed in [Table 23](#) are reserved for future use by the IB standard. Special symbols K28.1, K28.3, K28.4, and K28.7 are reserved. For this version of the IB standard, the following rules apply to reserved control symbols:

- 1) The use of these control symbols may be defined in a future revision of this specification.
- 2) Devices based on this version of the specification shall not transmit these control symbols.
- 3) When a device based on this version of the specification receives a reserved control symbol a symbol coding error shall be reported. (See [Section 5.6.2 on page 169](#))

5.5.1.2.1 VENDOR-SPECIFIC CONTROL SYMBOL

The IB standard reserves a special 8b/10b control symbol (K28.6) for vendor-specific use. The function and use of this symbol is vendor-defined, and interoperability between vendors is not guaranteed. The negotiation process for the use of this symbol is not defined by this specification.

The following rules apply to the use of the vendor-specific control symbols:

- 1) The use of the vendor-specific control symbols may be defined for functions that extend or enhance the standard architected functionality.

- 2) Devices supporting the vendor-specific control symbol shall not transmit the symbol until both ends have negotiated its use.
- 3) When an unsupported or un-negotiated vendor-specific control symbol is detected, a coding error shall be reported. (see [Section 5.6.2 on page 169](#))

5.5.2 CONTROL ORDERED-SETS 8B/10B ENCODING

In addition to the individual control symbols described above, IB defines control ordered-sets. The ordered-sets are used for link training and clock tolerance compensation. The first symbol of all ordered-sets shall be the “COM” symbol, with additional symbols unique to the set type. When ordered-sets are transmitted, the ordered-set shall be transmitted on all physical lanes.

The defined ordered-sets are illustrated in [Figure 38](#).

Ordered Set Formats - per Lane						
Byte	SKIP	TS1	TS2	TS3	HRTBT	TS-T
0	COM	COM	COM	COM	COM	COM
1	SKP	LANE ID	LANE ID	LANE ID	LANE ID	Reserved
2	SKP	D10.2 (4Ah)	D5.2 (45h)	D13.2 (4Dh)	D1.2 (41h)	D17.2 (51h)
3	SKP	D10.2 (4Ah)	D5.2 (45h)	D13.2 (4Dh)	D1.2 (41h)	D17.2 (51h)
4		D10.2 (4Ah)	D5.2 (45h)	D13.2 (4Dh)	D1.2 (41h)	D17.2 (51h)
5		D10.2 (4Ah)	D5.2 (45h)	D13.2 (4Dh)	OpCode	D17.2 (51h)
6		D10.2 (4Ah)	D5.2 (45h)	D13.2 (4Dh)	Reserved	D17.2 (51h)
7		D10.2 (4Ah)	D5.2 (45h)	D13.2 (4Dh)	PortNum	D17.2 (51h)
8		D10.2 (4Ah)	D5.2 (45h)	SpeedActive	GUID[63-56]	Speeds
9		D10.2 (4Ah)	D5.2 (45h)	HBR/ADD	GUID[55-48]	Test Opcode
10		D10.2 (4Ah)	D5.2 (45h)	DDSV/DDS	GUID[47-40]	Reserved
11		D10.2 (4Ah)	D5.2 (45h)	Reserved	GUID[39-32]	Reserved
12		D10.2 (4Ah)	D5.2 (45h)	Reserved	GUID[31-24]	TxCfg[15-8]
13		D10.2 (4Ah)	D5.2 (45h)	Reserved	GUID[23-16]	TxCfg[76-0]
14		D10.2 (4Ah)	D5.2 (45h)	Reserved	GUID[15- 8]	RxCfg[15-8]
15		D10.2 (4Ah)	D5.2 (45h)	Reserved	GUID[7- 0]	RxCfg[76-0]

Figure 38 Ordered-Sets

5.5.2.1 SKIP ORDERED-SET (SKIP)

When transmitted the Skip sequence (SKIP) is a four symbol ordered-set comprised of a comma (COM) and three consecutive “Skip” (SKP) symbols. A SKIP ordered-set may be inserted or repeated by a retiming repeater in the link. (See [Section 5.12 on page 241](#)).

5.5.2.2 TRAINING SEQUENCE ONE ORDERED-SET (TS1)

Link Training Sequence One (TS1) is a sixteen symbol ordered-set composed of a comma (COM), a lane identifier data symbol, and fourteen data symbols unique to training sequence one. The lane identifiers used for TS1 and TS2 shall use the definitions found in [Table 24](#) below.

Table 24 Lane Identifiers

Lane Identifier	Hex number	8b/10b Encoding	Description
0	00	D00.0	Physical lane 0 used by 1x, 4x 8x, and 12x links
1	01	D01.0	Physical lane 1 used by 4x 8x, and 12x links
2	02	D02.0	Physical lane 2 used by 4x 8x, and 12x links
3	04	D04.0	Physical lane 3 used by 4x 8x, and 12x links
4	08	D08.0	Physical lane 4 used by 8x and 12x links
5	0F	D15.0	Physical lane 5 used by 8x and 12x links
6	10	D16.0	Physical lane 6 used by 8x and 12x links
7	17	D23.0	Physical lane 7 used by 8x and 12x links
8	18	D24.0	Physical lane 8 used by 12x links
9	1B	D27.0	Physical lane 9 used by 12x links
10	1D	D29.0	Physical lane 10 used by 12x links
11	1E	D30.0	Physical lane 11 used by 12x links

The TS1 unique data symbol is D10.2 (or 4Ah), and the 10-bit encoded value is a toggling pattern (010101 0101) for both the positive and negative running disparity.

5.5.2.3 TRAINING SEQUENCE TWO ORDERED-SET (TS2)

Link Training Sequence Two (TS2) is a sixteen symbol ordered-set composed of a comma (COM), a lane identifier data symbol, and fourteen data symbols unique to training sequence two. The lane Identifiers used by TS2 are the same as for TS1 and are defined in [Table 24](#) above.

The TS2 unique data symbol is D5.2 (or 45h), and the 10-bit encoded value is the same pattern (101001 0101) for both the positive and negative running disparity.

5.5.2.4 TRAINING SEQUENCE THREE ORDERED-SET (TS3)

Link Training Sequence Three (TS3) is a sixteen symbol ordered-set composed of a comma (COM), a lane identifier data symbol, six data symbols unique to training sequence three, TS3 revision, a bit map of the active speeds, a bit map requesting transmitter de-emphasis and/or link heartbeat enabling, request extended speed change time and requested test time, and a byte describing the transmitter de-emphasis/preset or coefficients setting which should be used, FEC request bit, select of test pattern, fine tuning ability and max packet rate disable ability followed by four reserved bytes. The lane Identifiers used by TS3 are defined in [Table 24](#). The format of the TS3 ordered-set is shown in [Figure 39 on page 144](#) and [Figure 40 on page 145](#) below. The use of the TS3 ordered-set is described in [Section 5.8.4.6, “Configuration States - Enhanced Signaling,” on](#)

[page 203.](#)

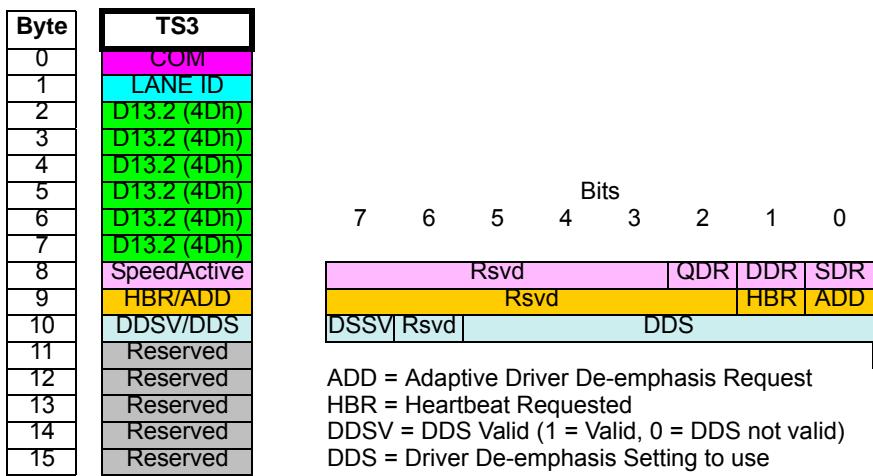
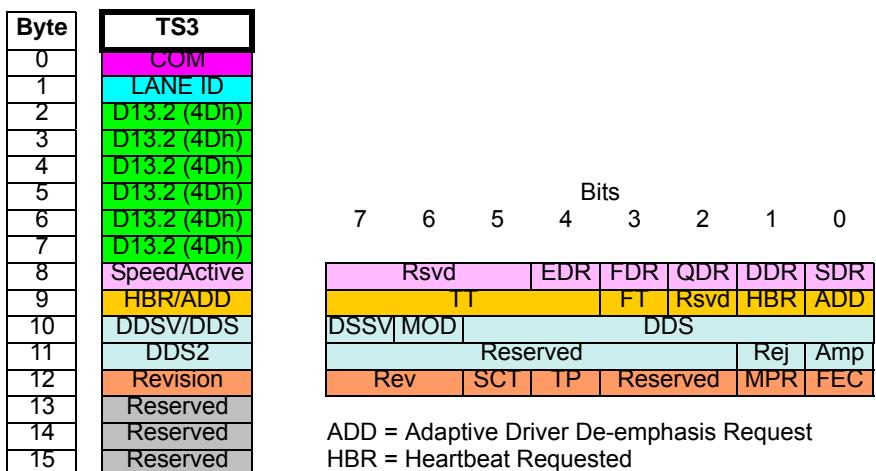
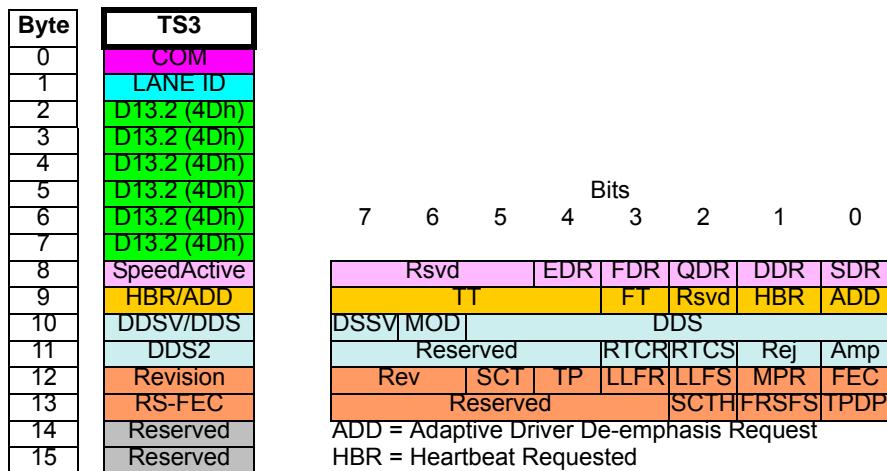


Figure 39 TS3 Ordered-Set: Detailed Format - Rev 0



ADD = Adaptive Driver De-emphasis Request
 HBR = Heartbeat Requested
 FT = Fine Tuning capability
 TT = requested Test Time
 DDSV = DDS Valid (1 = Valid, 0 = DDS not valid)
 DDS = Driver De-emphasis Setting to use
 MOD = Mode (0 Preset mode, 1 Coefficient mode)
 Rej - Reject DDS request
 Amp - Amplitude Bit
 FEC = FEC request
 MPR = Max Packet Rate disable capability
 TP = Test pattern (0 - PRBS23, 1- PRBS11)
 SCT - Extended Speed Change Time request
 Rev = TS3 revision (= 01, for Rev. 1)

Figure 40 TS3 Ordered-Set: Detailed Format - Rev 1



ADD = Adaptive Driver De-emphasis Request

HBR = Heartbeat Requested

FT = Fine Tuning capability

TT = requested Test Time

DDS = Driver De-emphasis Setting to use

MOD = Mode (0 Preset mode, 1 Coefficient mode)

DDSV = DDS Valid (1 = Valid, 0 = DDS not valid)

Amp - Amplitude Bit

Rej - Reject DDS request

RTCS = Remote Transmit CDR enable Support

RTCR = Remote Transmit CDR enable Request

Amp - Amplitude Bit

FEC = FEC request

MPR = Max Packet Rate disable capability

LLFS = Low Latency FEC Support

LLFR = Low Latency FEC Request

TP = Test pattern (0 - PRBS23, 1- PRBS11)

SCT - Extended Speed Change Time request

Rev = TS3 revision (= 10, for Rev. 2)

FRSFS = FDR RS-FEC Support

TPDP = Test Pattern with Different Polynomials

SCTH = SCT, High bit of {SCTH,SCT} 2-bit request

Figure 41 TS3 Ordered-Set: Detailed Format - Rev 2

5.5.2.4.1 REV 0 TS3

Symbol 8, the *SpeedActive* link speed identifier, is used to identify and advertise the speeds at which the Link/Phy is enabled to operate, which the link medium supports. The enumerated values used for advertising enabled link speeds are the same as described for the **SM.PortInfo(LinkSpeedEnabled)** field.

Symbol 9 indicates whether the sender is requesting adaptive driver de-emphasis (ADD), and whether heartbeat is enabled and requested (HBR).

Symbol 10 indicates whether the transmitter of the TS3 is requesting a driver de-emphasis setting different from the default setting on its peer transmitter, and a 4-bit identifier for the driver de-emphasis setting requested. The transmitter driver may therefore have 17 de-emphasis settings - a default de-emphasis setting (DDSV = 0), and up to 16 adaptive driver de-emphasis settings (DDSV=1, DDS specifies a driver de-emphasis setting number). Driver de-emphasis is described in [Section 6.5.2, “Equalization for InfiniBand Release 1.2.1 Devices.” on page 280](#) and [Section 6.5.3, “Equalization for InfiniBand Release 1.3 Devices.” on page 282](#).

Symbols 11 to 15 are Reserved. They are transmitted as D00.0, and valid Dxx.y symbols are ignored at the receiver.

The TS3 unique data symbol is D13.2 (or 4Dh), and the 10-bit encoded value is the same pattern (101100 0101) for both the positive and negative running disparity.

5.5.2.4.2 REV 1 TS3

Symbol 8, the *SpeedActive* link speed identifier, is used to identify and advertise the speeds at which the Link/Phy is enabled to operate, which the link medium supports. The enumerated values used for advertising enabled link speeds are the same as described for the **SM.PortInfo(LinkSpeedEnabled)** and **SM.PortInfo(LinkSpeedExtEnabled)** fields.

Symbol 9 indicates whether the sender is requesting adaptive driver de-emphasis (ADD), as described in [Table 25 on page 148](#), and whether heartbeat is enabled and requested (HBR). Symbol 9 is also used for the sender to indicate support of Fine Tuning and Test Time. Setting the FT bit to 1 indicates that the device is capable of the optional Fine Tuning algorithm, as described in [Section 5.13, “Fine Tuning.” on page 246](#). Fine Tuning shall only be enabled if both link partners set this bit. The 4 bits of TT are used to request a Test Time for equalization training, as defined in [Table 41 on page 191](#). The larger of the two Test Time values requested is used by both ports.

All lanes of a port must transmit the same FT, HBR, ADD, and TT values.

Symbol 10 indicates whether the sender of the TS3 is requesting a driver de-emphasis setting different from the current setting on its peer transmitter, and a 4-bit identifier indicating the new preset request or a new coefficient change request. The transmitter driver may therefore have 17 de-emphasis settings - a default de-emphasis setting (DDSV = 0), and up to 16 adaptive driver de-emphasis settings (DDSV=1, DDS specifies a driver de-emphasis setting number). Driver de-emphasis is described in [Section 6.5.3, “Equalization for InfiniBand Release 1.3 Devices.” on page 282](#) and [Section 6.6.5, “Host Driver Output Characteristics for FDR.” on page 293](#).

For Rev 1 TS3, the DDS is a 6 bit field. Bit 6 is the DDS mode (MOD) bit, and is used to choose if the requested DDS is in the de-emphasis/preset format or coefficient change format.

When MOD bit is cleared to 0, the DDS field is used to request a preset.

When MOD bit is set to 1, the DDS field is used to request an incremental change to the coefficient. In this incremental mode, the DDS bits take the following encoding:

bits 0-1: C₋₁ pre-cursor tap

bits 2-3: C₀ cursor tap

bits 4-5: C₊₁ post-cursor tap

The 2 bits use the following encoding:

00 - no change

01 - increase

10 - decrease

11 - reserved

Each lane may transmit a different value for the DDS, DDSV and MOD fields.

Table 25 DDSV, ADD and MOD Encoding

ADD	DDSV	MOD	
0	0	0	No Config.Test training request, no equalization request - the peer shall use its current transmitter setting in the Config.Test training
0	1	0	No Config.Test training request, preset request - the peer will use the transmitter preset requested in the DS field and proceed to linkup
1	0	0	Config.Test training request, no equalization request - the peer shall use the current transmitter setting in the Config.Test training
1	1	0	Config.Test training request, preset request - the peer shall use the requested transmitter preset based on the DDS value
0	0	1	No Config.Test training request, no equalization request - the peer shall use its current transmitter setting for linkup
0	1	1	No Config.Test training request, coefficient request - the peer shall use the transmitter coefficient change requested in the DDS field and proceed to linkup
1	0	1	Config.Test training request, no equalization request - the peer shall use the current transmitter setting in the Config.Test training
1	1	1	Config.Test training request, coefficient request - the peer shall use the requested transmitter coefficient change based on the DDS value

Implementation Note

With a multi lane port some lanes may complete the equalization training process before other lanes. When a single lane requires equalization training the port will request equalization training by setting the ADD bit in byte 9 on all lanes of the port.

Lanes that have completed the equalization training shall clear the DDSV bit in byte 10 to indicate than no change to the peer transmitter setting is required.

Symbol 11:

Bit 0 is the amplitude bit. When set to 1, the preset request in the Rev 1 TS3 is for the higher amplitude range, as defined in [Table 54 on page 299](#). When cleared to 0, the preset request in the Rev1 TS3 is for the lower amplitude range as defined in [Table 54 on page 299](#). This bit is only applicable for a preset request, i.e., when the MOD bit is cleared to 0 (Preset Mode). When the MOD bit is set to 1 (Coefficient Mode), the receiving port shall ignore the amplitude bit.

For preset 0 request, the amplitude bit shall be ignored by the receiving port.

Bit 1 is the Reject field - when setting this bit to 1 a port indicates that the previous transmitter setting was rejected by the port's transmitter. Each lane may transmit a different value for the Rej bit.

Bits 2 to 7 are Reserved.

Symbol 12:

Bit 0 is the Fire-Code Forward Error Correction (FEC) request - Fire-Code FEC shall be enabled if either link partner sets this bit.

Bit 1 - Max Packet Rate disable capability (MPR) - devices that are capable of receiving more than 1 packet every 64 Bytes - see [Section 5.15, "Max Packet Rate," on page 255](#) may set this bit. When this bit is set in the received TS3s the device may change the value of MPR_en to false.

Bits 2 to 3 are reserved.

Bit 4 - Test Pattern (TP) - PRBS11 request. PRBS11 test pattern shall only be enabled if both link partners set this bit. PRBS11 test pattern - see [Section 5.8.4.6.6, "Config.Test State - Release 1.2 Enhanced Signaling," on page 214](#). When this bit is set in both the received and transmitted TS3 the test pattern during Config.Test is PRBS11.

Bit 5 - Extended Speed Change Time (SCT). When set to 1, this bit indicates that the device is requesting to extend the speed change time to 16 ms when changing speed to FDR or EDR or when changing speed from FDR or EDR. When neither of the link partner requests extending the speed change time, a 4 ms value shall be used for

speed changes to or from FDR and EDR speeds. For all other speed changes a 2 ms value shall be used.

Bits 6 to 7 - Rev - TS3 revision

All lanes of a port must transmit the same value for the FEC, MPR, TP, SCT and Rev fields.

Symbols 13 to 15 are Reserved. They are transmitted as D00.0, and valid Dxx.y symbols are ignored at the receiver.

The TS3 unique data symbol is D13.2 (or 4Dh), and the 10-bit encoded value is the same pattern (101001 0101) for both the positive and negative running disparity.

Table 26 TS3 Rev1 and 2 Variables

Field	Byte in TS3	Bits	Per Lane	Description
Speed-Active	8	4:0	No	Port's Active Speed. MaxBothActive speed is achieved by performing an AND operation on the transmitted and received SpeedActive values. MaxSpeedAgreed is set to true when the maximum transmitted SpeedActive equals the maximum received SpeedActive. MaxSpeedAgreed is set to false when the maximum transmitted SpeedActive and the maximum received SpeedActive differ.
ADD	9	0	No	The ADD is used by a port to request equalization training through Config.Test. The port uses the OR of the transmitted and received ADD field values to identify if equalization training is to be performed.
HBR	9	1	No	Heartbeat request. The port performs an AND operation on the transmitted and received HBR values, and enables heartbeat if and only if both sides have requested
FT	9	3	No	Fine Tuning supported. The port performs an AND operation on the transmitted and received FT values, and may enable fine tuning if and only if both sides have reported support of fine tuning.
TT	9	7:4	No	Test Time request. The test time shall be the maximum of the test times requested by the two ports.
DDS	10	5:0	Yes	DDS field defines the transmitter setting requested by the lane - when MOD bit is clear the DDS field holds the preset index requested by the lane, when MOD bit is set the DDS field holds the coefficient change requested by the lane.
MOD	10	6	Yes	MOD bit defines if the transmitter setting requested by the lane is (0) a preset request or (1) a coefficient change request.
DDSV	10	7	Yes	DDS Valid - this bit when set indicated that the lane is requesting a new transmitter setting as defined by DDS and MOD.
AMP	11	0	Yes	Amplitude Bit - this bit is used by the port to signal to the peer port the amplitude range (high or low) for the preset requested in the DDS when MOD bit is cleared to 0 (Preset Mode).

Table 26 TS3 Rev1 and 2 Variables (Continued)

Field	Byte in TS3	Bits	Per Lane	Description
Rej	11	1	Yes	Reject transmitter request - a port that doesn't support a requested preset or cannot perform the requested coefficient change will set the Rej bit to indicate that the request was rejected Note: the port will report rejecting the coefficient request only after completing Config.Test and returning to Config.CfgEnhanced state.
RTCS	11	2	No	Remote Transmit CDR enable Support - when set indicates that a transmit CDR is present and supported for the maximum transmitted active speed and that remote transmit CDR control is supported.
RTCR	11	3	Yes	Remote Transmit CDR enable Request- when set indicates that remote transmit CDR enable is requested. A different RTCR value may be requested for different lanes. When the transmitted RTCS is set, the transmit CDR on the lane is enabled during MaxBoth-Active speed according to the received RTCR. When the transmitted RTCS is cleared, a transmit CDR may be enabled according to the local port policy.
FEC	12	0	No	Forward Error Correction request - The port performs an OR operation on the transmitted and received FEC values, and enables the FEC if either port requested FEC.
MPR	12	1	No	Max Packet Rate disable capability - this bit reports the receiver's capability to receive packets at a higher rate than defined in Max Packet Rate (Section 5.15) .A port may enable its transmitter to send packets at a higher rate if the received MPR bit is set.
LLFS	12	2	No	Low Latency FEC Supported - when set, indicates that the port supports RS(271,257) FEC. The port performs AND operation between the transmitted and received LLFS.
LLFR	12	3	No	Low Latency FEC Request - when set, indicates that the port requests to use RS(271,257) FEC instead of the RS(528,514) code. The port performs OR operation between the transmitted and received LLFR.
TP	12	4	No	Test Pattern - this field is the port's requested test pattern in Config.Test: 0 - PRBS23 1 - PRBS11 The port performs an AND operation on transmitted and received TP bits, and PRBS11 will be used if and only if both ports requested the PRBS11 pattern.
SCT	12	5	No	Extended Speed Change Time - a port may request extending the speed change time to and from rates higher than QDR by setting this bit. In Rev1, the ports perform OR on the transmitted and received SCT bit. If set by either port then the speed change time to and from rates higher than QDR is extended from 4 to 16 ms. In Rev2, the ports perform a MAX on the transmitted and received values of {SCTH,SCT}. The speed change time to and from rates higher than QDR is extended from 4 ms to either: 16 ms (if [SCTH,SCT]=01), or 32 ms ([SCTH,SCT]=10), or 64 ms ([SCTH,SCT]=11).
Rev	12	7:6	No	TS3 revision. BothRev2 is set to TRUE when both the received and transmitted TS3 revision is greater than or equal to 2.
TPDP	13	0	No	Test Pattern with Different Polynomials - when set, indicates that the port requests that different lanes will use different polynomials for the test pattern. The port performs AND operation between the transmitted and received TPDP, and different polynomials will be used if and only if both ports have requested it.
FRSFS	13	1	No	FDR RS-FEC support - when set, indicates that the port supports RS-FEC for FDR rate. The port performs AND operation between the transmitted and received FRSFS.
SCTH	13	2	No	See SCT discussion above.

5.5.2.4.3 Rev 2 TS3

Symbol 8, the *SpeedActive* link speed identifier, is used to identify and advertise the speeds at which the Link/Phy is enabled to operate, which the link medium supports. The enumerated values used for advertising enabled link speeds are the same as described for the ***SM.PortInfo(LinkSpeedEnabled)*** and ***SM.PortInfo(LinkSpeedExtEnabled)*** fields.

Symbol 9 indicates whether the sender is requesting adaptive driver de-emphasis (ADD), as described in [Table 25 on page 148](#), and whether heartbeat is enabled and requested (HBR). Symbol 9 is also used for the sender to indicate support of Fine Tuning and Test Time. Setting the FT bit to 1 indicates that the device is capable of the optional Fine Tuning algorithm, as described in [Section 5.13, “Fine Tuning,” on page 246](#). Fine Tuning shall only be enabled if both link partners set this bit. The four bits of TT are used to request a Test Time for equalization training, as defined in [Table 41 on page 191](#). The larger of the two Test Time values requested is used by both ports.

All lanes of a port must transmit the same FT, HBR, ADD, and TT values.

Symbol 10 indicates whether the sender of the TS3 is requesting a driver de-emphasis setting different from the current setting on its peer transmitter, and a 4-bit identifier indicating the new preset request or a new coefficient change request. The transmitter driver may therefore have 17 de-emphasis settings: a default de-emphasis setting (DDSV = 0), and up to 16 adaptive driver de-emphasis settings (DDSV=1, DDS specifies a driver de-emphasis setting number). Driver de-emphasis is described in [Section 6.5.3, “Equalization for InfiniBand Release 1.3 Devices,” on page 282](#) and [Section 6.6.5, “Host Driver Output Characteristics for FDR,” on page 293](#).

For Rev 1 TS3, the DDS is a 6-bit field. Bit 6 is the DDS mode (MOD) bit, and is used to choose whether the requested DDS is in the de-emphasis/preset format or coefficient change format.

When the MOD bit is cleared, the DDS field is used to request a preset.

When the MOD bit is set to 1, the DDS field is used to request an incremental change to the coefficient. In this incremental mode, the DDS bits take the following encoding:

bits 0-1: C₋₁ pre-cursor tap

bits 2-3: C₀ cursor tap

bits 4-5: C₊₁ post-cursor tap

The two bits use the following encoding:

00 - no change

01 - increase

10 - decrease

11 - reserved

Each lane may transmit a different value for the DDS, DDSV, and MOD fields.

Symbol 11

Bit 0 is the amplitude bit. When set to 1, the preset request in the Rev 1 TS3 is for the higher amplitude range, as defined in [Table 54 on page 299](#). When cleared to 0, the preset request in the Rev1 TS3 is for the lower amplitude range as defined in [Table 54 on page 299](#). This bit is only applicable for a preset request, i.e., when the MOD bit is cleared to 0 (Preset Mode). When the MOD bit is set to 1 (Coefficient Mode), the receiving port shall ignore the amplitude bit.

For preset 0 request, the amplitude bit shall be ignored by the receiving port.

Bit 1 is the Reject field - when setting this bit to 1 a port indicates that the previous transmitter setting was rejected by the port's transmitter. Each lane may transmit a different value for the Rej bit.

Bits 2, 3 are used for the remote transmit CDR enable (RTCS, RTCR). The remote transmit CDR control enables a port to request its peer port to enable or disable the transmit CDR on the attached media. This enables to reduce the system power consumption when the link can operate without a transmit CDR.

A port that supports remote transmit CDR control and is attached to a media with a transmit CDR that supports the maximum active speed, the port shall set the RTCS bit. When the port sets the RTCS, the transmit CDR on a lane is enabled when operating in the MaxBothActive speed according to the received RTCR. When the transmitted RTCS is cleared, a transmit CDR may be enabled according to the local port policy.

The transmit CDR enable when operating in SDR is outside the scope of the spec.

Bits 4 to 7 are Reserved.

Symbol 12

Bit 0 is the Reed-Solomon Forward Error Correction (RS-FEC) request when using Rev 2 TS3. RS-FEC shall be enabled if either link partner sets this bit. The RS-FEC code is decided based on bits 2 and 3 of this byte.

Bit 1 - Max Packet Rate disable capability (MPR) - devices that are capable of receiving more than 1 packet every 64 Bytes may set this bit - see [Section 5.15, "Max Packet Rate," on page 255](#). When this bit is set in the received TS3s, the device may change the value of MPR_en to false.

Bit 2 - LLFS - Low Latency FEC Supported. This bit indicates support of the lower latency RS-FEC RS(271,257) code. The lower latency FEC can only be enabled if both ports of the link support it.

Bit 3 - LLFR - Low Latency FEC Request. This bit indicates that the port requests that when using RS-FEC (based on the value of bit 0 Byte 12), the code to be used should be RS(271,257). If both ports support that code as indicated by the LLFS bit, and at least one port sets the LLFR, then the code to be used if FEC is enabled as decided by bit 0 (FEC) of Byte 12 shall be RS(271,257). Otherwise, it will be RS(528,514).

Bit 4 - Test Pattern (TP) - PRBS11 request. The PRBS11 test pattern shall only be enabled if both link partners set this bit. PRBS11 test pattern - see [Section 5.8.4.6.6, "Config.Test State - Release 1.2 Enhanced Signaling," on page 214](#). When this bit is set in both the received and transmitted TS3, the test pattern during Config.Test is PRBS11.

Bit 5 - Extended Speed Change Time (SCT). In Rev 2, the Extended Speed Change Time request is denoted as a two-bit field, comprising SCTH (Symbol 13, Bit 2) and SCT. When [SCTH,SCT] is set to a value other than 00, they indicate that the device is requesting to extend the speed change time to longer than 4 ms when switching to or from speeds faster than QDR. The ports perform a MAX operation of the transmitted and received values of [SCTH, SCT], and both ports use the maximum value 16 ms when changing speed to FDR or EDR, or when changing speed from FDR or EDR. When neither of the link partners requests extending the speed change time, a 4 ms value shall be used for speed changes to or from FDR and EDR speeds. For all other speed changes (i.e., between speeds QDR and slower), a 2 ms value shall be used. The requested values for [SCHT,SCT] correspond to: [0,0]: 4 ms, [0,1]: 16 ms, [1,0]: 32 ms, [1,1]: 64 ms.

Bits 6-7 - Rev - TS3 revision.

All lanes of a port must transmit the same value for the FEC, MPR, LLFS, LLFR, TP, SCT and Rev fields.

Symbol 13

Bit 0 - Test pattern with Different Polynomials (TPDP). When set to 1, indicates that the port requests that different lanes will use different polynomials for the test pattern (see [Table 43, "PRBS11 polynomials when Different Polynomials are used for Each Lane," on page 219](#) and [Table 44, "PRBS23 Polynomials when Different Polynomials are used for Each Lane," on page 219](#)). The different polynomials mode is enabled if only and only if both link partners set this bit.

Bit 1 - FDR RS-FEC support (FRSFS) - when set, indicates that the port supports RS-FEC for FDR rate. The port performs AND operation between the transmitted and received FRSFS.

Bit 2 - Extended Speed Change Time, High bit (SCTH). See discussion of SCT (Symbol 12, Bit 5) above.

Bits 3 to 7 are Reserved.

All lanes of a port must transmit the same value for the TPDP, FRSFS, and SCTH fields.

Symbols 14 to 15 are Reserved. They are transmitted as D00.0, and valid Dxx.y symbols are ignored at the receiver.

The TS3 unique data symbol is D13.2 (or 4Dh), and the 10-bit encoded value is the same pattern (101001 0101) for both the positive and negative running disparity.

In order to enable different policies per speed, the FEC request bit shall be updated after Config.test according to the last tested speed.

5.5.2.5 LINK HEARTBEAT ORDERED-SET (HRTBT)

Link Heartbeat ordered-set (HRTBT) is a sixteen symbol ordered-set composed of a comma (COM), a lane identifier data symbol, three data symbols unique to the Link Heartbeat ordered-set, a 1-symbol OpCode, a 1-symbol Reserved field, a 1-symbol PortNum value (only used for switch ports), and an eight-symbol Globally Unique ID (GUID). The format of the HRTBT ordered-set is shown in [Figure 42 on page 155](#) below. The use of the HRTBT ordered-set is described in [Section 5.14.1, “Operation of Link Heartbeats.” on page 253](#)

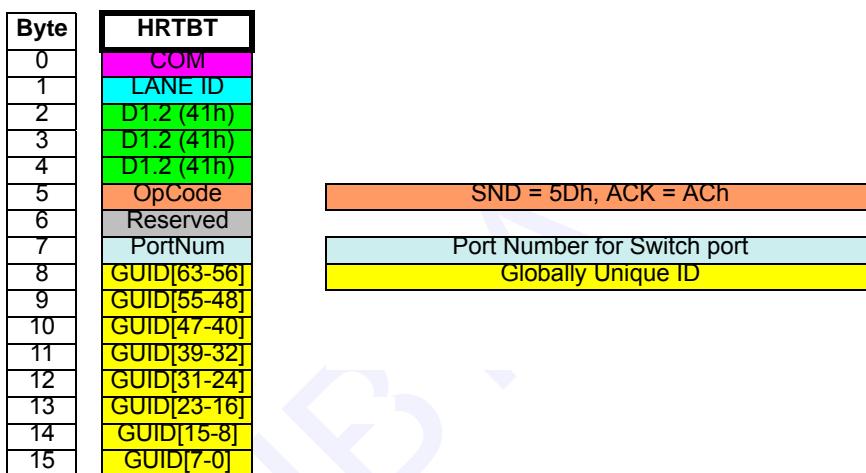


Figure 42 HRTBT Ordered-Set: Detailed Format

Symbol 5 is an OpCode, which signifies whether the Heartbeat is a Send Heartbeat (SND, OpCode = 5Dh), or an acknowledgment (ACK, OpCode = ACh).

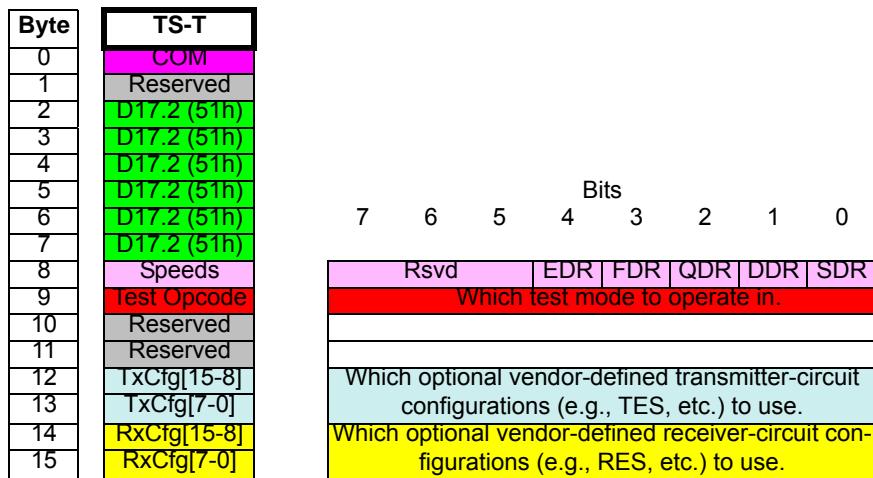
Symbol 7 and Symbols 8-15 indicates the port number and the GUID or the port that the SND heartbeat came from, and which port the ACK is going back to, and Symbols

The Link Heartbeat unique data symbol is D1.2 (or 41h), and the 10-bit encoded value is either 011101 0101 for negative running disparity, or 100010 0101 for positive running disparity.

5.5.2.6 TRAINING SEQUENCE FOR TEST ORDERED-SET (TS-T)

The Training Sequence for Test is a sixteen symbol ordered-set composed of a comma (COM), a reserved symbol replacing the usual Lane ID symbol, six data symbols unique to Training Sequence for Test, a bit map of the speed at which test phase is to occur (SDR, DDR, QDR, FDR, or EDR), a symbol indicating which of several defined test modes to use, and 2 16-bit transmitter configuration and receiver configuration fields, which define, in a vendor-specific way, how the transmitter and receiver are configured

during the test. The format of the TS-T ordered-set is shown in [Figure 43](#) below. The use of the TS-T ordered-set is described in [Section 5.17.2.2, “Use of the Training Sequence for Test Ordered-Set \(TS-T\).” on page 258.](#)



FDR and higher data rates: Preset Tx Equalization Settings:
See [Section 6.6.5.3, “Transmit Equalization Presets,” on page 298](#)

12	TxCfg[15-8]	Vendor-Defined	AMP	Preset 0-15
13	TxCfg[7-0]	Vendor-Defined		

Figure 43 TS-T Ordered-Set: Detailed Format

The Training Sequence for Test (TS-T) is only generated by test equipment. There is no need for an InfiniBand device to be able to generate this ordered-set - only the need to recognize it and behave appropriately when one complete and error free TS-T arrives at the receiver port on one or more lanes. Note that this is different than for TS2 or TS3, where 8 contiguous and valid copies of TS2 or TS3 are required to cause a receiver status change, in order to assure intact reception of link training configuration information. For TS1, a single ordered-set is usually enough to cause a state change, except in the Config.RcvfCfg state, where 8 are needed to establish that the receiver is fully configured.

Symbol 8 of a TS-T, the link speed identifier, is used to identify the speed at which the test equipment requests the IB device to operate. At least one bit of this bit map must be asserted to 1. If multiple bits are asserted, the test will be conducted at the highest *Link-SpeedEnabled* and *LinkSpeedExtEnabled* speed. This allows testing at the highest enabled speed by simply asserting all bits in this symbol.

Symbol 9 of a TS-T allows the test equipment to determine which testing mode the port will be placed in. The following values are defined.

0: SKIP-less Idle Data

For QDR or lower bit rates each transmitter lane transmits a pseudo-random sequence of data symbols, generated by the 11th order LFSR = $X^{11} + X^9 + 1$ with no insertion of SKIP ordered-sets. All lanes shall transmit data streams synchronized at the output of the transmitting IC.

For FDR and EDR bit rates each transmitter lane transmits a stream of scrambled idle blocks see [Section 5.5.3.5](#), with no insertion of SKIP ordered-sets.

1: SKIP-less back-to-back TS1s

2 Each transmitter lane transmits an unbroken string of TS1 ordered-sets, with 3 no insertion of SKIP ordered-sets. All lanes shall transmit data streams 4 synchronized at the output of the transmitting IC.

5 2: Receiver test.

6 For QDR or lower bit rates:

7 On each lane, the transmitter sends an indication of the validity of the data 8 received on the corresponding receiver lane.

9 VALID DATA: For each received symbol on a lane which decodes to a 10 valid 8b/10b code point with good running disparity, the corresponding 11 transmitter lane transmits a D10.2 (010101 0101) character.

12 LOGICAL ERROR: For each received symbol on a lane which decodes 13 with a logical error (e.g., bit error, or running disparity error), the corre- 14 sponding transmitter lane transmits a K28.5 D00.0 pair of symbols.

15 LOSS OF SIGNAL: For each received symbol on a lane which indicates 16 an inadequate signal (e.g., inadequate signal swing, all 0s, all 1s, or 17 noise), the corresponding transmitter lane transmits a K28.5 D01.0 pair of 18 symbols.

19 For FDR and EDR bit rates:

20 On each lane the transmitter sends a PRBS31 corresponding to the va- 21 lidity of the received PRBS31 on the corresponding receiver lane, when a 22 PRBS31 error was detected on the receiver lane the transmitter produces 23 a PRBS31 error on its transmitted stream.

24 3: High frequency pattern

25 Each transmitter lane transmits a high frequency pattern at the highest pos- 26 sible bit transition rate, a repeating 01010101 pattern. All lanes shall be syn- 27 chronized at the output of the transmitting IC.

28 4: PRBS31

29 Each transmitter lane transmits the PRBS31 pseudo-random sequence. Each 30 lane must send an uncorrelated PRBS31 pattern in order to simulate data 31 lane to lane cross talk conditions. The means by which a device achieves the 32 above requirement is implementation specific. It can be achieved by imple- 33 menting a different PRBS seed on each lane, by implementing a delay be- 34 tween the lanes or by any other method which ensures uncorrelated pattern 35 on each lane. Note that a device using delay to create the uncorrelated data 36

must delay by at least 132 UI to compensate for the maximum allowed lane to lane skew.

5: PRBS11

Each transmitter lane transmits the PRBS11 pseudo-random sequence. Each lane must send an uncorrelated PRBS11 pattern in order to simulate data lane to lane cross talk conditions. The means by which a device achieves the above requirement is implementation specific. It can be achieved by implementing a different PRBS seed on each lane, by implementing a delay between the lanes or by any other method which ensures uncorrelated pattern on each lane. Note that a device using delay to create the uncorrelated data must delay by at least 132 UI to compensate for the maximum allowed lane to lane skew.

6: PRBS9

Each transmitter lane transmits the PRBS9 pseudo-random sequence. Each lane must send an uncorrelated PRBS9 pattern in order to simulate data lane to lane cross talk conditions. The means by which a device achieves the above requirement is implementation specific. It can be achieved by implementing a different PRBS seed on each lane, by implementing a delay between the lanes or by any other method which ensures uncorrelated pattern on each lane. Note that a device using delay to create the uncorrelated data must delay by at least 132 UI to compensate for the maximum allowed lane to lane skew.

7: Back-to-back TS1s

All lanes shall transmit back-to-back TS1s with Skip ordered sets. The Skip ordered set insertion shall follow the rules defined in [Section 5.8.1, “Link De-skew Training Sequence and SKIP ordered sets,” on page 188](#). This mode may be used to measure the lane-to-lane skew. All lanes shall be synchronized at the output of the transmitting IC.

8-255: Optional Vendor-specific opcodes to allow other testing modes.

Symbol 10 and Symbol 11 of a TS-T are reserved. They are transmitted as D00.0 and valid Dxx.y symbols are ignored at the receiver.

For QDR or lower data rates:

Symbols 12 and 13 allow the test equipment to optionally configure the transmitter in one of 65,536 vendor-dependent states, and symbols 14 and 15 allow the test equipment to configure the receiver in one of 65,536 vendor-dependent states. The values 0000h are identified for “normal” or “default” operation, so that test equipment can be expected to get good and correct operation when these fields are set to 0. Values other than 0 are optional and vendor-dependent. Typically only a very small subset of these states will be valid and will provide specific unique behaviors.

For FDR or higher data rates:

Symbols 12 and 13 allow the test equipment to optionally configure the transmitter in one of the 16 Transmitter settings as defined in [Table 54, “Tx FIR Filter Coefficients and Amplitudes for 14.0625 Gb/s \(FDR\),” on page 299](#), using the 4 least significant bits of byte 12. Byte 12, bit number 4 shall represent the requested amplitude bit. Preset 0 shall represent the transmitter default preset used for active limiting or optical cables and must meet the host transmitter requirement as defined in [Table 55, “FDR host output specifications at Preset 0, for Limiting Active Cables,” on page 301](#). Presets 33 - 65,536 are vendor-dependent presets.

Symbols 14 and 15 allow the test equipment to configure the receiver in one of 65,536 vendor-dependent states. The values 0000h are identified for “normal” or “default” operation, so that test equipment can be expected to get good and correct operation when these fields are set to 0. Values other than 0 are optional and vendor-dependent. Typically only a very small subset of these states will be valid and will provide specific unique behaviors.

The TS-T unique data symbol is D17.2 (or 51h), and the 10-bit encoded value is the pattern (100011 0101) for both the positive and negative running disparity.

5.5.3 CONTROL BLOCK 64B/66B ENCODING

A number of control blocks (sync header of 10) that shall be used for packet delimiter and clock compensation are defined in [Table 27](#) below. The packet byte stream is created with Control Bytes and the control bit to indicate a block type of control. The 64b/66b encoder function uses the packet byte stream and the control bit to create the 66-bit control block.

Table 27 64b/66b Control Blocks

Ordered set	Block Type ^a	Size	description
SDP	78	8 Bytes	Start Data Packet block
SLP	55	8 Bytes	Start Link Packet block
EGP3	B4	8 Bytes	End Good Packet block with 3 data bytes
EGP7	FF	8 Bytes	End Good Packet block with 7 data bytes
EBP3	AA	8 Bytes	End Bad Packet block with 3 data bytes
EBP7	E1	8 Bytes	End Bad Packet block with 7 data bytes
Idle	1E	8 Bytes	Idle Block
SKP	4B	8 Bytes	Skip Block
TS1	4B	8 Bytes	Training Sequence 1
TS2	4B	8 Bytes	Training Sequence 2
FTB	4B	8 Bytes	Fine Tuning Block
HRTBT	4B	16 Bytes	HeartBeat ordered set
Vendor-specific	4B	8 Bytes	Vendor-specific Block

a. Block Types are in hexadecimal

5.5.3.1 START OF DATA PACKET DELIMITER (SDP)

The Start of Data Packet Block is transmitted to identify the start of a data packet. Packet formatting rules specify which physical lanes may be used by the “SDP” control block. (See [Section 5.7 on page 175](#)) The first 7 bytes of the packet are included in the SDP block at Byte position 1 through 7 as shown in [Table 28](#).

Table 28 64b/66b SDP Block

Type	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
SDP	78	Data 0	Data 1	Data 2	Data 3	Data 4	Data 5	Data 6

5.5.3.2 START OF LINK PACKET DELIMITER (SLP)

The Start of Link Packet Block is transmitted to identify the start of a link packet. Packet formatting rules specify which physical lanes may be used by the “SLP” control block. (See [Section 5.7 on page 175](#)) The link packet data is included in the SLP block at Byte position 1 through 7 as shown in [Table 29](#). For 64b/66b encoding the link packet size is a single block and the SLP delimiter represents the start and end of the link packet.

Table 29 64b/66b SLP Block

Type	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
SLP	55	Data 0	Data 1	Data 2	Data 3	Data 4	Data 5	Data 6

5.5.3.3 END OF GOOD PACKET DELIMITER (EGP)

The End of Good Packet Block is used to mark the end of each packet as it is transmitted by the originating port. Packet length rules restrict which physical lanes may be used to transmit the “EGP” control block. (See [Section 5.7 on page 175](#)). As all packets are an integer number of words (4 bytes) there are 2 different EGP blocks.

5.5.3.3.1 EGP3

The End of Good Packet 3 Block, shown in [Table 30](#), is used to allow termination of a packet with the last 3 bytes of data along with 4 bytes of PAD symbol (00h). The EGP3 block includes the last 3 bytes of the packet at Byte position 1 through 3.

Table 30 64b/66b EGP3 Block

Type	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
EGP3	B4	Data n-3	Data n-2	Data n-1	00	00	00	00

5.5.3.3.2 EGP7

The End Good Packet 7 Block is used to allow termination of a packet with the last 7 bytes of data. The EGP7 block includes the last 7 bytes of the packet at Byte position 1 through 7.

Table 31 64b/66b EGP7 Block

Type	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
EGP7	FF	Data n-7	Data n-6	Data n-5	Data n-4	Data n-3	Data n-2	Data n-1

5.5.3.4 END OF BAD PACKET DELIMITER (EBP)

The End of Bad Packet Block is used to mark the end of a bad packet forwarded by a switch or router node. When an error (e.g.: sync header error, CRC error, etc.) is detected in a data packet it is marked bad by replacing the original “EGP” block with a “EBP” block. The Error Detection Per Lane (EDPL) CRC-8 calculation is not used for packet-level error detection, and is not used in the algorithm for replacing EGP with EBP. A port that wishes to nullify a packet due to error shall do so by terminating the packet with the EBP3 or EBP7. Receiving end nodes are required to recognize either EGP or EBP as the end of packet delimiter. Any data packet terminated with “EBP” block must be treated as if it had a CRC error. (See [Section 5.7 on page 175](#)). There are 2 different EBP Blocks.

5.5.3.4.1 EBP3

The End of Bad Packet 3 Block is used to mark the end of a bad packet with the last 3 bytes of data and 4 bytes of PAD symbol (00h).

Table 32 64b/66b EBP3 Block

Type	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
EBP3	AA	Data n-3	Data n-2	Data n-1	00	00	00	00

5.5.3.4.2 EBP7

The End of Bad Packet 7 Block is used to mark the end of a bad packet with the last 7 bytes of data.

Table 33 64b/66b EBP7 Block

Type	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
EBP7	E1	Data n-7	Data n-6	Data n-5	Data n-4	Data n-3	Data n-2	Data n-1

5.5.3.5 IDLE BLOCK

The Idle block is transmitted between packets when no data or link packets are scheduled for transmission.(See [Section 5.7 on page 175](#))

Idle blocks may be removed or added by a switch or retiming device between packets.

Table 34 64b/66b Idle Block

Type	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
IDL	1E	00	00	00	00	00	00	00

5.5.3.6 SKIP BLOCK (SKP)

For 64b/66b encoding a SKIP ordered-set is a series of Skip Blocks, shown in [Table 35](#), which is used for clock tolerance compensation and for lane to lane de-skew at the receiver. SKIP ordered-set are added to the Packet Byte Stream on all lanes at the same block time. (See [Section 5.11 on page 239](#))

A SKIP ordered-set consists of a group of 1 to 5 consecutive Skip Blocks. A transmitter must transmit a SKIP ordered-set of 3 Skip Blocks.

The Skip Block is also used to identify errors on a lane using the EDPL field in the SKP block. (See [Section 5.3.4, “Error Detection Per Lane,” on page 104](#))

When RS-FEC is enabled, the EDPL field in the Skip Block should be sent as Rsvd. When RS-FEC is enabled, the number of corrected symbols can be used as an indication for the errors on the lane. See [Section 5.6.3, “Port Performance Counters,” on page 173](#)

Table 35 64b/66b Skip Block

Type	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
SKP	4B	1E	1E	1E	Rsvd	Rsvd	Rsvd	EDPL

5.5.3.7 TRAINING SEQUENCE ONE (TS1)

The Link Training Sequence One (TS1) is a single block composed of 3 bytes of TS1 identifier (4Ah) and 4 reserved bytes.

The TS1 is added in the Packet Byte Stream for each lane independently.

The 64b/66b TS1 does not contain a lane identifier.

Table 36 64b/66b TS1 Ordered-Set

Type	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
TS1	4B	4A	4A	4A	Rsvd	Rsvd	Rsvd	Rsvd

5.5.3.8 TRAINING SEQUENCE TWO (TS2)

The Link Training Sequence Two (TS2) is a single block composed of 3 bytes of TS2 identifier (45h) and 4 reserved bytes.

The TS2 is added in the Packet Byte Stream for each lane independently.

The 64b/66b TS2 does not contain a lane identifier.

Table 37 64b/66b TS2 Ordered-Set

Type	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
TS2	4B	45h	45h	45h	Rsvd	Rsvd	Rsvd	Rsvd

5.5.3.9 FINE TUNING BLOCK (FTB)

The Fine Tuning Block is used to perform the linkup and pre-linkup fine tuning algorithm. The FTB is inserted to the Packet Byte Stream by replacing the IDLE control and data. (See [Section 5.13 on page 246](#)).

Table 38 Fine Tuning Block

Type	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
FTB	4B	48h	CMD (command)	STAT (status)	Rsvd	Lane Requested	Rsvd	Rsvd

5.5.3.10 LINK HEARTBEAT ORDERED-SET (HRTBT)

The Link Heartbeat ordered-set (HRTBT) is a two block ordered-set composed of unique Ordered-set identifier (41h), a 1-byte OpCode, a 1-byte Reserved field, a 1-byte PortNum value (only used for switch ports), and an eight-symbol Globally Unique ID (GUID). The format of the HRTBT ordered-set is shown in [Table 39](#) below. The two blocks of the HRTBT ordered set shall be transmitted sequentially on the wire, with no blocks in between the two blocks of the HRTBT ordered set. The use of the HRTBT ordered-set is described in [Section 5.14, “Link Heartbeat,” on page 253](#).

Table 39 64b/66b Heartbeat Ordered-Set

Type	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
HRTBT	4B	41h	41h	Opcode	Rsvd	port num	GUID[63:56]	GUID[55:48]
	4B	41h	GUID[47:40]	GUID[39:32]	GUID[31:24]	GUID[23:16]	GUID[15:8]	GUID[7:0]

5.5.3.11 VENDOR-SPECIFIC BLOCK

[Table 40, "Vendor-specific Block," on page 164](#) describes a Vendor-specific block.

Table 40 Vendor-specific Block

Type	Byte 0	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7
Vendor-specific	4B	F4h				vendor-specific		

5.5.4 LOGICAL INTERFACES

This section describes the logical interfaces of the Link/Phy layer. This section is not intended to describe an actual implementation of a device but rather to explain the logical interface between protocol layers. This interface may or may not be implemented as an internal or external device interface.

The logical interface section defines three logical interfaces:

- 1) Data/Status interface to the upper layer, Logical Link to Link/Phy;
- 2) Interface to the lower layer, Link/Phy to Physical;
- 3) Control and Status interface to the Link/Phy layer.

5.5.4.1 LOGICAL LINK TO LINK/PHY DATA INTERFACE

The Logical Link to Link/Phy data interface is defined in the *InfiniBand Architecture Specification*, [Volume 1, Chapter 6: Physical Layer Interface](#).

5.5.4.1.1 PACKET BYTE STREAM

For 64b/66b encoding the Link/Phy layer create a Packet Byte Stream from the Transmit Stream received from the link logic. The Packet Byte Stream Function creates the stream by replacing the received message with the appropriate control symbol, aligning the packet stream to 8 bytes, and enforcing the Max Packet Rate as defined in [Section 5.15, "Max Packet Rate," on page 255](#).

5.5.4.2 LOGICAL LINK/PHY TO PHYSICAL INTERFACE

The Link/Phy to Physical interface consists of the following logical signals:

- 1) [PhyTxData\[11:0\]](#) - Transmit bit stream(s) 1x, 4x, 8x or 12x.
- 2) [PhyRxData\[11:0\]](#) - Receive bit stream(s) 1x, 4x, 8x or 12x.

5.5.4.3 LINK/PHY TO CONTROL AND STATUS INTERFACE

The Link/Phy control and status interface consists of the following logical signals:

- 1) PowerOnReset - Port Power on Reset input,
- 2) LinkPhyReset - Link Physical Reset control input,
- 3) LinkPhyRecover - Link Physical Recovery control input;

-
- 4) LinkPhyStat - Link Physical State (up or down) status output.

5.6 MANAGEMENT DATAGRAM CONTROL AND STATUS INTERFACE

The Management Datagram Control and Status interface is subdivided into two sub-sections: Control Inputs, and Status Outputs.

Implementation Note

Designers should not rely on the absence or characteristics of any features or commands marked “reserved” or “undefined”. The InfiniBand™ Trade Association reserves these for future definition.

C5-7: This compliance statement is obsolete and has been replaced by [C5-7.2.1](#):

C5-7.2.1: All ports shall implement the control and status management interface defined by [Section 5.6, “Management Datagram Control and Status Interface,” on page 165](#), excluding DDR/QDR interoperability, 8x interoperability, Phy Test compliance testing, and LinkRoundTripLatency.

o5-7.2.1: All ports claiming compliance with InfiniBand Rel. 1.2 shall implement the control and status management interface defined by [Section 5.6, “Management Datagram Control and Status Interface,” on page 165](#), including DDR/QDR interoperability, 8x interoperability, Phy Test compliance testing, and LinkRoundTripLatency.

o5-7.2.2: All ports claiming compliance with InfiniBand Rel. 1.3 shall implement the control and status management interface defined by [Section 5.6, “Management Datagram Control and Status Interface,” on page 165](#), including FDR operation and LinkSpeedExt functionality.

o5-7.2.3: All ports claiming compliance with InfiniBand Rel. 1.3.1 shall implement the control and status management interface defined by [Section 5.6, “Management Datagram Control and Status Interface,” on page 165](#), including EDR operation and LinkSpeedExt functionality.

5.6.1 CONTROL INPUTS (MAD SET)

Specific implementations may provide control information via proprietary mechanisms.

Multiple commands may be sent in the same **SM.PortInfo(component)** Management Datagram. Simultaneous commands, one of which changes the state of the Port Training State machine, shall set the other associated state variable(s) before the port training state change occurs.

The Control Input interface consists of the following Management Datagram to variable or logical signal mappings:

- 1) A Management Datagram **SM.PortInfo(PortPhysicalState)** set shall cause the Port
Training State machine to transition state based on the following enumerated
values:
 - 1 0: No State Change (NOP)
 - 2 1: Sleeping
 - 3 2: Polling
 - 4 3: Disabled
 - 5 4-15: Reserved (Ignored)

Refer to [Section 5.6.2](#) for fields returned by the MAD get operation. The default value
following power on shall be set to **Polling**.
- 2) A Management Datagram **SM.PortInfo(LinkDownDefaultState)** set shall set the
logical signal **LinkDownDefaultState** based on the following enumerated values:
 - 12 0: No State Change (NOP)
 - 13 1: Sleeping
 - 14 2: Polling
 - 15 3-15: Reserved (Ignored)

Refer to [Section 5.6.2](#) for fields returned by the MAD get operation. The default value
following power on shall be set to **Polling**.
- 3) A Management Datagram **SM.PortInfo(LinkWidthEnabled)** set shall set the **Link-
WidthEnabled** variable. No action shall be taken upon the variable until the Port
Training State machine transitions to the **Configuration** state. The port shall only at-
tempt to configure the link to width(s) based on the following enumerated values:
 - 22 0: No State Change (NOP)
 - 23 1: 1x
 - 24 2: 4x
 - 25 3: 1x or 4x
 - 26 4: 8x
 - 27 5: 1x or 8x
 - 28 6: 4x or 8x
 - 29 7: 1x or 4x or 8x
 - 30 8: 12x
 - 31 9: 1x or 12x
 - 32 10: 4x or 12x
 - 33 11: 1x, 4x or 12x
 - 34 12: 8x or 12x

- 13: 1x or 8x or 12x
- 14: 4x or 8x or 12x
- 15: 1x or 4x or 8x or 12x
- 16-254: Reserved (Ignored)
- 255: Set to **LinkWidthSupported** value

Refer to [Section 5.6.2](#) for fields returned by the MAD get operation. The default value following power on shall be set to **LinkWidthSupported**.

- 4) A Management Datagram **SM.PortInfo(LinkSpeedEnabled)** set shall set the **LinkSpeedEnabled** variable. No action shall be taken upon the variable until the Port Training State machine transitions to the **Configuration** state. The port shall only attempt to configure the link to a speed based on the following enumerated values:

- 0: No State Change (NOP)
- 1: 2.5 Gb/s (SDR) (Also, when **SMSupportsExtendedSpeeds** is set to 0 in request/response AM, disable any extended speeds)
- 3: 2.5 or 5.0 Gb/s (SDR or DDR) (Also, when **SMSupportsExtendedSpeeds** is set to 0 in request/response AM, disable any extended speeds)
- 5: 2.5 or 10.0 Gb/s (SDR or QDR)
- 7: 2.5 or 5.0 or 10.0 Gb/s (SDR or DDR or QDR)
- 2, 4, 6, 8-14: Reserved (ignored)

15: Set to **LinkSpeedSupported** value; Response contains actual **LinkSpeedSupported**. In addition to the **LinkSpeedSupported** value, when **SMSupportsExtendedSpeeds** is set to 0 in request/response AM, response contains QDR indication if an extended speed is enabled.

Note that the SDR speed (2.5 Gb/s) must always be enabled, since link initialization occurs at this speed. Refer to [Section 5.6.2](#) for fields returned by the MAD get operation. The default value following power on shall be set to **LinkSpeedSupported**.

- 5) If the port supports the extended link speed option as indicated in **SM.PortInfo.CapabilityMask2.IsExtendedSpeedsSupported**, a Management Datagram **SM.PortInfo(LinkSpeedExtEnabled)** set shall set the **LinkSpeedExtEnabled** variable. No action shall be taken upon the variable until the Port Training State machine transitions to the **Configuration** state. The port shall only attempt to configure the link to a speed based on the following enumerated values:

- 0: No State Change (NOP)
- 1: 14.0625 Gb/s (FDR)
- 2: 25.78125 Gb/s (EDR)
- 3: 14.0625 Gb/s (FDR) or 25.78125 Gb/s (EDR)
- 4-29: Reserved (ignored)

30: Disable extended link speeds	1
31: Set to LinkSpeedExtSupported value; response contains actual Link-SpeedExtSupported.	2
6) If the port supports Forward Error Correction coding, as indicated in SM.PortInfoExtended.CapabilityMask.IsFECModeSupported, a Management Datagram set shall set the Port FEC Enabled variable for operation at FDR and/or EDR speeds, using the SM.PortInfoExtended(FDRFECModeEnabled) and SM.PortInfoExtended(EDRFECModeEnabled) fields. No action shall be taken upon the variable until the Port Training State machine transitions to the Configuration state. The port shall only attempt to configure the link to FEC types based on the following enumerated values. Enumerated values that include Fire-Code FEC shall only be used for operation at FDR speed. Setting of the Fire-code bit shall be ignored for operation at other speeds.	3
0: No change	4
1: Enable operation only without any FEC	5
2: Fire-Code forward error correction (FDR only)	6
3: Fire-Code and No FEC enabled (FDR only)	7
4: RS-FEC(528,514) - Reed-Solomon (528,514) forward error correction	8
5: RS-FEC(528,514) and No FEC enabled	9
6: RS-FEC(528,514) and Fire-Code enabled (FDR only)	10
7: RS-FEC(528,514), Fire-Code, and NoFEC enabled (FDR only)	11
8: RS-FEC(271,257) - Low Latency Reed-Solomon (271,257) forward error correction	12
9: RS-FEC(271,257) and No FEC enabled	13
10: RS-FEC(271,257) and Fire-Code enabled (FDR only)	14
11: RS-FEC(271,257), Fire-Code and No FEC enabled (FDR only)	15
12: RS-FEC(271,257) and RS-FEC(528,514) enabled	16
13: RS-FEC(271,257), RS-FEC(528,514) and No FEC enabled	17
14: RS-FEC(271,257), RS-FEC(528,514) and Fire-Code enabled (FDR only)	18
15: RS-FEC(271,257), RS-FEC(528,514), Fire-Code, and No FEC enabled (FDR only)	19
16-65535: Reserved	20

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

Implementation Note

MAD responders are required to respond to a **SM.PortInfo** Set() command with a GetResp() MAD. (*InfiniBand Architecture Specification, Volume 1, Section 13.4.6.1*) The **SM.PortInfo(PortPhysicalState)** set will cause the port training state machine to down the port when implementing the command. When this MAD is received on the port affected by the port state change command, the GetResp() MAD may not be transmitted before the port responds to the command. When a command is issued in this manner, the requester should not expect a GetResp() MAD. Command execution can be verified by observing the state change at the other end of the link.

Implementation Note

Subnet Manager software designers should be aware that when a powered port is commanded to the Disabled State, the port will no longer respond to received packets or beacons. The port can only be re-enabled by using an alternate path, a second port on the same node, or some out-of-band connection. If there is no alternate path, the port must be reset to recover normal operation. For example, a single port IB device without an out-of-band management path would have to be physically reset to exit the Disable State.

5.6.2 STATUS OUTPUTS (MAD GET)

Specific implementations may provide status via proprietary mechanisms.

The Status Output interface consists of the following Management Datagram to variable or logical signal mappings:

- 1) A Management Datagram **SM.PortInfo(PortPhysicalState)** get shall return the Port Training State machines current state as one of the following enumerated values:
 - 1: Sleeping
 - 2: Polling
 - 3: Disabled
 - 4: Configuration
 - 5: LinkUp
 - 6: Recovery
 - 7: Phy Test
 - 0, 8-15: Reserved
- 2) A Management Datagram **SM.PortInfo(LinkDownDefaultState)** get shall return the **LinkDownDefaultState** current value based on the following enumerated values:
 - 1: Sleeping
 - 2: Polling

- 0, 3-15: Reserved
- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
- 3) A Management Datagram **SM.PortInfo(LinkWidthEnabled)** get shall return the previously written **LinkWidthEnabled** value based on the following enumerated values:
- 1: 1x
2: 4x
3: 1x or 4x
4: 8x
5: 1x or 8x
6: 4x or 8x
7: 1x or 4x or 8x
8: 12x
9: 1x or 12x
10: 4x or 12x
11: 1x, 4x or 12x
12: 8x or 12x
13: 1x or 8x or 12x
14: 4x or 8x or 12x
15: 1x or 4x or 8x or 12x
- 0, 16-255: Reserved
- 4) A Management Datagram **SM.PortInfo(LinkSpeedEnabled)** get shall return the previously written **LinkSpeedEnabled** value based on the following enumerated values:
- 1: 2.5 Gb/s (SDR)
3: 2.5 or 5.0 Gb/s (SDR or DDR)
5: 2.5 or 10.0 Gb/s (SDR or QDR) (or higher when **SMSupportsExtended-Speeds** is set to 0 in request/response AM)
7: 2.5 or 5.0 or 10.0 Gb/s (SDR or DDR or QDR) (or higher when **SMSupportsExtendedSpeeds** is set to 0 in request/response AM)
- 0, 2, 4, 8-15: Reserved
- 5) A Management Datagram **SM.PortInfo(LinkWidthSupported)** get shall return the supported width(s) based on the following enumerated values:
- 1: 1x
3: 1x or 4x
7: 1x, 4x or 8x

- 11: 1x, 4x, or 12x (not valid for products supporting Rel. 1.2.1 or higher) 1
15: 1x, 4x, 8x or 12x 2
0, 2, 4-6, 8-10,12-14, 16-255: Reserved 3
6) A Management Datagram **SM.PortInfo(LinkWidthActive)** get shall return the Port 4
Training State machines currently configured width based on the following enum- 5
erated values: 6
1: 1x 7
2: 4x 8
4: 8x 9
8: 12x 10
0, 3, 5-7, 9-255: Reserved 11
7) A Management Datagram **SM.PortInfo(LinkSpeedSupported)** get shall return the 14
supported speeds based on the following enumerated values: 15
1: 2.5 Gb/s (SDR) 16
3: 2.5 or 5.0 Gb/s (SDR or DDR) 17
5: 2.5 or 10.0 Gb/s (SDR or QDR) (or higher when **SMSSupportsExtended- 18
Speeds** is set to 0 in request/response AM) 19
7: 2.5 or 5.0 or 10.0 Gb/s (SDR or DDR or QDR) (or higher when **SMSSupport- 20
sExtendedSpeeds** is set to 0 in request/response AM) 21
0, 2, 4, 6, 8-15: Reserved 22
8) A Management Datagram **SM.PortInfo(LinkSpeedActive)** get shall return the Port 24
Training State machines currently configured speed based on the following enum- 25
erated values: 26
1: 2.5 Gb/s (SDR) 27
2: 5.0 Gb/s (DDR) 28
4: 10 Gb/s (QDR) (or higher when **SMSSupportsExtendedSpeeds** is set to 0 in 29
request/response AM). 30
3, 5-15: Reserved 31
0: Reserved (when **SMSSupportsExtendedSpeeds** is set to 0 in request/re- 32
sponse AM). 33
0: Extended Speed Active (when **SMSSupportsExtendedSpeeds** is set to 1 in re- 34
quest/response AM). 35
Note: Either **LinkSpeedActive** or **LinkSpeedExtActive** should be non zero 36
when **SMSSupportsExtendedSpeeds** is set to 1 in request/response AM. 37
9) A Management Datagram **SM.PortInfo(LinkRoundTripLatency)** get shall return 38
the minimum measured round trip link latency from the port to the connected port on 39
the opposite side of the link. The **LinkRoundTripLatency** is a 32-bit value repre- 40
41
42

senting the minimum measured time for a bit in Link Heartbeat ordered-set to tra-
verse the link in both directions, and can take on the following values:

FFFF_FFFFh: Link round-trip latency not yet measured, following reset on entry
to **LinkDownDefaultState**.

Others (0000_0000h to FFFF_FFFEh): Link round-trip latency between when a
SND HEARTBEAT is transmitted and its corresponding ACK HEARTBEAT is re-
turned, reported in 4 nanosecond intervals.

Measurement of LinkRoundTripLatency using HRTBT ordered-sets is described in
[Section 5.14.1, “Operation of Link Heartbeats,” on page 253](#).

- 10) If the port supports the extended link speed option as indicated in **SM.PortInfo.Ca-**
pabilityMask2.IsExtendedSpeedsSupported, a Management Datagram
SM.PortInfo(LinkSpeedExtEnabled) get shall return the previously written **Link-**
SpeedExtEnabled value based on the following enumerated values:

- 0: No extended speed enabled
- 1: 14.0625 Gb/s (FDR)
- 2: 25.78125 Gb/s (EDR)
- 3: 14.0625 Gb/s (FDR) or 25.78125 Gb/s (EDR)
- 4-29: Reserved
- 30: Extended link speeds disabled
- 31: Reserved

- 11) If the port supports the extended link speed option as indicated in **SM.PortInfo.Ca-**
pabilityMask.IsExtendedSpeedsSupported, a Management Datagram
SM.PortInfo(LinkSpeedExtSupported) get shall return the supported extended
speeds based on the following enumerated values:

- 0: No extended speed supported
- 1: 14.0625 Gb/s (FDR)
- 2: 25.78125 Gb/s (EDR)
- 3: 14.0625Gb/s (FDR) or 25.78125Gb/s (EDR)
- 4-15: Reserved

- 12) If the port supports the extended link speed option as indicated in **SM.PortInfo.Ca-**
pabilityMask2.IsPortInfoExtendedSupported, a Management Datagram
SM.PortInfo(LinkSpeedExtActive) get shall return the Port Training State ma-
chines currently configured extended speed based on the following enumerated
values:

- 0: No extended speed active
- 1: 14.0625 Gb/s (FDR)
- 2: 25.78125 Gb/s (EDR)
- 3-15: Reserved

- 13) If the port supports Forward Error Correction coding, as indicated in **SM.PortInfoExtended.CapabilityMask.IsFECModeSupported**, a Management Datagram **SM.PortInfoExtended(FDRFECModeSupported)** or **SM.PortInfoExtended(EDRFECModeSupported)** get shall return the Port supported FEC types in **FDRFECModeSupported** and **EDRFECModeSupported** fields:
- 1: Operation with No forward error correction coding is supported
 - 2: Fire-Code forward error correction (FDR only)
 - 3: Fire-Code and No FEC supported (FDR only)
 - 4: RS-FEC(528,514) - Reed-Solomon (528,514) forward error correction
 - 5: RS-FEC(528,514) and No FEC supported
 - 6: RS-FEC(528,514) and Fire-Code supported (FDR only)
 - 7: RS-FEC(528,514), Fire-Code, and NoFEC supported (FDR only)
 - 8: RS-FEC(271,257) - Low Latency Reed-Solomon (271,257) forward error correction
 - 9: RS-FEC(271,257) and No FEC supported
 - 10: RS-FEC(271,257) and Fire-Code supported (FDR only)
 - 11: RS-FEC(271,257), Fire-Code and No FEC supported (FDR only)
 - 12: RS-FEC(271,257) and RS-FEC(528,514) supported
 - 13: RS-FEC(271,257), RS-FEC(528,514) and No FEC supported
 - 14: RS-FEC(271,257), RS-FEC(528,514) and Fire-Code supported (FDR only)
 - 15: RS-FEC(271,257), RS-FEC(528,514), Fire-Code, and No FEC supported (FDR only)
 - 16: Reserved
 - 17: Reserved
 - 18: Reserved
 - 19: Reserved
 - 20: Reserved
 - 21: Reserved
 - 22: Reserved
 - 23: Reserved
 - 24: Reserved
 - 25: Reserved
 - 26: Reserved
 - 27: Reserved
 - 28: Reserved
 - 29: Reserved
 - 30: Reserved
 - 31: No FEC active
 - 32: Fire-Code forward error correction active (only valid for operation at FDR speed)
 - 33: RS-FEC(528,514) - Reed-Solomon (528,514) forward error correction active
 - 34: RS-FEC(271,257) - Low Latency Reed-Solomon (271,257) forward error correction active
 - 35: Reserved
 - 36: Reserved
 - 37: Reserved
 - 38: Reserved
 - 39: Reserved
 - 40: Reserved
 - 41: Reserved
 - 42: Reserved
- 14) If the port supports Forward Error Correction coding, as indicated in **SM.PortInfoExtended.CapabilityMask.IsFECModeSupported**, a Management Datagram **SM.PortInfoExtended(FECModeActive)** get shall return the Port current configured FEC based on the following enumerated values:
- 0: No FEC active
 - 1: Fire-Code forward error correction active (only valid for operation at FDR speed)
 - 2: RS-FEC(528,514) - Reed-Solomon (528,514) forward error correction active
 - 3: RS-FEC(271,257) - Low Latency Reed-Solomon (271,257) forward error correction active
 - 4-65535: Reserved

5.6.3 PORT PERFORMANCE COUNTERS

Each port implements the following performance counters. These counters are accessed using the Performance Management command defined in *InfiniBand Architecture Spec-*

ification, [Volume 1, Section 15.2](#). The Link Physical Performance Counters shall implement both the “get” and “set” performance management methods. These counters do not rollover, but shall stop at their maximum count, and a “set” operation is required to re-enable error counting. Specific implementations may provide performance information via proprietary mechanisms. Some of the port performance counters are valid only when FEC is enabled, in addition, some of the performance counters interpretation depends on the FEC type in use. The FEC enable state and the FEC type being used can be queried in **SM.PortInfoExtended(FECModeActive)**. The Performance Management interface shall consist of the following performance counters:

- 1) A Management Datagram **Perf.PortCounters(SymbolErrorCounter)** read shall return the current value of the 16-bit counter **SymbolErrorCounter**. This counter is incremented each time an error is detected on one or more lanes. (See [Section 5.9.2, “Minor Link Physical Errors Events,” on page 236](#))
- 2) A Management Datagram **Perf.PortCounters(LinkErrorRecoveryCounter)** read shall return the current value of the 8-bit counter **LinkErrorRecoveryCounter**. This counter is incremented each time the Port Training State machine successfully completes the link error recovery process. (See [Section 5.8.4.8, “Link Error Recovery States,” on page 222](#))
- 3) A Management Datagram **Perf.PortCounters(LinkDownedCounter)** read shall return the current value of the 8-bit counter **LinkDownedCounter**. This counter is incremented each time the Port Training State machine fails the link error recovery process and downs the link. (See [Section 5.8.4.8, “Link Error Recovery States,” on page 222](#)). This counter is also incremented each time a Link Heartbeat error returns the Link Training State Machine to the **LinkDownDefaultState**. (See [Section 5.14.2, “Heartbeat Error Handling,” on page 254](#)).
- 4) A Management Datagram **Perf.PortExtendedSpeedsCounters(SynchHeader-ErrorCounter)** read shall return the current value of the 16-bit counter **SynchHeaderErrorCounter**. This counter is incremented each time an invalid sync header is detected on one or more lanes. (See [Section 5.9.2, “Minor Link Physical Errors Events,” on page 236](#)).
- 5) A Management Datagram **Perf.PortExtendedSpeedsCounters(UnknownBlock-Counter)** read shall return the current value of the 16-bit counter **UnknownBlockCounter**. This counter is incremented each time one or more of the following is true:
 - a) An invalid sync header is detected on one or more lanes.
 - b) A Control block with unknown Block type is detected on one or more lanes.
 - c) An Errorred control block is detected on one or more lanes - e.g., Idle block with one or more bytes not equal to PAD byte (00h).See [Section 5.9.2, “Minor Link Physical Errors Events,” on page 236](#).
- 6) A Management Datagram **Perf.PortExtendedSpeedsCounters(ErrorDetection-LaneCounter)** read shall return the current value of the per lane 16-bit counter **ErrorDetectionCounterLane<n>**. The per lane counter is incremented each time an EDPL error is detected on that lane, as described in [Section 5.3.4, “Error Detection Per Lane,” on page 104](#) and [Section 5.9.2, “Minor Link Physical Errors Events,” on page 236](#).

[page 236](#). When RS-FEC is active, the EDPL counters are reserved, and should be ignored, as the RS-FEC separately detects errors per lane (**FECCorrectedSymbolCounterLane< n >**). When RS-FEC is not active or **IsRSFECCountersSupported** is set to 0, counters are supported as **ErrorDetectionPerCounterLane< n >**.

- 7) A Management Datagram **Perf.PortExtendedSpeedsCounters(FECCorrectableBlockCounterLane< n >)** read shall return the current value of the per lane for Fire-Code FEC and per port for RS-FEC 32-bit counter **PortFECCorrectableBlockCounter**. For Fire-Code FEC, the per lane counter is incremented each time a correctable error is detected and corrected on that lane - See [Section 5.4.1.3.1, “Fire-Code FEC Error Counters,” on page 111](#). For RS-FEC, the counter increments each time a correctable codeword is corrected by the RS FEC. See [Section 5.4.2.2.2, “RS-FEC Decoder Function,” on page 136](#). FEC correctable error counter is an informative counter and is not considered an error.
- 8) A Management Datagram **Perf.PortExtendedSpeedsCounters(FECUncorrectableBlockCounterLane< n >)** read shall return the current value of the per lane for Fire-Code FEC and per port for RS-FEC 32-bit counter **PortFECUncorrectableBlockCounter**. For Fire-Code, the per lane counter is incremented each time an uncorrectable error is detected on that lane, as described in [Section 5.4.1.3.1, “Fire-Code FEC Error Counters,” on page 111](#) and [Section 5.9.2, “Minor Link Physical Errors Events,” on page 236](#). For RS-FEC, the per port counter is incremented each time the RS-FEC decoder is unable to correct the FEC codeword. See [Section 5.4.2.2.2, “RS-FEC Decoder Function,” on page 136](#). The FEC enable state and the FEC type being used can be queried in **SM.PortInfoExtended(FECMode-Active)**
- 9) A Management Datagram **Perf.PortExtendedSpeedCounters(FECCorrectedSymbolCounterLane< n >)** read shall return the per lane RS-FEC symbol error counter. This 32-bit counter is incremented for each symbol received on the lane that the RS-FEC decoder corrects. See [Section 5.4.2.2.2, “RS-FEC Decoder Function,” on page 136](#).
- 10) A Management Datagram **Perf.PortExtendedSpeedCounters (PortFECCorrectedSymbolCounter)** read shall return the per port RS-FEC symbol error counter. This 32-bit counter is incremented for each symbol that the RS-FEC decoder corrects, on any lane. See [Section 5.4.2.2.2, “RS-FEC Decoder Function,” on page 136](#) and [Section 5.9.2, “Minor Link Physical Errors Events,” on page 236](#).

5.7 PACKET FORMATS FOR SINGLE AND MULTI LANE SUPPORT

This section describes the distribution and translation of the Link Layer packet byte stream to the physical lanes.

In addition to the control ordered-sets defined earlier, data symbols/blocks are used as the payload of link and data packets and as link idle data. Packet data symbols/blocks are framed by start of packet symbols/blocks SDP and SLP and end of packet symbols/blocks EGP, EGP3 or EGP7 and EBP, EBP3 or EBP7. Idle data symbols are not part of a link or data packet and are not framed by packet delimiters symbols/blocks.

5.7.1 LINK PACKET ORDERING

The upper layers of the Protocol provide a stream of packets. The packets in the stream are composed of:

- 1) A Local Routing Header and other optional headers.
- 2) A packet payload, type dependent.
- 3) A Invariant CRC, type dependent.
- 4) And a Variant CRC.

The combined length of packet header(s), payload, and Invariant CRC are a multiple of four bytes. Two bytes of a Variant CRC plus packet start delimiter and packet end delimiter ensure that all packet are a multiple of four bytes in length. See *InfiniBand Architecture Specification, Volume 1, Chapters 6 & 7* for complete details of internal packet formatting. The Link/Physical layer forwards these packets in the order received. As required, the Link/Physical layer will insert SKIP ordered-sets between packets. (See [Section 5.11.2](#)) An example of packet ordering on a 4x link is depicted below in [Figure 44](#) and [Figure 45](#). When there is no packet or SKIP ordered-sets set to transmit the Link/Physical layer will fill the link Idle with a pseudo-random sequence of data symbols (idle data) for 8b/10b encoding and with Idle blocks for 64b/66b encoding.

C5-8: All ports shall implement link packet ordering as defined by [Section 5.7.1, “Link Packet Ordering,” on page 176](#).

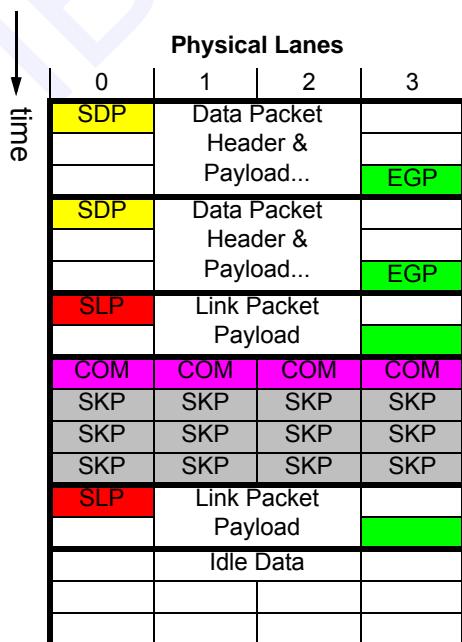


Figure 44 Packet Ordering Example 8b/10b

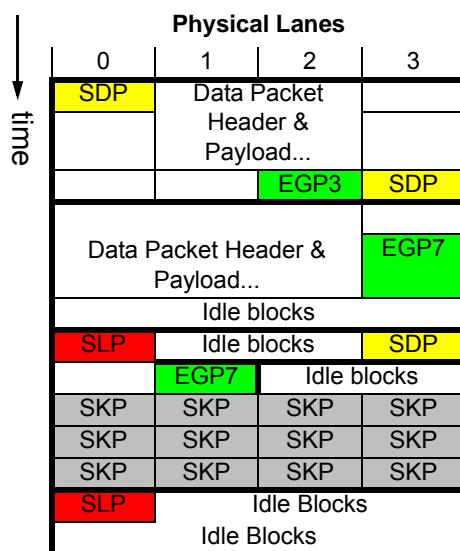


Figure 45 Packet Ordering Example 64b/66b

5.7.1.1 PACKET ORDERING RULES

- 1) Packets contain un-interruptible packet data content and delimiters. Ordered-sets and control symbols/blocks shall not be inserted into a packet's data content except as described in rule 4 below.
- 2) Packets received from the upper layers shall be transmitted in the order received.
- 3) Scheduled SKIP ordered-sets shall only be inserted between packets for clock tolerance compensation.
- 4) To initiate the error recovery process, TS1 ordered-sets shall be inserted at any symbol/block boundary, possibly interrupting the current packet.
- 5) When the link is idle (no packets or control ordered-sets to transmit)-
 - For 8b/10b encoding- a pseudo-random sequence of data symbols (idle data) shall be transmitted on all lanes.
 - For 64b/66b encoding - Idle blocks shall be transmitted on all lanes. Idle blocks may be replaced by FTB blocks if fine tuning of the link is performed - see [Section 5.13, "Fine Tuning," on page 246](#)
- 6) The 8b/10b link idle pseudo-random data sequence shall be generated by the 11th order LFSR = $X^{11} + X^9 + 1$. For 64b/66b the idle blocks are idle symbols (00h) scrambled using the scrambler defined in [Section 5.3.1](#)
- 7) Each lane may start the link idle pseudo-random data pattern at an arbitrary valid value. No lane-to-lane dependence is specified for the link idle pseudo-random data pattern.
- 8) Idle data shall be terminated at any time there is a packet or control ordered-set to transmit.

When the Fine Tuning protocol is supported, Idle blocks shall be terminated any time a FTB is scheduled to be transmitted

- 9) For 64b/66b encoding an SDP delimiter shall not be scheduled for transmission within less than 64 bytes of the previous SDP delimiter, unless MPR_en is set to false. (See [Section 5.15, "Max Packet Rate," on page 255](#)).

5.7.2 PACKET FORMATS

Packets including start and end delimiters are formed by the upper layers of the protocol. This stream of packets is striped across the available physical lane(s) (1x, 4x, 8x, or 12x).

C5-9: All ports shall implement packet formatting as defined by [Section 5.7.2, "Packet Formats," on page 178](#).

5.7.2.1 PACKET FORMATTING RULES

- 1) Total length of data packets including packet delimiter symbols/blocks shall be integer multiples of four symbols.
- 2) Data packets shall have an SDP symbol as the first symbol of the packet.
- 3) In 8b/10b encoding, link packets shall be eight symbols long including the packet delimiter symbols. In 64b/66b encoding, a link packet shall be wholly contained within one 66-bit block.
- 4) Link packets
 - For 8b/10b encoding - shall have a SLP symbol as the first symbol of the packet.
 - For 64b/66b encoding - shall be a single SLP block composed of one block type byte (55h) and 7 bytes of data.
- 5) When transmitted, all packets shall be terminated by an EGP or EBP symbol for 8b/10b encoding. For 64b/66b encoding all data packets shall be terminated by EGP3, EGP7, EBP3 or EBP7.
- 6) For 8b/10b encoding there are no per-lane even or odd alignment restrictions. Packets are not required to start with even or odd alignments. Comma symbols do not force even alignment.
- 7) For 8b/10b encoding the starting running disparity of packet delimiters and ordered-sets is not specified; a packet or ordered-set may start with positive or negative disparity. The disparity of all symbols shall comply with 8b/10b encoding rules. (See [Section 5.2.3](#))

5.7.3 1x PACKET FORMAT

The 1x link is composed of a single physical lane. For 8b/10b encoding, the combined packet symbol stream is serialized into a single stream of symbols. For 64b/66b encoding, the Packet Byte Stream is serialized into a single lane Byte Stream. The SKIP ordered-set is inserted between packets as needed for clock tolerance compensation. For SDR, DDR and QDR data rates the combined symbol stream (packets, SKIP ordered-sets, and idle data) is encoded using the 8b/10b code defined in [Section 5.2](#) above. For higher data rate the Lane Byte Stream is scrambled into a Scrambled Lane Stream and then encoded using a 64b/66b encoder into a Lane Block Stream defined in

[Section 5.3](#) above. A 1x symbol/Lane Blocks stream containing data packets, link packet, idle data, and SKIP ordered-sets is illustrated in [Figure 46](#) and [Figure 47](#) below.

C5-10: All 1x ports and 4x and 12x ports when configured as a 1x port shall implement packet formatting as defined by [Section 5.7.3, “1x Packet Format,” on page 178](#).

o5-10.2.1: All 8x ports when configured as a 1x port shall implement packet formatting as defined by [Section 5.7.3, “1x Packet Format,” on page 178](#).

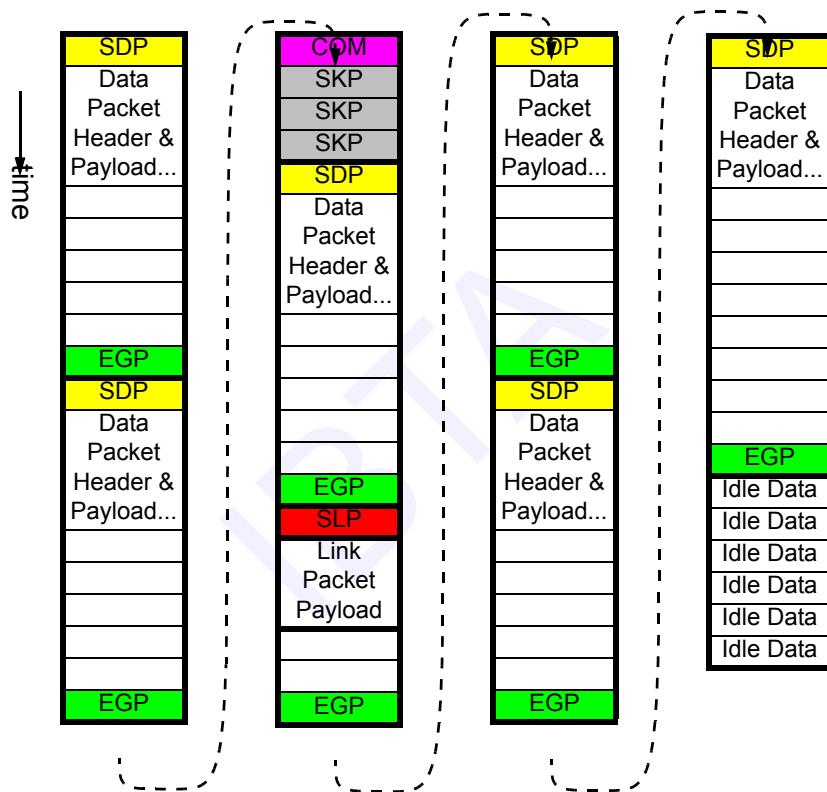


Figure 46 1x Packet Formats - 8b/10b

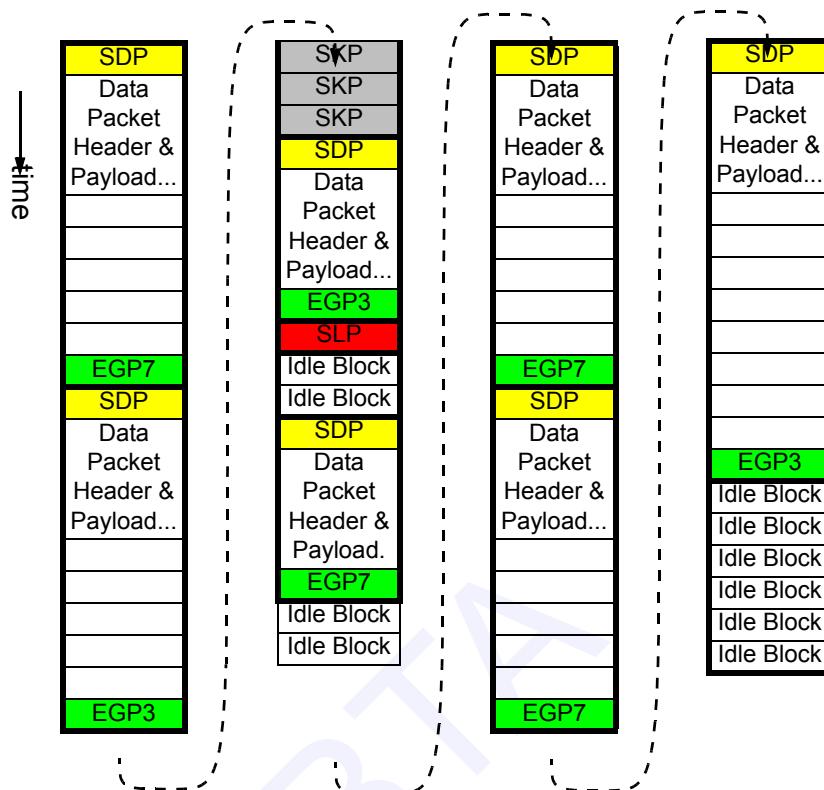


Figure 47 1x Packet Formats - 64b/66b

5.7.3.1 1x LINK FORMATTING RULES

- 1) The start of packet delimiters (SDP & SLP) shall be transmitted in lane zero only.
- 2) The end of packet delimiters (EGP & EBP and EGP3, EGP7, EBP3 & EBP7) shall be transmitted in lane zero only.
- 3) For 64b/66b encoding if MPR_en is true - an SDP shall be transmitted at least 64 bytes (8 Blocks) after the transmission of the previous SDP delimiter.

5.7.4 4x PACKET FORMAT

The 4x link is composed of four physical lanes (0 through 3). For 8b/10b encoding, the combined data and link packet symbol stream (control and data) are byte striped across physical lanes 0 through 3. For 64b/66b encoding, Packet Byte Stream (control and data) are 8 byte (block) striped across physical lanes 0 through 3 to form a Lane Byte Stream. In the 4x configuration: for 8b/10b encoding the start packet delimiters will always be transmitted on physical lane 0 and the end packet delimiters will always be transmitted on physical lane 3; for 64b/66b encoding the start and end delimiters may be transmitted on any lane. When SKIP ordered-sets are needed for clock tolerance compensation, they are inserted between packets simultaneously on all physical lanes. For SDR, DDR and QDR data rates the four symbol streams composed of packets, SKIP ordered-sets, and

idle data are individually encoded using the 8b/10b code defined in [Section 5.2](#) above. For higher data rates each Lane Byte Stream is scrambled into a Scrambled Lane Stream and encoded using 64b/66b encoding into a Lane Block Stream as defined in [Section 5.3](#) above. The 4x symbol/Lane Block streams containing data packets, link packets, and SKIP ordered-sets are illustrated in [Figure 48](#) and [Figure 49](#) below.

C5-11: All 4x ports and 12x ports when configured as a 4x port shall implement packet formatting as defined by [Section 5.7.4, “4x Packet Format,” on page 180](#).

o5-11.2.1: All 8x ports when configured as a 4x port shall implement packet formatting as defined by [Section 5.7.4, “4x Packet Format,” on page 180](#).

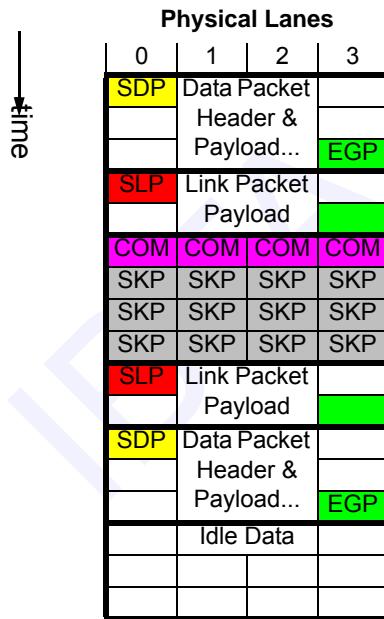


Figure 48 4x Packet Formats - 8b/10b

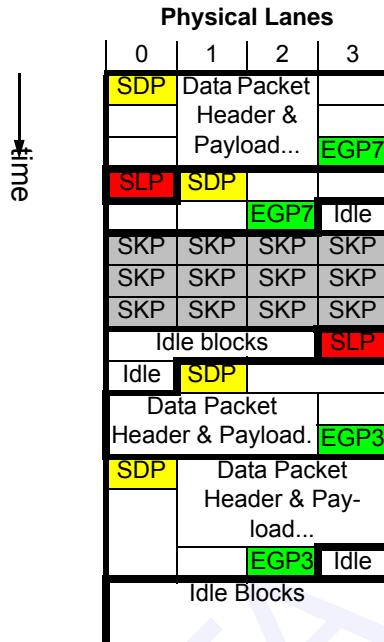


Figure 49 4x Packet Formats - 64b/66b

5.7.4.1 4X PACKET FORMATTING RULES

- 1) For 8b/10b encoding the start of packet delimiters (SDP & SLP) shall be transmitted in lane zero (0) only.
For 64b/66b encoding the SDP and SLP block may be transmitted on any lane.
- 2) For 8b/10b encoding the end of packet delimiters (EGP & EBP) shall be transmitted in lane three (3) only. For 64b/66b encoding the EGP3, EGP7, EBP3, and EBP7 blocks may be transmitted on any lane.
- 3) SKIP ordered-sets shall be transmitted on all (4) lanes simultaneously.
- 4) For 8b/10b encoding when the idle data is placed on the link, the idle data pattern shall be inserted on all four(4) lanes simultaneously. For 64b/66b encoding Idle blocks may be placed on any lane when there is no packet or ordered set scheduled for transmission.
- 5) For 8b/10b encoding the pseudo-random idle data for each lane should start at a different point in the sequence.

5.7.5 8X PACKET FORMAT

The 8x link is composed of eight physical lanes (0 through 7). For 8b/10b encoding the combined data and link packet symbol stream (control and data) are byte striped across physical lanes 0 through 7, for 64b/66b encoding the Packet Byte Stream (control and data) are 8 byte (block) striped across physical lanes 0 through 7 to form a Lane Byte

Stream. For 8b/10b encoding in the 8x configuration the start packet delimiters will always be transmitted on physical lane 0, and the end packet delimiters will always be transmitted on physical lane 3 or 7. The link will be padded (PAD control symbols) as necessary to maintain the start of packet alignment rule. For 64b/66b encoding in the 8x configuration the start packet and end packet delimiter may be transmitted on any lane. When SKIP ordered-sets are needed for clock tolerance compensation, they are inserted between packets simultaneously on all physical lanes. For SDR, DDR and QDR the eight symbol streams composed of packets, SKIP ordered-sets, and idle data are individually encoded using the 8b/10b code defined in [Section 5.2](#) above. For higher data rates each Lane Byte Stream is scrambled into a Scrambled Lane Stream and encoded using 64b/66b encoding into a Lane Block Stream as defined in [Section 5.3](#) above. The 8x symbol streams containing data packets, link packets, and SKIP ordered-sets are illustrated in [Figure 50](#) and [Figure 51](#) below.

o5-11.2.1: All 8x ports and 12x ports when configured as a 8x port shall implement packet formatting as defined by [Section 5.7.5, “8x Packet Format.” on page 182](#).

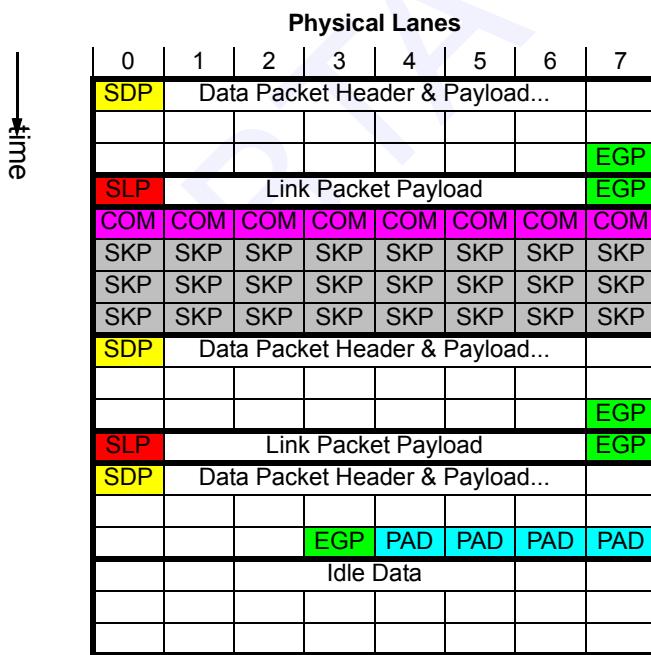


Figure 50 8x Packet Formats - 8b/10b

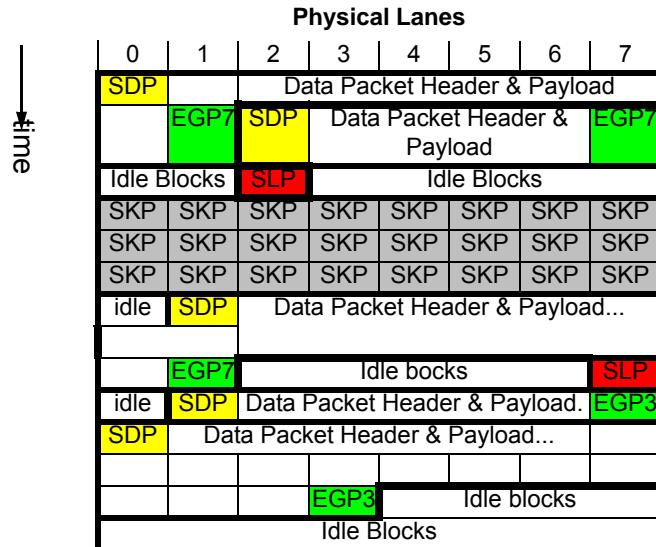


Figure 51 8x Packet Formats - 64b/66b

5.7.5.1 8X PACKET FORMATTING RULES

- 1) For 8b/10b encoding the start of packet delimiters (SDP & SLP) shall be transmitted in lane zero (0) only. For 64b/66b encoding the start of packet delimiter (SDP & SLP) may be transmitted on any lane.
- 2) For 8b/10b encoding the end of packet delimiters (EGP & EBP) shall be transmitted in lane three or seven (3 or 7) only. For 64b/66b encoding the end of packet delimiters (EGP3, EGP7, EBP3 & EBP7) may be transmitted on any lane.
- 3) For 8b/10b encoding when the currently transmitted packet does not end on lane 7, four PAD control symbols shall be inserted to align the link for the next transmit operation.
- 4) SKIP ordered-sets shall be transmitted on all eight (8) lanes simultaneously.
- 5) For 8b/10b encoding when the idle data is placed on the link, the idle data pattern shall be inserted on all eight (8) lanes simultaneously. For 64b/66b encoding Idle blocks may be placed on any lane when there is no packet or ordered set scheduled for transmission
- 6) For 8b/10b encoding the pseudo-random idle data for each lane should start at a different point in the sequence.

5.7.6 12X PACKET FORMAT

The 12x link is composed of twelve physical lanes (0 through 11). For 8b/10b encoding the combined data and link packet symbol stream (control and data) are byte striped

across physical lanes 0 through 11. For 64b/66b encoding the Packet Byte Stream (control and data) are 12 byte (block) striped across physical lanes 0 through 11 to form a Lane Byte Stream. For 8b/10b encoding the start packet delimiters will always be transmitted on physical lane 0, and the end packet delimiters will always be transmitted on physical lane 3, 7, or 11. The link will be padded (PAD control symbols) as necessary to maintain the start of packet alignment rule. For 64b/66b encoding the start of packet and end of packet blocks may be transmitted on any lane. When SKIP ordered-sets are needed for clock tolerance compensation, the link is padded as necessary to allow all twelve SKIP ordered-sets to start on the same symbol/block time. SKIPS are inserted between packets simultaneously on all physical lanes. For SDR, DDR and QDR data rates the twelve symbol streams, composed of packets, SKIP ordered-sets, and idle data are individually encoded using the 8b/10b code defined in [Section 5.2](#) above. For FDR and higher data rates each Lane Byte Stream is scrambled into a Scrambled Lane Stream and encoded using 64b/66b encoding into a Lane Block Stream as defined in [Section 5.3](#). The 12x symbol/Lane Block streams containing data packets, link packet, idle data and SKIP ordered-sets are illustrated in [Figure 52](#) and [Figure 53](#) below.

C5-12: All 12x ports shall implement packet formatting as defined by [Section 5.7.6, “12x Packet Format,” on page 184](#).

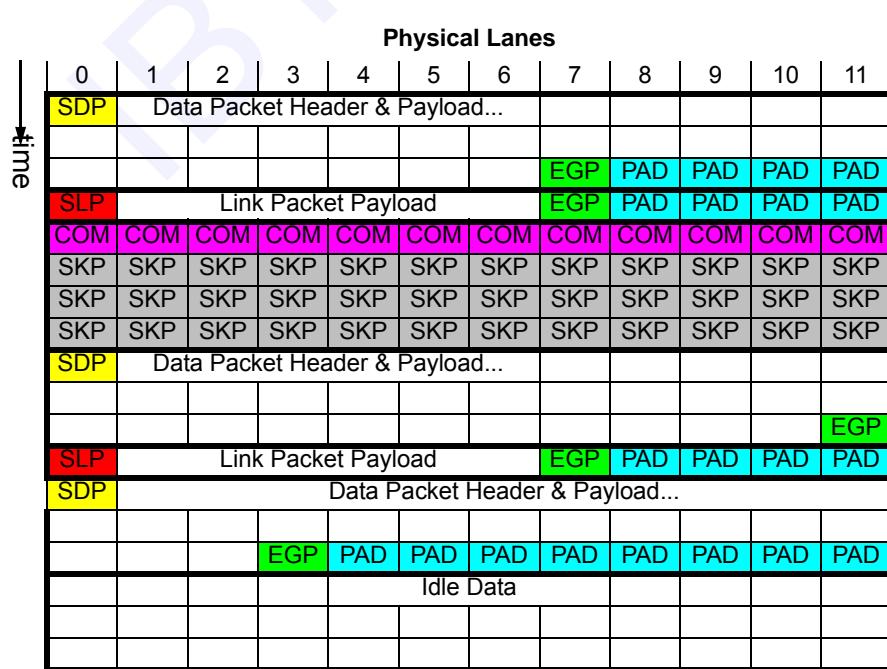


Figure 52 12x Packet Formats - 8b/10b

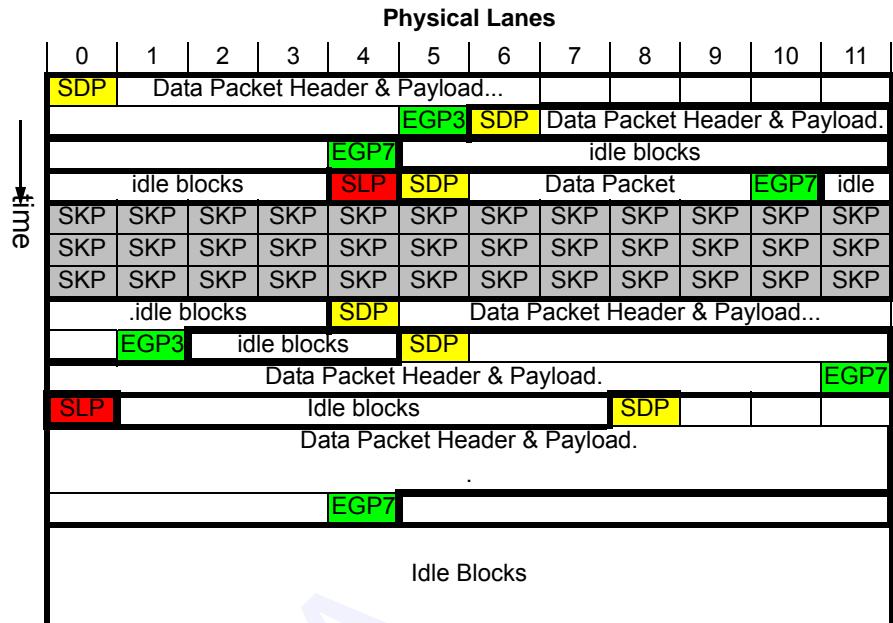


Figure 53 12x Packet Formats - 64b/66b

5.7.6.1 12x PACKET FORMATTING RULES

- 1) For 8b/10b encoding, the start of packet delimiters (SDP & SLP) shall be transmitted in lane zero (0) only. For 64b/66b encoding the start of packet blocks (SDP & SLP) may be transmitted on any lane.
- 2) For 8b/10b encoding, the end of packet delimiters (EGP & EBP) shall be transmitted in lane three, seven, or eleven (3, 7, or 11) only. For 64b/66b encoding the end of packet block (EGP3, EGP7, EBP3 & EBP7) may be transmitted on any lane.
- 3) For 8b/10b encoding, when the currently transmitted packet does not end on lane 11, four or eight PAD control symbols shall be inserted to align the link for the next transmit operation.
- 4) SKIP ordered-sets shall be transmitted on all twelve(12) lanes simultaneously.
- 5) For 8b/10b encoding, when the idle data is placed on the link, the idle data pattern shall be inserted on all twelve(12) lanes simultaneously. For 64b/66b encoding Idle blocks may be placed on any lane when there is no packet or ordered set scheduled for transmission.
- 6) For 8b/10b encoding, the pseudo-random idle data for each lane should start at a different point in the sequence.

5.8 LINK INITIALIZATION AND TRAINING

This section defines the link/physical control process that configures and initializes a link for normal operation. A fully-powered port implements this process by accomplishing the following:

- 1) Waking a remote port on auxiliary power.
- 2) Configuring and initializing the link.
- 3) Supporting normal packet transfers
- 4) Recovering from transient link errors.

Waking a remote port is a special case operation with the local port operating fully-powered and the remote port operating on auxiliary power. The local port polls (see [Section 5.8.4.2, "Polling States," on page 197](#)) the remote port. The remote port detects the presence of this signal and initiates fully-powered operation. In this special case the signal detected by the end operating on auxiliary power is referred to as a beacon. Its use is described further in [Chapter 6: High Speed Electrical Interfaces](#) and in [InfiniBand Architecture Specification, Volume 2-DEPR, Chapter 6: OS Power Management](#).

During the training process, the ports at each end of the link learn each other's capabilities and configure the following parameters:

- 1) Link width (1x, 4x, 8x or 12x).
- 2) Link speed (SDR-2.5 Gb/s, DDR-5.0 Gb/s, QDR-10.0 Gb/s, FDR-14.0625 Gb/s or EDR-25.78125 Gb/s).
 - Note - from the supported speeds a port may learn the peer port's encoding capabilities - a port supporting speeds greater than QDR will support the 64b/66b encoding as well as the 8b/10b encoding.
- 3) Optional correction of lane reversal.
- 4) Optional correction of inverted received serial data (crossed differential signals).
- 5) Forward Error Correction (FEC) enable
- 6) FEC code selection
- 7) Clear Maximum Packet Rate enable.
- 8) Fine Tuning enable

C5-13: This compliance statement is obsolete and has been replaced by [C5-13.2.1](#).

C5-13.2.1: All ports shall implement link initialization and training as defined by [Section 5.8, "Link Initialization and Training," on page 187](#). Ports are not required to implement the lane reversal and serial data inversion options. Implementations that do not support Rel. 1.2 Enhanced Signaling must use the Link Training State Machine described in [Figure 55 Link Training State Machine - Legacy](#), and shall not implement any Enhanced Signaling functions related to TS3 ordered-sets, Heartbeat ordered-sets, Link-RoundTripLatency, TS-T ordered-sets, or Phy Test compliance testing.

o5-13.2.1: All ports that claim compliance with the Rel. 1.2 Enhanced Signaling shall implement link initialization and training as defined by [Section 5.8, “Link Initialization and Training.” on page 187](#) and shall implement all functions related to TS3 ordered-sets, heartbeat ordered-sets, LinkRoundTripLatency measurement, TS-T ordered-sets, or Phy Test compliance testing.

o5-13.2.2: All ports that claim compliance with the Rel. 1.3 shall comply with Rel. 1.2 Enhanced Signaling, and shall, in addition, implement FDR or EDR link initialization and training functions dependent on 64b/66b coding as defined by [Section 5.8, “Link Initialization and Training.” on page 187](#).

o5-1: This compliance statement is obsolete and has been replaced by [o5-13.2.1](#).

o5-13.2.1: Ports that implement the serial data inversion option shall implement the following: [Section 5.8.7.3, “RxCMD = EnConfig,” on page 231](#) rule #2.

o5-2: This compliance statement is obsolete and has been replaced by [o5-13.2.1](#).

o5-13.2.1: All 4x, 8x and 12x ports that implement the lane reversal option shall implement the following: [Section 5.8.4.6.3, “Config.WaitRmt State,” on page 210](#) rule #7 and [Section 5.8.4.6.4, “Config.TxRevLanes State,” on page 211](#) all rules and [Section 5.8.7.3, “RxCMD = EnConfig,” on page 231](#), rule #4.

5.8.1 LINK DE-SKEW TRAINING SEQUENCE AND SKIP ORDERED SETS

The terms link de-skew and training sequence are used throughout this section. Both are briefly explained below. Additionally, the characteristics of the training sequence ordered-sets (TS1, TS2, and TS3) are summarized herein.

Link de-skew: A multi-lane link (4x, 8x and 12x) may have many sources of lane-to-lane skew. These sources include but are not limited to chip I/O drivers and receivers, printed wiring boards, electrical and optical cables, serialization and de-serialization logic, and retiming repeaters. Although symbols are transmitted simultaneously on all lanes, they cannot be expected to arrive at the receiver without lane-to-lane skew. The lane-to-lane skew may include components which are less than a bit time, bit time units (400 ps at SDR, 200 ps at DDR, 100 ps at QDR, 71.11 ps at FDR and 38.79 ps at EDR), or full symbol/block time units (4 ns at SDR, 2 ns at DDR, 1 ns at QDR, 4.69 ns at FDR and 2.56 ns at EDR) of skew caused by the retiming repeaters’ insert/delete operations. A link may have up to two retiming repeaters. Each repeater operates independently on multiple lanes. Because of this independent operation, they may insert or delete a skip symbol/block independently on each lane. Refer to [Section 5.12 on page 241](#) for a complete description of retiming repeaters. The receiving node is required to remove this lane-to-lane skew in order to receive and process data on all lanes simultaneously. This process is called “link de-skew”. Receivers use TS1, TS2, TS3, or SKIP ordered-sets to perform link de-skew functions.

TS1 and TS2 Ordered Sets: The training sequences TS1 and TS2 provide three fundamental types of information at 8b/10b encoding:

1) Lane-to-lane skew:

- At 8b/10b encoding - Information is provided by the unique structure and length of the two training sequences and the fact that they are transmitted without lane-to-lane skew. Both training sequences' (TS1 and TS2) sixteen symbol length (comma and 15 data symbols) allows unambiguous de-skew of up to seven symbol times of skew.
- At 64b/66b encoding - Information is provided by the unique structure and length of the two training sequences and the fact that they are transmitted without lane-to-lane skew. Both training sequences (TS1 and TS2) of 66 bit length allows unambiguous de-skew of up to one block time of skew.
- When RS-FEC is enabled - Information is provided by the unique structure and position of the alignment sequences at the start of every 16th, 32nd, or 48th FEC codeword (x4, x8 and x12, respectively) when alignment is enabled. This allows an unambiguous de-skew of 2 FEC codeword lengths.

Lane Identification: The second symbol of the training sequence at 8b/10b encoding contains the lane number information. The lane number uniquely identifies each of the twelve possible lanes.
2) Lane serial data polarity: The TS1 and TS2 at 8b/10b encoding data symbols contain serial data polarity information. When the bit stream on a lane is inverted, the comma and lane number symbols swap running disparity but still decode to the same value. The TS1 data symbol changes from a D10.2 (4Ah) to a D21.5 (B5h), and the TS2 data symbol changes from a D5.2 (45h) to a D26.5 (BAh) providing a clear indication of lane serial data polarity inversion.

TS3 Ordered Set: The training sequence TS3 provides the same three fundamental types of information provided by TS1 and TS2. In addition, it provides the following types of information:

- 1) Enabled and Supported Link Speeds: Information is provided by a Link/Physical layer about the speed or speeds that the layer supports and is enabled to operate at. This information is provided in the Symbol 8, the SpeedActive symbol, of the sixteen symbols of the TS3 training sequences. During link training, each Link/Physical layer transmits a series of TS3 training sequences to the Link/Physical layer on the other side of its associated link, as part of the link speed auto-negotiation procedure. A link always operates at the highest speed that is supported and enabled on both ends and supported by the physical channel.
- 2) Link Heartbeat Enabled: The Link Heartbeat function is required at data rates greater than SDR, and expected at SDR speed. A port may disable use of the Link Heartbeat function, e.g., for interoperation with legacy devices at SDR speed or working in loopback for testing. Link Heartbeat will only be put into effect when both ports on a negotiation exchange support and enable it.
- 3) TS3 revision: The TS3 revision defines the format of the TS3 and the link training. The revision number is indicated in the two most significant bits of Symbol 12 of the TS3. A value of 00b is used for Rev 0 (specified in Rel. 1.2.1 and older releases of the specification), a value of 01b is used for Rev 1 (specified in Rel. 1.3) and a value of 10b is used for Rev 2 (specification in Rel. 1.3.1).

4) Adaptive Driver De-emphasis Request:

- For Rev 0 TS3: Allows a port to request handshake for a simple adaptive driver de-emphasis setting for equalization, with one of 17 equalization settings: either a default driver de-emphasis setting or one of 16 adaptive driver de-emphasis settings negotiated between transmitter and receiver.
- For Rev 1 and Rev 2 TS3: Allows a port to request a handshake for a single adaptive driver de-emphasis setting for transmitter equalization, with either a preset setting or a coefficient change requested by the receiver.

5) Driver De-emphasis Setting request:

For Rev 0 TS3 - Allows a receiver to transmit to its peer transmitter the requested driver de-emphasis setting out of the 17 possible settings.

For Rev 1 and Rev 2 TS3 - Allows a receiver to transmit to its peer transmitter an equalization request of either a preset as defined in [Table 54 on page 299](#) or a coefficient change as defined in [Section 5.5.2.4](#).

Fields from [6\)](#) through [12\)](#) below are only defined for Rev 1 and Rev 2 TS3.

6) Driver De-emphasis Setting Mode: The DDS Mode chooses the type of the Driver De-emphasis Setting between preset request and coefficient change request. DDS Mode is bit 6 of symbol 10.

- DDS mode 0 - Driver De-emphasis Request is one of the 16 defined presets, as described in [Section 6.6.5.3](#)
- DDS mode 1 - Driver De-emphasis Request is coefficient change request, as described in [Section 6.6.5.4](#).

7) Test Time: Test time is the requested test time by the port, requested time is defined in [Table 41](#). The test time value used is the maximum of the two values requested by the two ports.

8) FEC request: A request to use Forward Error Correction - Fire-Code FEC when using Rev 1 TS3 and RS-FEC when using Rev 2 TS3.

9) Test Pattern: Test pattern is the request to use the PRBS11 test pattern instead of the default PRBS23 pattern. PRBS11 pattern is used if both ports request it.

10) Different polynomials: A request to use different polynomials in each lane, used if both ports request it.

11) Speed Change Time: Speed Change time is a request to extend the FDR/EDR speed change time from 4 ms to 16 ms. The speed change is extended if either side requests extending the time.

12) Max Packet Rate: This field allows for a port to indicate that its receiver is capable of receiving data at a packet rate greater than the max packet rate defined for FDR and higher speeds, as described in [Section 5.15, "Max Packet Rate," on page 255](#).

13) Low Latency FEC Support and Request (LLFS and LLFR): Support of the optional RS(271,257) RS-FEC code and request to use the optional RS(271,257) rather than the mandatory RS(528,514) code.

- 14) Remote Transmit CDR enable Support and Request (RTCS and RTCR): Support of
an optional control of peer port's transmit CDR and request to enable the peer port's
transmit CDR. By default, a port shall request its peer port to enable the transmit
CDR. A port may choose to disable the remote port transmit CDR in order to reduce
the CDR's power consumption. When the Remote Transmit CDR is not supported,
the transmit CDR shall be enabled according to the local port policy.

Table 41 Test Time Values

TT setting	Delay in Config.Test	TT setting	Delay in Config.Test
0	16 ms	8	512 ms
1	32 ms	9	640 ms
2	48 ms	A	768 ms
3	64 ms	B	896 ms
4	100 ms	C	1024 ms
5	128 ms	D	2048 ms
6	256 ms	E	3072 ms
7	384 ms	F	4096 ms

SKIP Ordered Set: The SKIP ordered set may be used for de-skew in 8b/10b. SKIP ordered-set shall be used for de-skew at 64b/66b encoding when RS-FEC is not enabled.

For 64b/66b encoding when RS-FEC is not enabled, the contents of the SKIP block, its length, and the fact that SKIP blocks are transmitted at the same block time on all lanes allow the receiver to perform lane-to-lane deskew. A port shall support de-skew capability of at least 2 block times (132 bits).

5.8.2 LINK INITIALIZATION AND TRAINING OPTIONS

Two independent optional features are defined as part of link initialization and training. Both are intended to allow on chip logic to correct non-optimum pin assignments which may occur when connecting an IBA link chip-to-chip or chip-to-connector. These options can eliminate the two common printed wiring board (PWB) layout issues of crossed differential pair and bus bow ties. Implementing these optional features allows the PWB layout to focus on signal integrity with simplified connections.

5.8.2.1 INVERTED SERIAL DATA CORRECTION (OPTIONAL)

The inverted serial data correction feature allows the receiver logic to correct a receive differential pair that is crossed in the PWB layout. Such a receiver will need to test the polarity of the received training sequences and correct inverted data as part of link configuration. This option does not provide the capability to correct the polarity of transmit data.

To be inter-operable with ports that do not implement inverted serial data correction, the polarity at the transmitting connector must be as specified in the InfiniBand Architecture Specification, [Chapter 7: Electrical Connectors for Modules and Cables](#) and [Volume 2-DEPR, Chapter 2: Backplane Connector Specification](#).

5.8.2.2 LANE REVERSAL CORRECTION (OPTIONAL)

The lane reversal feature allows a PWB layout to connect 4x, 8x or 12x ports in reversed lane number order. The receiver logic uses the lane number symbol in the training sequences (TS1 or TS2) at 8b/10b encoding to detect and correct the reversed lane connections in the PWB layout. When the remote port is incapable of correcting receive lane reversal (link width is less than the local port or the port does not implement this option) the local port will reverse its transmit lane as part of its configuration process. (see [Section 5.8.4.6, “Configuration States - Enhanced Signaling,” on page 203](#))

5.8.3 INTERACTIONS WITH OTHER ENTITIES

Link training is an involved process that requires interactions with other entities. This section briefly describes these interfaces and refers the reader to other sections for more detailed information. [Figure 54](#) illustrates all entities that interact with the link training state machine in the link initialization and configuration process.

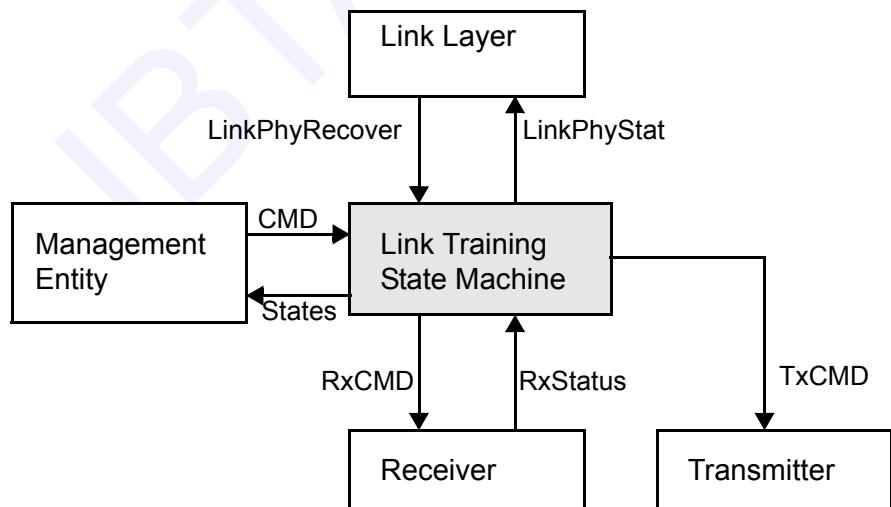


Figure 54 Link Training State Machine Interactions

The interface between the link layer and the Link/Phy layer is comprised of the signals LinkPhyRecover and LinkPhyStat. These logical signals are described in [Section 5.5.4, “Logical Interfaces,” on page 164](#).

The link training state machine and the unit’s management entity communicate over a set of logical signals. Although this logical interface is not defined completely in this specification, functions that need to be carried out over this interface are defined. The manage-

ment messages related to the link training state machine operations are described in [Section 5.6, “Management Datagram Control and Status Interface,” on page 165.](#)

The link training state machine also interfaces to the transmitter and receiver logic to coordinate the link initialization and configuration operations. The interface to the transmitter is in the form of commands (TxCMD) for the transmitter to perform. Similarly, the receiver interface has a set of commands (RxCMD) or configurations and a set of receiver status (RxStatus) values.

These commands and status conditions are listed in [Table 42](#) below and are described in detail in [Section 5.8.6](#) and [Section 5.8.7](#). The receiver status conditions (RxStatus) are not mutually exclusive conditions.

Table 42 Transmitter and Receiver Interface

Transmitter	Speed	Receiver	
TxCMD		RxCMD	Status (RxStatus)
Disable		Disable	
SendTS1, SendTS2, SendTS3	SPEED = MinEnabledSpeed, SPEED = MaxBothActive	WaitTS1, WaitTS2, WaitTS3	RcvdTS1, RcvdTS2, RcvdTS3
RevLanes		EnConfig, EnDeSkew	RxTrained
SendIdle		WaitIdle	RcvdIdle
Enable		Enable	RxMajorError, RxHeartbeatError
		WaitTS-T	RcvdTS-T
SendPRBS23	SPEED = MaxBothActive	WaitPRBS23	RcvdPRBS
SendPRBS11	SPEED = MaxBothActive	WaitPRBS11	RcvdPRBS

5.8.4 LINK TRAINING STATE MACHINE

Link initialization, configuration, and link error recovery operations are performed by the link training state machine. This state machine is described in this section in the form of hierarchical states. As depicted in [Figure 56](#), the link training state machine has seven primary states. Some of these states are super states composed of two or more states.

The link training state machine has the following primary states:

- 1) **Disabled:** Port drives its outputs to quiescent levels and does not respond to received data.
- 2) **Sleeping:** In this super state, the port drives its outputs to quiescent levels and responds to received training sequences.
- 3) **Polling:** In this super state, the port transmits training sequences and responds to received training sequences. This is the default state following power on.

- 4) **Configuration:** A transient super state with both the transmitter and receiver active.
The port is attempting to configure and transition to the LinkUp state.
- 5) **LinkUp:** This is the normal link operation state. The port is available to transfer packets.
- 6) **Recovery:** The recovery super state attempts to re-synchronize the link and return it to normal operation. This state is entered when a port experiences loss of link synchronization, a major error, or when a link layer error triggers error recovery.
- 7) **Phy Test:** The Physical Layer test state allows simplified testing to determine compliance of physical layer transmitter driver and receiver circuitry with specified requirements. This state is defined for Enhanced Signaling, and is not implemented in products not supporting Enhanced Signaling.

These primary states directly map to the enumerated port states defined in [Section 5.6, "Management Datagram Control and Status Interface," on page 165](#).

The status of the link training state machine is also reported to the link layer in a more abstract fashion. The “LinkPhyStat” logical signal over this interface provides the values “Up” and “Down”. This logical signal is given the “Up” value when this state machine is in the LinkUp state. It is given the “Down” value in all other states.

The initial or post-reset state of the link training state machine is Polling. This insures that a newly connected and powered on port will be recognized by any remote port. Management commands can force the link training state machine to the following states: Disabled, Polling, or Sleeping. Management commands also set a default state for link down. All other state transitions are under the control of the link training state machine. Test equipment can control the transition in and out of the Phy Test state by injecting patterns in the port’s receiver.

[Figure 55](#) shows the Link Training State Machine for products not supporting Enhanced Signaling. [Figure 56](#) shows the Link Training State Machine for products that support Enhanced Signaling, which adds the Phy Test state, and additional substates within the Configuration super state.

In [Figure 56](#), all state transition events are based on management commands, delay events, or receiver status. These receiver status conditions are described later in [Section 5.8.7 on page 230](#).

Operations in certain states involve delays or time-out periods. These events are denoted in the state diagrams as state transitions labeled with DelayTimeOut (time). These time-out periods start as the state is entered and cleared when a transition to a different state is taken. Time-out periods cannot be re-started within a state. (See [Section 5.8.5 on page 225](#).)

Note: This is the legacy state diagram, for products that do not support Enhanced Signaling. It has been modified for Enhanced Signaling as shown in the diagram in [Figure 56](#).

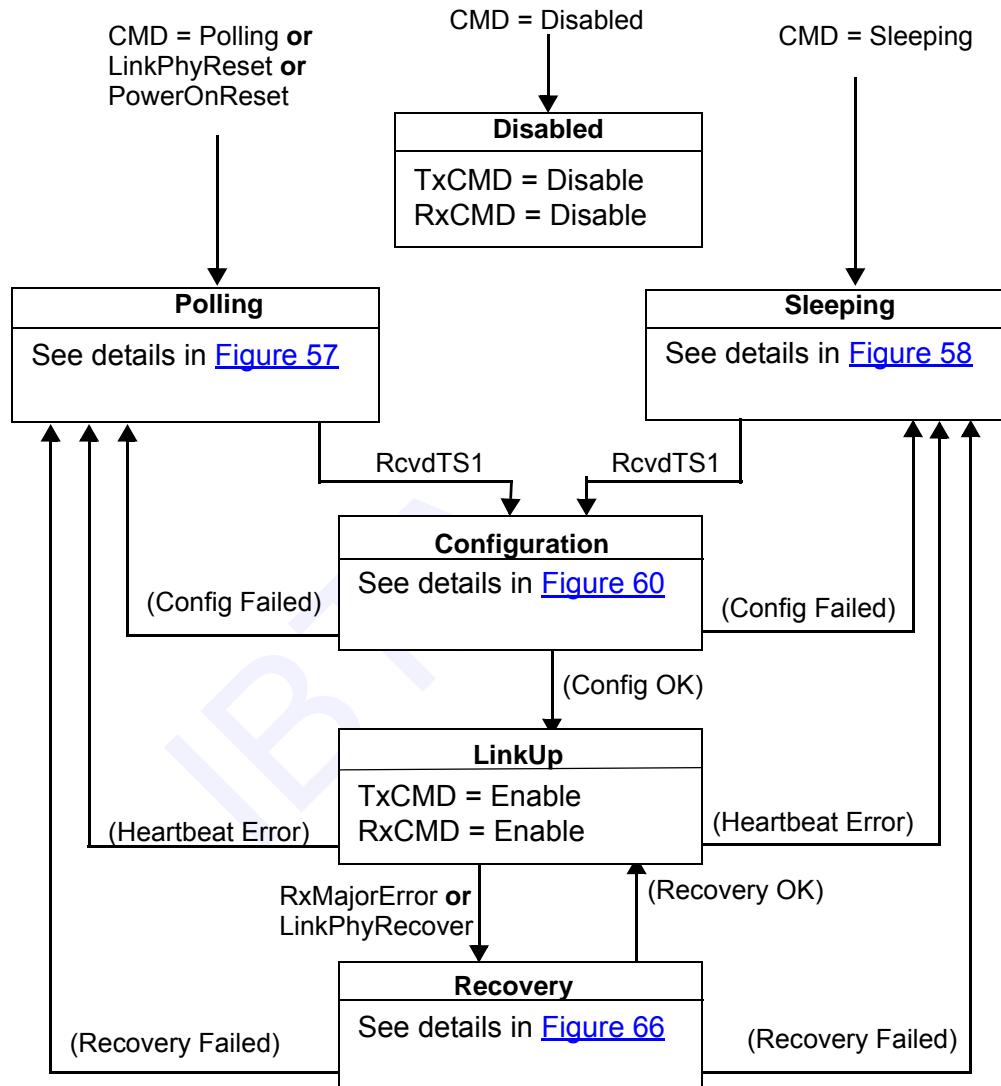


Figure 55 Link Training State Machine - Legacy

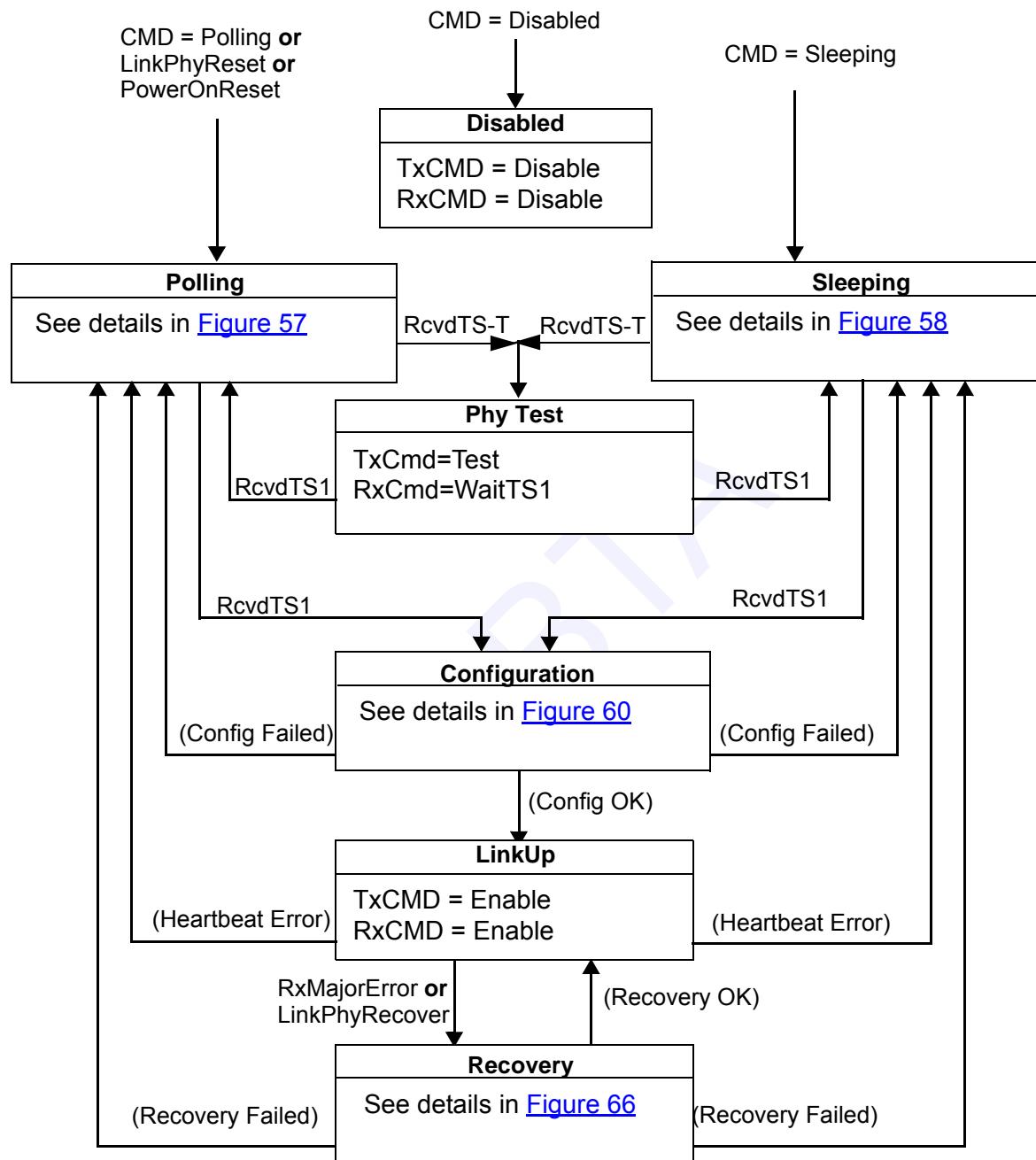


Figure 56 Link Training State Machine - Enhanced Signaling

5.8.4.1 DISABLED STATE

As the state name implies, there is no physical layer activity in this state, and both the transmitter and receiver are disabled. The transmitter outputs are forced to a quiescent condition, and receiver inputs are ignored. Entrance to and exit from the Disabled state are controlled by management commands. In the Port Disabled State the following rules and commands shall be active:

- 1) TxCMD = Disable
- 2) RxCMD = Disable
- 3) LinkPhyStat = Down
- 4) **SymbolErrorCounter** is inhibited

5.8.4.2 POLLING STATES

The Polling States define the polling process used to initiate link configuration. The link training state machine cycles between the two states. These states are transmitting repeated TS1(s) for 2 ms in Polling.Active followed by 100 ms of a quiescent output in Polling.Quiet. In both states, the port's receiver is configured to detect TS1 ordered-sets. When TS1 is received, the link training state machine transfers control to the Configuration Super State. The states of the Polling Super State are expanded in the state graph shown in [Figure 57](#).

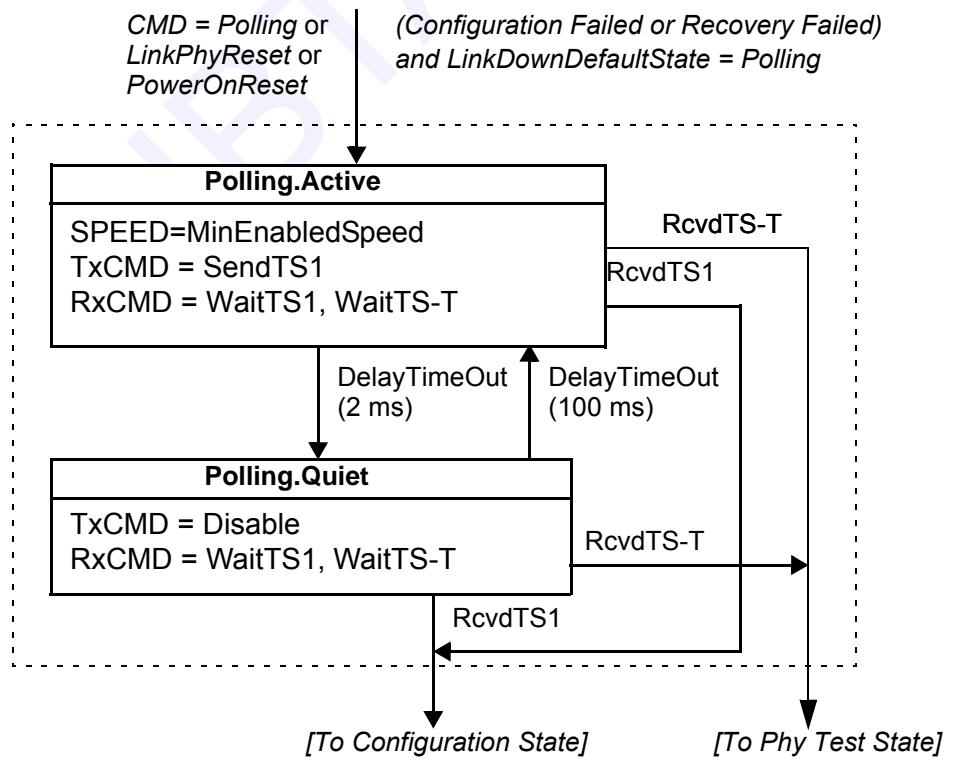


Figure 57 Polling Super State (Expanded)

5.8.4.2.1 POLLING.ACTIVE STATE

The Polling.Active state sends a stream of TS1s and at the same time enables detection of TS1s on all receiver lanes. The transmitter speed is set to the minimum enabled speed. In the Polling.Active state the following rules and commands shall be active:

- 1) SPEED = MinEnabledSpeed
- 2) TxCMD = SendTS1
- 3) RxCMD = WaitTS1, WaitTS-T
- 4) LinkPhyStat = Down
- 5) **SymbolErrorCounter** is inhibited.
- 6) **LinkSpeedActive = LinkSpeedEnabled**
- 7) **LinkSpeedExtActive = LinkSpeedExtEnabled**
- 8) TS1 and SKIP ordered-sets are transmitted on all lanes.
- 9) If RxStatus = RcvdTS1, the next state is Configuration.
- 10) If RxStatus = RcvdTS-T, the next state is Phy Test
- 11) Else if DelayTimeOut(2 ms), the next state is Polling.Quiet

5.8.4.2.2 POLLING.QUIET STATE

The Polling.Quiet state inhibits the transmitter and continues to enable detection of TS1 on all receiver lanes. When TS1 is detected, configuration can be started. In the Polling.Quiet state, the following rules and commands shall be active:

- 1) TxCMD = Disable
- 2) RxCMD = WaitTS1, WaitTS-T
- 3) LinkPhyStat = Down
- 4) **SymbolErrorCounter** is inhibited
- 5) If RxStatus = RcvdTS1, the next state is Configuration
- 6) If RxStatus = RcvdTS-T, the next state is Phy Test
- 7) Else If DelayTimeOut(100 ms), the next state is Polling.Active

5.8.4.3 SLEEPING STATES

The Sleeping states are used to deactivate a link without powering off the port. When a port enters the sleeping state, it first disables its transmitter and receiver and then delays, allowing all link activity to cease. Following the delay, the receiver is enabled to detect TS1 ordered-sets. When a TS1 is detected, port configuration is started. The states of the Sleeping Super State are expanded in the state graph shown in [Figure 58](#).

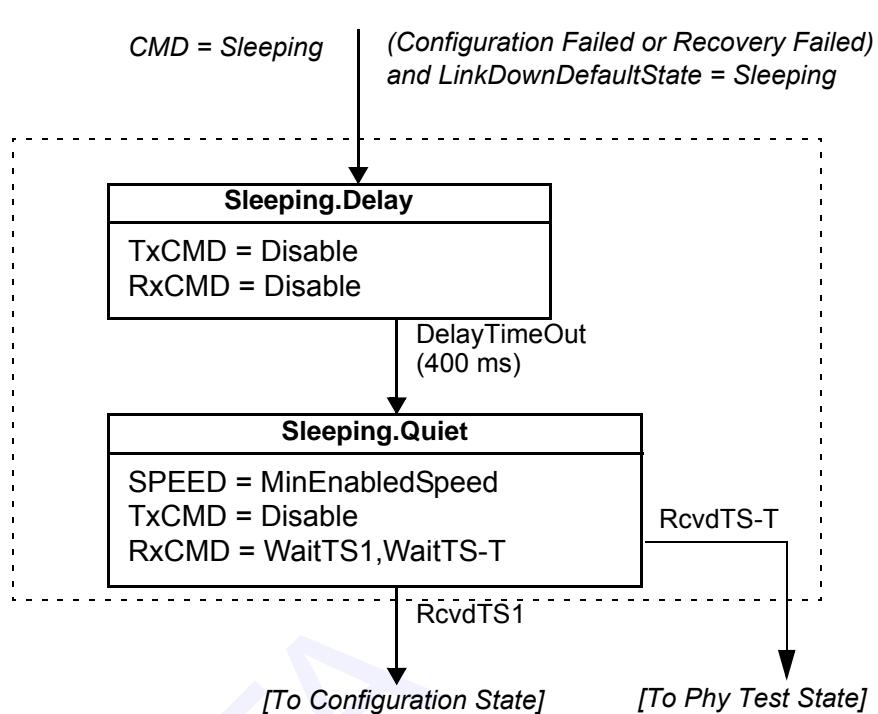


Figure 58 Sleeping Super State (Expanded)

5.8.4.3.1 SLEEPING.DELAY STATE

In the Sleeping Delay (Sleeping.Delay) state the transmitter is quiescent, and the receiver is disabled. The next state following the delay is Sleeping.Quiet. In the Sleeping.Delay state the following rules and commands shall be active:

- 1) TxCMD = Disable
- 2) RxCMD = Disable
- 3) LinkPhyStat = Down
- 4) **SymbolErrorCounter** is inhibited
- 5) If DelayTimeOut(400 ms), the next state is Sleeping.Quiet

5.8.4.3.2 SLEEPING.QUIET STATE

The Sleeping.Quiet state inhibits the transmitter and at the same time enables detection of TS1 on all receiver lanes. The transmitter speed is set to the minimum enabled speed for when transmission resumes. When TS1 is detected, configuration can be started. In the Sleeping.Quiet state, the following rules and commands shall be active:

- 1) SPEED = MinEnabledSpeed
- 2) TxCMD = Disable
- 3) RxCMD = WaitTS1, WaitTS-T

-
- | | |
|---|---|
| 4) LinkPhyStat = Down | 1 |
| 5) SymbolErrorCounter is inhibited | 2 |
| 6) LinkSpeedActive = LinkSpeedEnabled | 3 |
| 7) LinkSpeedExtActive = LinkSpeedExtEnabled | 4 |
| 8) If RxStatus = RcvdTS1, the next state is Configuration | 5 |
| 9) If RxStatus = RcvdTS-T, the next state is Phy Test | 6 |

5.8.4.4 PHY TEST STATE

The Phy Test state allows the transmitter and receiver circuitry to be tested by external test equipment for compliance with transmitter and receiver specifications. This state is not used for normal operations. Operation of this state is described further in [Section 5.17, “Physical Layer Compliance Testing,” on page 256](#). Entry to this state occurs on receipt of a TS-T (Training Sequence for Test) ordered-set. In the Phy Test state, the following rules and commands shall be active:

- | | |
|--|----|
| 1) SPEED = Value determined in TS-T ordered-set | 15 |
| 2) For QDR and lower bit rates, RxCMD = WaitTS1 | 16 |
| 3) Transmitted signals will follow rules defined in Section 5.17, “Physical Layer Compliance Testing,” on page 256 . | 17 |
| 4) For QDR and lower bit rates, if RxStatus = RcvdTS1, the next state is the Link-DownDefaultState (Polling or Sleeping). | 18 |

5.8.4.5 CONFIGURATION STATES - LEGACY OPERATION

The Configuration Super State controls the configuration and training of the link. A low level protocol using TS1, TS2, and Idle Data is used to communicate the state of the two ports on the link.

The first step is to send a extended stream of TS1 ordered-sets indicating that the port has started the configuration process. This allows time for the physical connection to stabilize (debounce delay).

Following the delay, the second step of configuration begins by enabling the receivers to auto-configure and de-skew. The remote port may have completed receiver training and be transmitting TS2 ordered-sets. The receiver can auto-configure and de-skew while receiving TS1 or TS2 ordered-sets. The transmitter continues to send a stream of TS1 ordered-sets and waits for the receiver to complete configuration and de-skew (training).

The third step starts when the receiver has completed training, and the transmitter begins to send a stream of TS2 ordered-sets indicating that the local receiver is trained. A delay is started, and the port waits for the remote port to indicate that its receiver has completed training.

At the end of the delay, if the remote port has not indicated that it is trained, a port implementing the optional lane reversal will reverse its transmit lanes, start a short delay, and

1 continue to wait for the remote port to report that its receiver has completed training. If
2 both ports are not trained at the end of the delay(s), configuration has failed. If the port
3 completes training before the end of the delay(s), the final step of configuration starts
4 when the port is both sending and receiving TS2 ordered-sets.

5 The final step of link configuration transmits the idle data stream and waits to receive idle
6 data. When the port is both sending idle data and receiving idle data, configuration is
7 complete, and the link is up. Receiving TS1 in this state will force the restart of link con-
8 figuration.

9 The states of Configuration Super State are expanded in the state graph shown in [Figure](#)
10 [59](#).

Note: This is the legacy state diagram, for products that do not support Enhanced Signaling. It has been modified for Enhanced Signaling as shown in the diagram in [Figure 60](#).

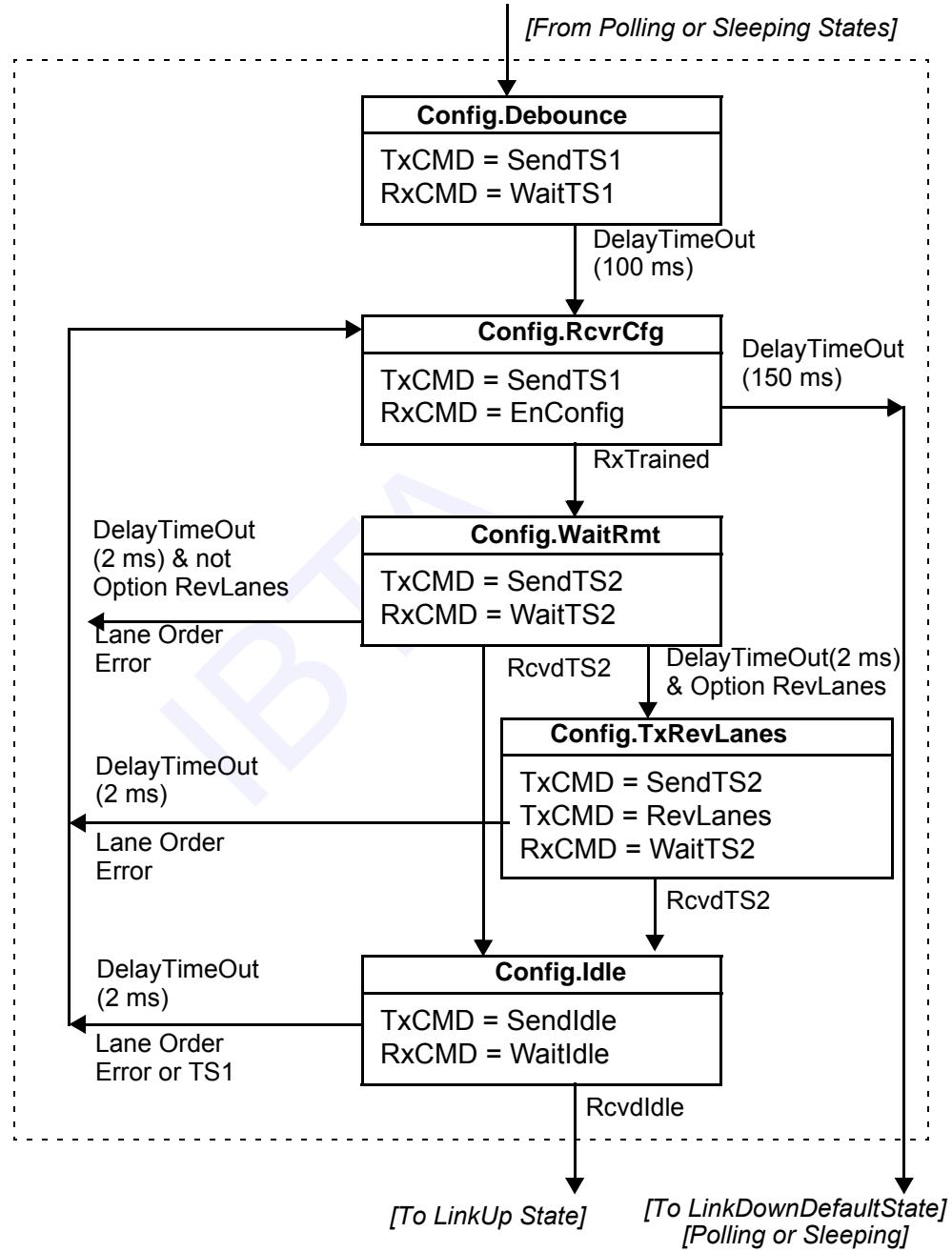


Figure 59 Configuration Super State (Expanded) - Legacy

5.8.4.6 CONFIGURATION STATES - ENHANCED SIGNALING

In this release of this specification, additional functions are added to the Configuration states to support enhanced signaling capabilities of the links, including higher speed operation at DDR (5.0 Gb/s), QDR (10.0 Gb/s), FDR (14.0625Gb/s) and EDR (25.78125Gb/s) bit rates, Link Heartbeat function (described in [Section 5.14, "Link Heartbeat," on page 253](#)), and support for transmitter and receiver equalization to better enable operation at higher bit rates over a variety of media. Several additional states and an additional training sequence (TS3) are added for the purpose of negotiating common use of these enhanced capabilities between the two peer ports on a link.

The first three steps of link initialization are the same as for legacy devices: a TS1 exchange assures that both transmitters are active, a TS2 exchange assures that both receivers are trained, and lane reversal assures correct polarity at each end of the link.

Following these steps, both ports transmit a TS3 training-set, indicating which enhanced capabilities they are enabled to support (DDR/QDR/FDR/EDR, Equalization, FEC, MPR, Fine Tuning, and Link Heartbeat). Each transmitter compares its own capabilities with the capabilities indicated in the TS3 received from the other end of the link. When both sides have sent and received TS3s, indicating agreement on common enhanced capabilities, they act to enable those capabilities. These negotiations occur at the lowest enabled speed, which will generally be the SDR speed.

The TS3 exchange for negotiating link speed operates by having each port send a bit map of the active speeds, in the format used in the **SM.PortInfo(LinkSpeedEnabled)** or the **SM.PortInfo(LinkSpeedExtEnabled)**, and **SM.PortInfo(LinkSpeedSupported)** or the **SM.PortInfo(LinkSpeedExtSupported)** MADs. The first time the ports traverse the states negotiating enhanced capabilities, they send TS3 ordered-sets indicating the link speeds enabled by the **SM.PortInfo(LinkSpeedEnabled)** and **SM.PortInfo(LinkSpeedExtEnabled)**. If, however, a particular speed is attempted, and found to be unsupportable (for example, link attenuation is too high for intact data transmission), the bit corresponding to this speed is cleared to 0 in the **LinkSpeedActive** or in the **LinkSpeedExtActive** field, and in the following TS3 ordered-sets. A logical AND operation on the transmitted and received fields is used to determine the speed(s) supported on both ports, and further operation is attempted at the highest speed that is active on both.

For ports supporting TS3 revision 2 or higher, when the highest received active speed differs from the highest transmitted active speed the ports shall skip the Config.Test state and transition directly to the Config.WaitRmtTest state. This allows the ports to traverse the Config.CfgEnhanced again, adjust the negotiated training parameters to the highest speed active on both ports before entering the Test state. Re-negotiating the training parameters based on the test speed enables to match parameters such as driver de-emphasis, test pattern and test time to tested speed.

A link pair will go through TS3 exchange at least twice, to ensure that the supported capabilities on both ends of the link are consistent, and supported by the link media. Links will always be *Active* at the highest speed which is *Enabled* in both ports driving the link, and supported by the link transmission medium. During the negotiation phase of the con-

figuration super state, multiple speeds bits may be set to 1 in the ***LinkSpeedActive*** and ***LinkSpeedExtActive*** bit map exchanged in TS3 ordered-sets, but once consistent and correct operation has been established, only one ***LinkSpeedActive*** and ***LinkSpeedExtActive*** bit will be set to 1. Links always operate at the same speed in both directions, and a port always configures its transmitter and receiver to operate at the same speed.

The TS3 exchange is also used to establish whether adaptive driver de-emphasis equalization training is requested. If either side requested adaptive equalization, both sides will participate.

At QDR and lower bit rates for equalization training, each transmitter sends idle data at the maximum speed commonly active in both ports (typically DDR or QDR, but possibly SDR as well), for a 100 millisecond period. The 100 ms period is split into a 36 ms period where data is sent with a default transmitter equalization setting, for receivers which don't support adaptive equalization, and 16 separate adaptive transmitter equalization setting slots, of 4 ms duration each. On each 4 ms slot, the transmitter sends data using one of a set of vendor-dependent pre-emphasis settings. The receiver, on each slot, attempts to receive intact data, optionally adjusting a receiver equalization setting. The receiver determines which is the first slot that provides a signal good enough to be equalized at the receiver.

At bit rates higher than QDR for equalization training, each transmitter sends PRBS23 or PRBS11 at the maximum speed commonly active in both ports (FDR or EDR), for a pre-negotiated time. The generation of the transmitted PRBS is described in [Section 5.8.4.6.6.14. "PRBS Test Pattern." on page 217](#). The pre-negotiated time is achieved by choosing the highest Test Time (TT) advertised in the TS3 by both ports, the test encoding is described in [Table 41 on page 191](#). The receiver should perform a receiver test to determine if the current transmitter setting is optimal and sufficient for the tested speed, the receiver test, equalization and decision protocol is outside the scope of this specification. The receiver should allow for stabilization time at the start and end of the test time. At the start a receiver shall allow for at least 4 ms stabilization time when [SCTH,SCT]=[0,0] and 16, 32, or 64 ms, when [SCTH,SCT]=01, 10, or 11, respectively. At the end of the test time the receiver stabilization time is undefined but it is recommended to use a 4 ms stabilization time. The stabilization time shall not be used by the receiver for receiver testing.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

Implementation Note

On definition of the terms *Enabled*, *Supported*, and *Active*, in the context of describing link widths and speeds:

The set of *Enabled* widths or speeds is determined by the subnet manager, and will be equal to the set of or a subset of the widths or speeds *Supported* by the design and implementation of the port. The *Active* width or speed is determined by a combination of the implemented capabilities in the two ports of a link, and by the link used to connect them. For example, a pair of ports sharing a link may both *Support* and be *Enabled* for SDR/DDR/QDR operation (7), but the cable over which they are connected may only allow intact data transmission at SDR speed. In this case, the port Link/Phy logic would set the **Link-SpeedActive** value to 1 on both ports. The actual operating link speed or link width, as determined by the Link/Phy logic during the Configuration super state, is reflected in the *Active* fields.

Implementation Note:

Equalization is used to compensate for signal distortion (attenuation or dispersion) that occurs in the medium between transmitter and receiver. The distortion being compensated for may take a variety of forms, including attenuation of high-frequency components, frequency-dependent crosstalk, or frequency-dependent reflections. Signal-processing circuitry in the transmitter, or the receiver, or (typically) both, can equalize out some of the effect of some distortions. This equalization may, for example, take the form of emphasizing high frequencies (fast toggles) or de-emphasizing low frequencies (strings of 1s or 0s), or of de-emphasizing high frequencies, in cases where frequency-dependent crosstalk is the major distortion. The equalization circuitry should therefore be adaptive to the characteristics of the particular link medium.

Achieving ideal adaptive equalization over different backplanes and over the wide variety of copper and optical media of the various distances used for InfiniBand™ links is a very complex problem. The algorithm described here for QDR and lower speeds (16 different adaptive driver de-emphasis settings, which the receiver selects from, after attempting adaptive receiver equalization on each), and the algorithm for FDR and EDR speeds (combination of Preset setting options, High- and Low-Amplitude ranges, and incremental changes to the C₋₁, C₀, and C₊₁ tap weights during link initialization, as well the Fine Tuning Algorithm for incremental adjustment while the link is operational) provides mechanisms for both transmitter and receiver equalization with moderate complexity. The algorithms also allow interoperability with ports that don't implement adaptive equalization.

Since the maximum physical length of an InfiniBand link at Rel. 1.3, using single mode optical fiber links, is 10 km, with speed-of-light round-trip latency on the order of 100 microseconds (0.1 ms), the two ends of a link may be un-synchronized with each other by this amount. This synchronization period determines, along with the (-1us/ +126us) clock tolerance and 100 ppm variation in clock frequency described in [Section 5.8.5, “State Machine Delays and Timeouts,” on page 225](#), the window of proper transmitter-equalized data within the 4 ms slot period, and must be taken into account at the receiver when determining transmitter equalization setting slots.

Following attempted transmission at the maximum speed that is active on both ports, including transmitter and receiver adaptive equalization, the ports then transition through a Config.WaitRmtTest state, to ensure that both ports have completed attempting operation at the maximum active speed, and then do another round of TS3 exchange.

For QDR and lower speeds: this round allows both ports to indicate whether they were able to receive intact data at the attempted speed, and if so, to allow each port's receiver to indicate to its peer port's transmitter which of the 16 4 ms slots was the first that had a transmitter equalization setting with good enough signal for receiver data recovery. The use of the first good slot allows the transmitter/receiver pair to optimize the link's overall power usage.

For FDR and EDR speeds: This round allows both ports to either request another round through test with the same or different transmitter settings or to indicate the optimal transmitter setting has been found during the equalization procedure.

The TS3 exchange also establishes the use of Link Heartbeat. Link Heartbeat is used (a) to determine that both receivers on a link are receiving data from their peer transmitters rather than from crosstalk, and (b) to determine the link round-trip latency for data to traverse the link in both directions. Link Heartbeat is required at DDR and higher speeds, due to the small received signal swings and lack of a signal detect capability at the higher bit rates. Link Heartbeat is also useful at SDR speed, and is required for devices supporting Enhanced Signaling, even if they only support SDR. A bit in the exchanged TS3 indicates that Link Heartbeat is requested. If both ports request the use of Link Heartbeat, it is used when the link transitions to the LinkUp state. Further details on Link Heartbeat may be found in [Section 5.14, "Link Heartbeat," on page 253](#).

The TS3 exchange establishes Forward Error Correction (FEC) negotiation. FEC is used to reduce the bit error ratio of a given media. A bit in the exchanged TS3 indicates that FEC is requested. If either port requests the use of FEC, it is used when the link transitions to the Config.WaitCfgEnhanced state. The FEC code is determined by both the TS3 revision and the LLFS and LLFR bits. Further details on FEC may be found in [Section 5.4, "Forward Error Correction," on page 106](#).

The TS3 exchange establishes Fine Tuning (FT) algorithm negotiation support. Fine Tuning is used to allow fine grain equalization adjustments while the LinkPhy is in the LinkUp state. A bit in the exchanged TS3 indicates that FT is supported. If both ports support the FT then FT_en is set to true and the port may use the FT algorithm when in LinkUp state. Further details on the FT algorithm may be found in [Section 5.13, "Fine Tuning," on page 246](#).

The TS3 exchange also allows a receiver to report ability to receive packets at a rate higher than the Max Packet Rate (MPR), as defined in [Section 5.15](#). A bit in the exchanged TS3 indicates that disabling of the MPR is supported. A port receiving TS3 with MPR bit set may transmit packets at a rate higher than 1 packet per 64 Bytes (8 blocks), in LinkUp state. Further details on the MPR may be found in [Section 5.15, "Max Packet Rate," on page 255](#).

Once both ports on a link have exchanged consistent TS3s, and assured correct operation at the highest commonly-*Active* speed, they will transition to the Config.WaitCfgEnhanced state, to indicate completion of configuration for enhanced operation, and from there to Config.Idle. If the FEC is enabled it will be turned on when entering Config.WaitCfgEnhanced.

On connection with a port that does not support Enhanced Signaling, or during legacy operation, the port state machine would traverse the states directly from top to bottom, at SDR rate, bypassing the Config.Test, Config.WaitRmtTest, and Config.WaitCfgEnhanced states.

The states of Configuration Super State for ports supporting [Section 5.8.4.6](#) operation are expanded in the state diagram shown in [Figure 60](#). A simplified diagram of the State graph is shown in [Figure 61](#), along with a typical control flow through the states, showing negotiation from SDR up to higher speed operation.

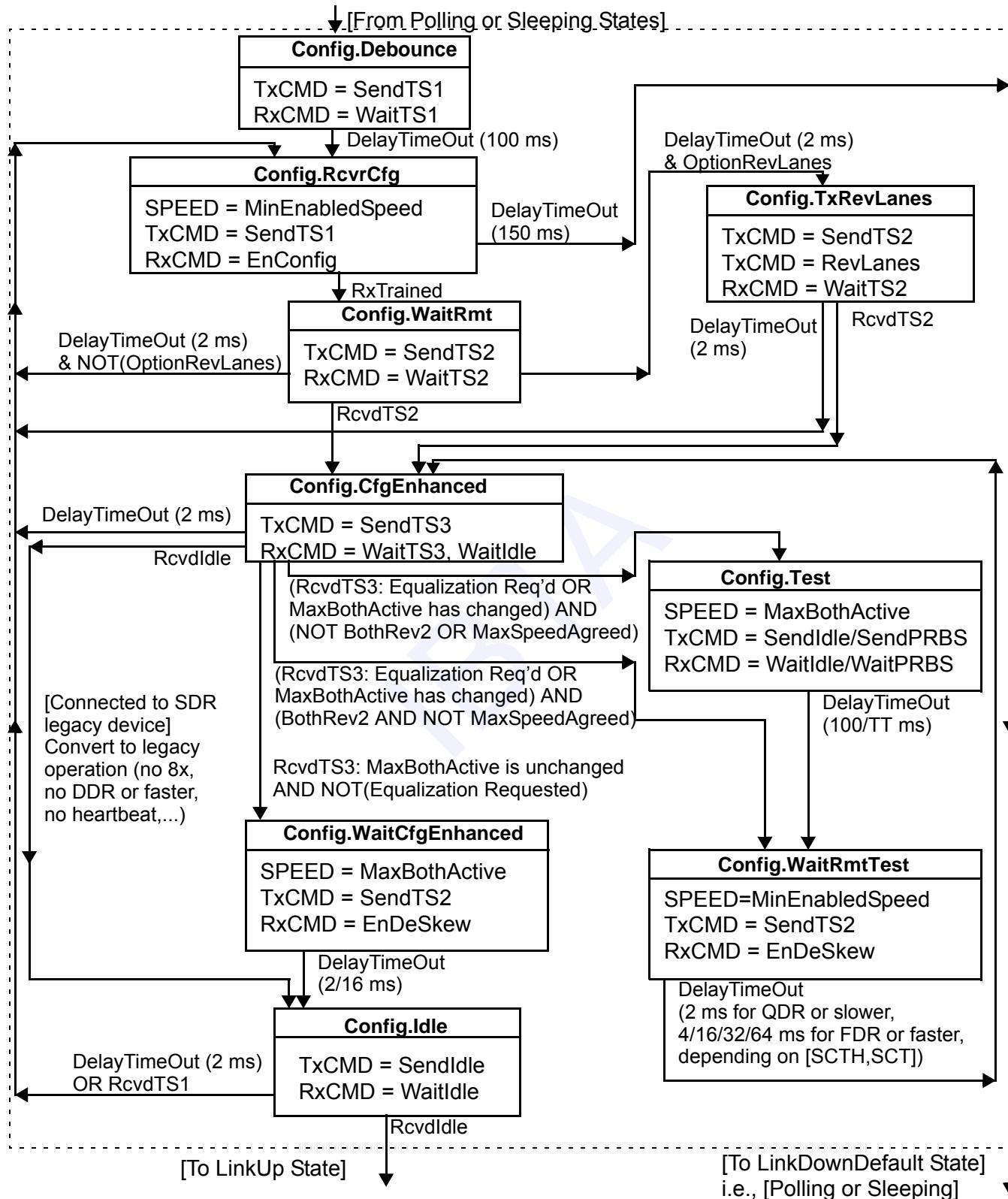


Figure 60 Configuration Super State (Expanded) - Enhanced Signaling

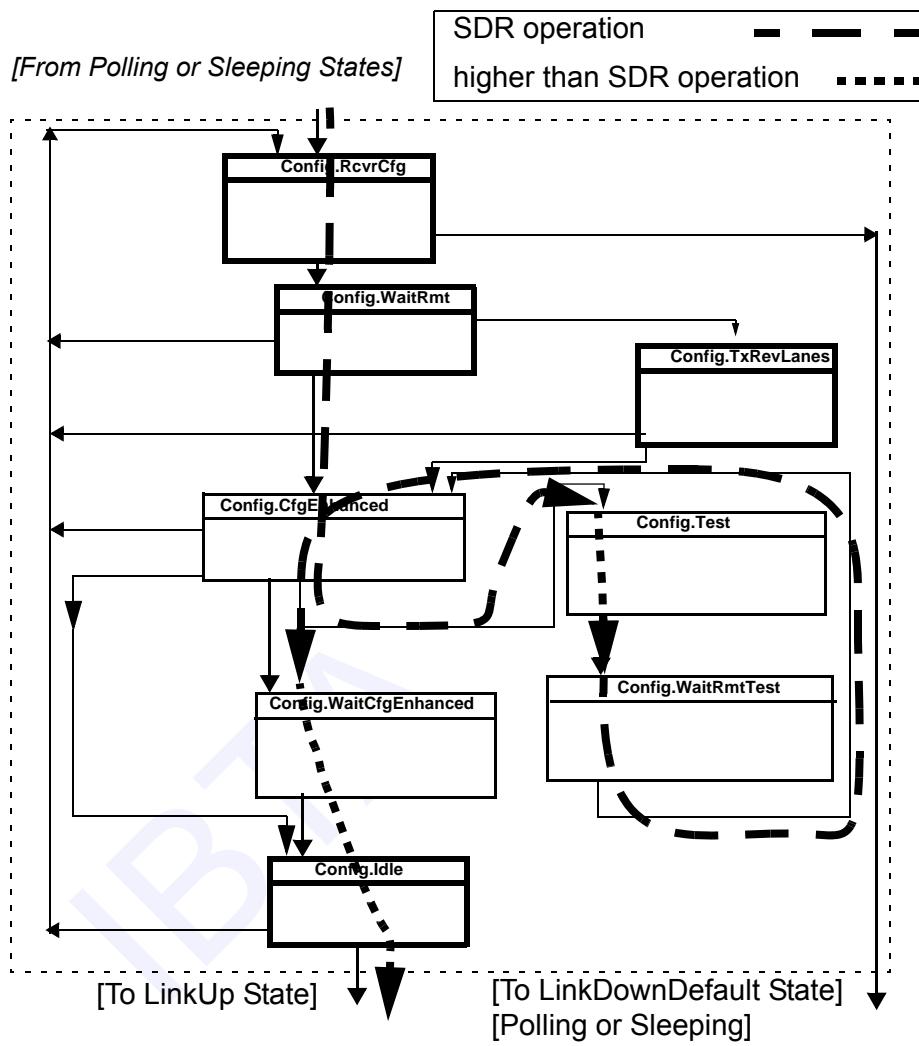


Figure 61 Typical Control Flow through Configuration Super State - Enhanced Signaling

5.8.4.6.1 CONFIG.DEBOUNCE STATE

In the debounce state (Config.Debounce), the transmitter sends a series of TS1 ordered-sets on all lanes. The receiver status is not checked in this state. In receivers with adaptive receiver equalization, the 100 ms duration of this state may optionally be used for determining receiver equalization settings. Operation in Config.Debounce is always at SDR speed. In the Config.Debounce state, the following rules and commands shall be active:

- 1) TxCMD = SendTS1
- 2) RxCMD = WaitTS1
- 3) LinkPhyStat = Down
- 4) TS1 and SKIP ordered-sets are transmitted on all lanes
- 5) **SymbolErrorCounter** is inhibited
- 6) If DelayTimeOut(100 ms), next state is Config.RcvrCfg.

5.8.4.6.2 CONFIG.RCVRCFG STATE

In the Receiver Configure (Config.RcvrCfg) state, the transmitter sends a stream of TS1 ordered-sets on all lanes. The receiver uses received TS1 or TS2 ordered-sets to identify the width of the remote port and to optionally correct lane reversal and inverted lane data. When the receiver has completed configuration and training, it reports Receiver Trained status. The transmitter speed is set to the minimum enabled speed. In the Config.RcvrCfg state, the following rules and commands shall be active:

- 1) SPEED = MinEnabledSpeed
- 2) TxCMD = SendTS1
- 3) RxCMD = EnConfig
- 4) LinkPhyStat = Down
- 5) TS1 and SKIP ordered-sets are transmitted on all lanes
- 6) **SymbolErrorCounter** is enabled
- 7) **LinkSpeedActive = LinkSpeedEnabled**
- 8) **LinkSpeedExtActive = LinkSpeedExtEnabled**
- 9) If the RxStatus = RxTrained, the next state is Config.WaitRmt.
- 10) Else if DelayTimeOut (150 ms) the next state is the **LinkDownDefaultState** (Polling or Sleeping).

5.8.4.6.3 CONFIG.WAITRMT STATE

In the Wait Remote (Config.WaitRmt) state, the transmitter sends a series of TS2 ordered-sets on all lanes. The receiver monitors the configured physical lanes waiting for a transition from TS1 to TS2 ordered-sets indicating the remote port has completed its configuration. If the lane reversal option is supported and the delay has timed out, the next state will reverse the port's transmit lanes. If the lane reversal option is not supported

and the delay has timed out, the next state will be retry receiver configuration. In the Config.WaitRmt state the following rules and commands shall be active:

- 1) TxCMD = SendTS2 (minimum of 16 TS2s)
- 2) RxCMD = WaitTS2
- 3) LinkPhyStat = Down
- 4) TS2 and SKIP ordered-sets are transmitted on all lanes
- 5) **SymbolErrorCounter** is enabled
- 6) For 64b/66b the **ErrorDetectionPerLaneCounter** is disabled Else If the RxStatus = RcvdTS2, the next state is Config.CfgEnhanced for ports implementing Rel. 1.2 Enhanced Signaling capability. For ports implementing only legacy functionality, the next state is Config.Idle.
- 7) Else If “Lane Reversal Option” and DelayTimeOut(2 ms), the next state is Config.TxRevLanes.
- 8) Else if not “Lane Reversal Option” and DelayTimeOut(2 ms), the next state is Config.RcvrCfg.

5.8.4.6.4 CONFIG.TXREVLANES STATE

In the reverse transmit lanes (Config.TxRevLanes) state, the transmitter reverses its lanes and continues to sends a series of TS2 ordered-sets on all lanes. The receiver monitors the configured physical lanes waiting for a transition from TS1 to TS2 ordered-sets, indicating the remote port has completed its configuration. In the Config.TxRevLanes state, the following rules and commands shall be active:

- 1) TxCMD = SendTS2 (minimum of 16 TS2s)
- 2) TxCMD = RevLanes
- 3) RxCMD = WaitTS2
- 4) LinkPhyStat = Down
- 5) TS2 and SKIP ordered-sets are transmitted on all lanes
- 6) **SymbolErrorCounter** is enabled For 64b/66b the **ErrorDetectionPerLaneCounter** is disabled
- 7) Else If the RxStatus = RcvdTS2, the next state is Config.CfgEnhanced for ports implementing Rel. 1.2 Enhanced Signaling capability. For ports implementing only legacy functionality, the next state is Config.Idle.
- 8) Else If DelayTimeOut(2 ms), the next state is Config.RcvrCfg.

5.8.4.6.5 CONFIG.CFGENHANCED STATE - RELEASE 1.3.1 ENHANCED SIGNALING

In the Configure Enhanced Capabilities state (Config.CfgEnhanced) the transmitter sends TS3 (Training Sequence 3) ordered-sets to negotiate with the link peer port what enhanced capabilities will be used.

In TS3 Rev 0, as described in [Section 5.5.2.4, "Training Sequence Three Ordered-Set \(TS3\)," on page 143](#), a TS3 ordered-set includes fields which allow the transmitter to indicate:

- which link speeds may be attempted (in a form consistent with the **SM.PortInfofo(LinkSpeedEnabled)** and **SM.PortInfo(LinkSpeedSupported)** fields described in [Section 5.6.2, "Status Outputs \(MAD get\)," on page 169](#)),
- whether or not adaptive transmitter pre-emphasis is requested, and
- whether or not the LINK HEARTBEAT functionality, required for DDR and QDR operation as described in [Section 5.14, "Link Heartbeat," on page 253](#), is enabled, and
- which transmitter equalization setting should be used, if equalization training has already occurred in the Config.Test state.

In TS3 Rev 1 and Rev 2, as described in [Section 5.5.2.4](#), a TS3 ordered-set includes fields which allow the transmitter to indicate:

- which link speeds may be attempted (in a form consistent with the **SM.PortInfofo(LinkSpeedEnabled)**, **SM.PortInfo(LinkSpeedExtEnabled)**, **SM.PortInfo(LinkSpeedSupported)** and **SM.PortInfo(LinkSpeedExtSupported)** fields described in [Section 5.6.2, "Status Outputs \(MAD get\)," on page 169](#)),
- whether or not adaptive transmitter pre-emphasis is requested,
- whether or not the LINK HEARTBEAT functionality, required for DDR and higher data rates operation as described in [Section 5.14, "Link Heartbeat," on page 253](#), is enabled,
- which transmitter equalization setting should be used for linkup, if equalization training has already completed in the Config.Test state or which equalization setting should be used in Config.Test,
- what is the requested Test Time (TT) in Config.Test,
- if PRBS11 test pattern is requested, rather than the default PRBS23 test pattern,
- whether 4 ms or 16 ms timer should be used for the FDR/EDR speed change time,
- whether or not Forward Error Correction (FEC) is requested,
- For Rev 2 only - whether Low Latency FEC (RS(271,257)) is supported,
- For Rev 2 only - whether Low Latency FEC is requested,
- For Rev 2 only - whether different polynomials mode is requested,
- For Rev 2 only - whether RS-FEC is supported at FDR rate,
- For Rev 2 only - whether Remote Transmit CDR enable is supported,
- For Rev 2 only - whether Remote Transmit CDR enable is requested,
- whether or not Fine Tuning algorithm is supported, and
- whether or not Max Packet Rate is supported.

In the Config.CfgEnhanced state the following rules and commands shall be active:

- 1) TxCMD = SendTS3
- 2) RxCMD = WaitTS3
- 3) LinkPhyStat = Down
- 4) TS3 and SKIP ordered-sets are transmitted on all lanes
- 5) **SymbolErrorCounter** is inhibited
- 6) For 64b/66b the **ErrorDetectionPerLaneCounter** is disabled
- 7) If RxStatus = RcvdTS3, and the exchanged TS3 indicate (a) a request for equalization on either port, OR (b) **MaxBothActive** changed from previous visit (i.e., first time on state since Config.RcvrCfg, or equalization failed on one of the ports), then:
 - For Rev 1 TS3 - The next state is Config.Test.
 - For Rev 2 TS3 - If the TS3 revision of both the received and transmitted TS3 is Rev 2 or higher (BothRev2 is set) and The maximum received SpeedActive differs from the maximum transmitted SpeedActive (MaxSpeedAgreed is cleared), then the next state is Config.WaitRmtTest. In addition, the ports shall clear the SpeedActive bits of speeds that are higher than MaxBothActive and shall request equalization in the next TS3 transmission. If the TS3 revision of the received TS3 is Rev 1 or lower (BothRev2 is cleared) or the maximum received SpeedActive is equal to the maximum transmitted SpeedActive (MaxSpeedAgreed is set), then the next state is Config.Test.
- 8) Also, **LinkSpeedActive** and **LinkSpeedExtActive** is set equal to the AND of the current **LinkSpeedActive** and **LinkSpeedExtActive** with the **SpeedActive** field of the received TS3 to determine the speed(s) active on both ports.
- 9) Else if RxStatus = RcvdTS3, and both no further equalization is requested AND **MaxBothActive** has not changed from previous visit, and **LinkSpeedActive** and **LinkSpeedExtActive** is set to the highest bit set on both ports, the next state is Config.WaitCfgEnhanced.
 - If both ports exchange Rev 1 or Rev 2 TS3:
 - If MaxBothActive is greater than QDR and FEC request is set on either received or transmitted TS3 then FEC_en is set to true
 - For Rev 1 TS3 - FEC type is set to Fire-Code.
 - For Rev 2 TS3 - If the LinkSpeedActive equals to EDR or LinkSpeedActive equals to FDR and both the transmitted and received FRSFS are set, then RS-FEC is set to true. Otherwise, FEC type is set to Fire-Code.
 - For Rev 2 TS3 - If RS-FEC is set and either the received or transmitted LLFR are set and both the transmitted and received LLFS are set, then FEC type is set to RS(271,257). Otherwise, FEC type is set to RS(528,514).
 - If MaxBothActive is greater than QDR and Fine Tuning (FT) support is set on both received and transmitted TS3 then FT_en is set to true.

- If the MPR bit on the received TS3 is set than MPR_en may be set to false.
- 10) Else If RxStatus = RcvIdle (indicating the port is connected to a legacy device port), the next state is Config.Idle.
- For legacy compatibility, an 8x device should assert RxStatus = RcvIdle if it receives Idles on only lanes 0...3, as described in [Section 5.8.7.7, "RxCMD = Wait-Idle," on page 233](#).
 - After following this transition, since the port is connected to a legacy device, all Rel. 1.2 and Rel. 1.3 Enhanced Signaling capabilities (DDR/QDR/FDR/EDR, FEC, heartbeat, etc.) are inactivated. Furthermore, an 8x device reverting to legacy operation shall operate with 4x or 1x link width, with only lanes 0-3 or lane 0 configured.
- 11) Else if DelayTimeOut (2 ms), the next state is Config.RcvrCfg.

5.8.4.6.6 CONFIG.TEST STATE - RELEASE 1.2 ENHANCED SIGNALING

For QDR and lower data rates:

If either the sent or received TS3 in Config.CfgEnhanced were Rev 0 TS3: In the Configure Test state (Config.Test) the transmitter will send idle data at the maximum speed indicated in the **LinkSpeedActive** bit map, for a 100 millisecond period. The 100 ms period is split into a 36 ms period of default transmitter equalization time, and 16 separate adaptive transmitter equalization setting slots of 4 ms duration each. On each 4 ms slot, the transmitter sends one of a vendor-dependent set of adaptive equalization settings. The receiver, on each slot, attempts to receive intact data (valid and supported 8b/10b code groups, with correct running disparity), while optionally adjusting receiver equalization settings.

The transmitter is expected to select the 16 settings on the basis of its knowledge of the transmission channel, so that the settings will, to the extent possible, be well-matched to the channel.

The receiver equalization techniques and settings used are vendor-dependent, don't require transmitter/receiver handshaking across a link. Therefore, the receiver equalization is unspecified here.

Once the 100 ms period of the Config.Test state is over, the receiver decides which of the 16 adaptive transmitter equalization settings was the first usable one. This request is indicated in the TS3 ordered-sets transmitted in the Config.CfgEnhanced state following this Config.Test state.

Transmitter and receiver adaptive equalization capabilities are optional, and may not be implemented in some devices. In this case, the non-implementing device will return a TS3 with DDSV and the DDS bits all cleared to 0, and the attached device will use the default equalization setting of the first 36 ms.

If both the sent and received TS3 in Config.CfgEnhanced were Rev 1 TS3: In the Configure Test state (Config.Test) the transmitter will send idle data at the maximum speed indicated in the **LinkSpeedActive** bit map, for a TT time as defined in [Table 41 Test Time Values](#). The transmitter shall use the transmitter equalization preset or coefficient requested in the TS3 Rev 1 ordered-sets received in Config.CfgEnhanced.

The receiver shall attempt to receive intact data (valid and supported 8b/10b code groups, with correct running disparity), while optionally adjusting receiver equalization settings.

The receiver equalization techniques and settings used are vendor-dependent, don't require transmitter/receiver handshaking across a link. Therefore, the receiver equalization is unspecified here.

Once the TT time period of the Config.Test state is over, the receiver should evaluate the quality of the transmitter equalization setting that was tested. The receiver should then communicate the results of Config.Test in the next instance of Config.CfgEnhanced state using the TS3 ordered-sets by requesting one of the following:

- a) The same transmitter equalization to be tested again.
- b) A different transmitter equalization to be tested.
- c) Receiver equalization has completed- no further equalization is required.

For FDR and higher data rates:

In the Configure Test state (Config.Test) the transmitter will send either PRBS23 or PRBS11 at the maximum speed indicated in the **LinkSpeedActive** and **LinkSpeedExtActive** bit map, for a test time of TT as defined in [Table 41 on page 191](#). The first 4, 16, 32, or 64 ms (depending on the value of the Extended Speed Change Time (SCTH, SCT)) is used for speed change and stabilization time.

A port may be connected to a medium that requires speed dependent configuration such as equalization configuration or CDR control. In addition, a port that supports remote Remote Transmit CDR enable (Transmitted RTCS is set) shall set the medium transmit CDR setting based on the received Transmit CDR enable Request (RTCS). Setting the attached medium properties may require additional setting time and stabilization time. The transmitter port shall guarantee that the medium settings are applied and that a valid training pattern is transmitted out of the attached physical medium by the end of the 4, 16, 32, or 64 ms speed change time.

The remaining test time is used for receiver test. It is highly recommended that an implementation will stop its equalization process some time before the completion of the test time to ensure that all the equalization is performed while the peer port is in Config.Test. The transmitter equalization is set to the preset or coefficients requested in the received TS3 in Config.CfgEnhanced state. The receiver attempts to receive PRBS23 or PRBS11 while optionally adjusting receiver equalization settings.

The receiver equalization techniques and settings used are vendor-dependent. The receiver equalization doesn't require transmitter/receiver handshaking across a link. Therefore, the receiver equalization is unspecified here.

Once the Test Time of the Config.Test state is over, the receiver should evaluate the quality of the transmitter equalization setting that was tested. The receiver should then communicate the results of Config.Test in the next instance of Config.CfgEnhanced state using the TS3 ordered-sets by requesting one of the following:

- a) The same transmitter equalization to be tested again.

b) A different transmitter equalization to be tested.

c) Receiver equalization has completed- no further equalization is required.

A receiver that determine that the current speed (**LinkSpeedActive** and **LinkSpeedExtActive**) can not be supported may chose to test a lower speed. The receiver shall communicate the new speed request by clearing the current higher speed bit in the **SpeedActive** symbol of the transmitted TS3 ordered-sets in the next instance of Config.CfgEnhanced state.

In the Config.Test state the following rules and commands shall be active:

1) SPEED = MaxBothActive

2) TxCMD = SendIdle for QDR and lower bit rates.

3) TxCMD = SendPRBS for FDR and higher bit rates - if the TP bit was set in both the received and transmitted TS3 in Config.CfgEnhanced state then TxCMD = SendPRBS11, Else TxCMD = SendPRBS23.

4) RxCMD = WaitIdle for QDR and lower bit rates.

5) RxCMD = WaitPRBS for FDR and higher bit rates - if TP bit was set in both the received and transmitted TS3 in Config.CfgEnhanced state then RxCMD = WaitPRBS11, Else RxCMD = WaitPRBS23.

6) Idle and SKIP ordered-sets are transmitted on configured lanes for QDR and lower bit rates

7) LinkPhyStat = Down

8) **SymbolErrorCounter** is enabled for QDR and lower bit rates. For FDR and higher bit rates **PRBSErrorCounter** is enabled

9) For QDR and lower bit rates - If either the sent or received TS3 in Config.CfgEnhanced was Rev 0 TS3 and DelayTimeOut(100 ms), next state is Config.WaitRmtTest. Else if both the sent and received TS3 in Config.CfgEnhanced were Rev 1 TS3 and DelayTimeOut(TT ms), next state is Config.WaitRmtTest.

- If the receiver was unable to recover data from the received signal without incurring any minor link errors, the bit indicating the currently-active speed, MaxBothEnabled, should be cleared to 0 in **LinkSpeedActive**. This modified Speeds mask, indicating that the current speed is not supported, is reflected in further transmitted TS3 ordered-sets.

- A receiver may use SKIP ordered-sets to ensure that the requirement of total link skew of 6 symbol times (60 UI) or less at SDR, DDR, and QDR rates has been met. Please see [Table 70](#), [Table 71](#), [Table 155](#), [Table 156](#), and [Table 157](#). If the total link skew is measured to be more than the maximum allowed lane-to-lane skew value, the bit indicating the currently-active speed, MaxBothEnabled, should be cleared to 0 in **LinkSpeedActive**. This modified **LinkSpeedActive** mask, indicating that the current speed is not supported, is reflected in further transmitted TS3 ordered-sets.

For FDR and higher bit rates - If DelayTimeOut(TT ms), next state is Config.Wait-RmtTest.

- If the receiver was unable to receive a PRBS23/11 on any of the requested transmitter equalization, the bit indicating the currently-active speed, MaxBoth-Enabled, should be cleared to 0 in **LinkSpeedActive**. This modified Speeds mask, indicating that the current speed is not supported, is reflected in further transmitted TS3 ordered-sets.

5.8.4.6.6.14 PRBS TEST PATTERN

At bit rates higher than QDR each transmitter sends PRBS23 or PRBS11 pattern at the maximum speed commonly active in both ports (FDR or EDR) during for a pre-negotiated time during Config.Test.

A port may request the shorter PRBS11 pattern to be used for the equalization if support of this optional test pattern is reported in the TS3 ordered-set - see [Section 5.5.2.4](#), if the PRBS11 request bit is set on both transmitted and received TS3 then the PRBS23 is replaced by a PRBS11.

A port may request that different polynomials will be used in the PRBS pattern for each lane (TPDP bit of TS3 set to 1). When both the receiver and the transmitter have requested to use different polynomials, the pattern generator shall use for each lane the generator polynomial provided in Table 43 if PRBS11 was chosen or in Table 44 if PRBS23 was chosen. The polynomials shall be used for both the generator and the checker. If the link width is x8 or x12, then lanes 5:8 and lanes 9:12 use the same polynomials as lanes 0:3. An EDR-capable port shall support different polynomials on different lanes.

The PRBS23 pattern generator shall produce the same result as the implementation shown in [Figure 62](#), this implements the inverted version of the bit stream produced by the polynomial in [Figure 62](#).

$$g(x)=x^{23} + x^{18} + 1$$

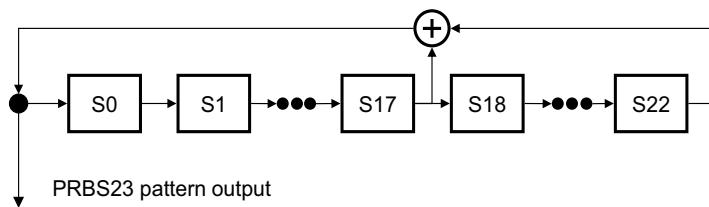


Figure 62 PRBS23 Pattern Generator

The PRBS23 pattern checker shall produce the same result as the implementation shown in [Figure 63](#).

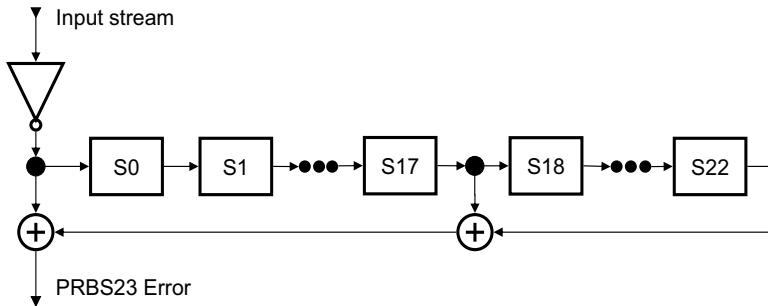


Figure 63 PRBS23 checker

The PRBS11 pattern generator shall produce the same result as the implementation shown in [Figure 64](#), this implements the inverted version of the bit stream produced by the polynomial in [Figure 64](#).

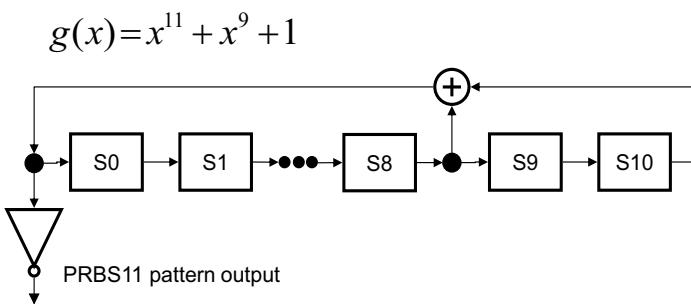


Figure 64 PRBS11 Pattern Generator

The PRBS11 pattern checker shall produce the same result as the implementation shown in [Figure 65](#).

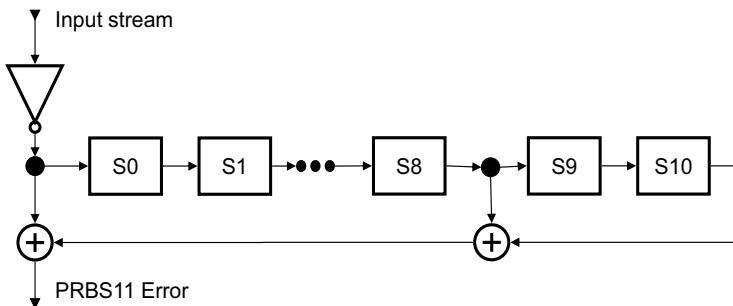


Figure 65 PRBS11 checker

Table 43 PRBS11 polynomials when Different Polynomials are used for Each Lane

Lane	Polynomial $g(x)$
0	$x^{11} + x^{10} + x^6 + x^5 + 1$
1	$x^{11} + x^9 + x^6 + x^5 + 1$
2	$x^{11} + x^8 + x^6 + x^4 + 1$
3	$x^{11} + x^7 + x^6 + x^4 + 1$

Table 44 PRBS23 Polynomials when Different Polynomials are used for Each Lane

Lane	Polynomial $g(x)$
0	$x^{23} + x^{18} + 1$
1	$x^{23} + x^{18} + x^{12} + x^6 + 1$
2	$x^{23} + x^{22} + x^{19} + x^{18} + 1$
3	$x^{23} + x^{19} + x^{18} + x^{11} + 1$

5.8.4.6.7 CONFIG_WAITRMTTEST STATE - RELEASE 1.2 ENHANCED SIGNALING

In the Configure Wait for Remote Test state (Config.WaitRmtTest) the transmitter sends a series of TS2 ordered-sets at the MinEnabledSpeed transmitter speed to indicate to the link peer port that it has completed attempting operation at the maximum speed commonly active in both ports, with attempted equalization training. In the Config.WaitRmtTest state the following rules and commands shall be active:

- 1) SPEED = MinEnabledSpeed
- 2) TxCMD = SendTS2
- 3) RxCMD = EnDeskew
- 4) LinkPhyStat = Down
- 5) TS2 and SKIP ordered-sets are transmitted on all lanes at the currently-active transmitter speed.
- 6) **SymbolErrorCounter** is inhibited
- 7) For 64b/66b the **ErrorDetectionPerLaneCounter** is disabled
- 8) If DelayTimeOut (2,4 or 16 ms), the next state is Config.CfgEnhanced
 - If the speed in Config.Test state was QDR or lower then the DelayTimeOut is 2 ms.
 - Else if SCT was cleared to 0 in both transmitted and received TS3s in Config.CfgEnhanced then DelayTimeOut is 4 ms, else DelayTimeOut is 16 ms.

5.8.4.6.8 CONFIG_WAITCFGENHANCED STATE - RELEASE 1.2 ENHANCED SIGNALING

In the Configure Wait Remote Configure Enhanced state (Config.WaitCfgEnhanced) the transmitter sends a series of TS2 ordered-sets at the currently-active transmitter speed to indicate to the link peer port that it has (a) completed TS3 exchange for negotiating enhanced capabilities, and (b) committed to operation at the speed negotiated in the Config.CfgEnhanced TS3 exchange and attempted in a previous Config.Test state.

Note: If the current speed is higher than QDR then the 64b/66b encoding and scrambling are enabled for both the transmitter and receiver and the 8b/10b encoding is disabled.

If FEC_en is true then at this state the FEC encoder and decoder are enabled, the transmitter encodes the sent bit stream, the receiver acquires FEC lock and decodes the in-bound bit stream. The LinkPhy state machine will set the align_enable variable when entering this state to enable the RS-FEC alignment state machine and FEC lock state machine to acquire lock.

A port that is connected to a medium that requires speed dependent setting such as equalization configuration or CDR control. In addition, a port that supports remote Remote Transmit CDR enable (Transmitted RTCS is set) shall set the medium transmit CDR setting based on the received Transmit CDR enable Request (RTCS). Setting the attached medium properties may require additional setting time and stabilization time. The port shall config the attached medium during the Config.WaitCfgEnhanced state.

In the Config.WaitCfgEnhanced state the following rules and commands shall be active:

- 1) SPEED = MaxBothActive
- 2) TxCMD = SendTS2
- 3) RxCMD = EnDeSkew
- 4) LinkPhyStat = Down
- 5) TS2 and SKIP ordered-sets are transmitted on all lanes at the currently-active transmitter speed.
- 6) **SymbolErrorCounter** is inhibited
- 7) For 64b/66b the **ErrorDetectionPerLaneCounter** is disabled
- 8) For EDR-capable ports, **FECPortCorrectedSymbolCounter** and FECLaneCorrectedSymbolCounter are disabled.
- 9) If FEC_en is true, transmitter sends FEC encoded data, and receiver acquires FEC lock and enables the FEC decode.
- 10) Set align_enable to true.
- 11) If DelayTimeOut (2 or 16 ms), the next state is Config.Idle.
 - If MaxBothActive is QDR or lower then the DelayTimeOut is 2 ms.
 - Else if MaxBothActive is higher than QDR then the DelayTimeOut is 16 ms.

5.8.4.6.9 CONFIG.IDLE STATE

In the configuration idle (Config.Idle) state, both the local and remote ports have completed configuration and training. The port now transmits link idle data at the configured active speed. The receiver monitors the configured physical lanes, waiting for a transition from TS2 ordered-sets to Idle Data. When Link Idle Data is received the link is up. In the Config.Idle state, the following rules and commands shall be active:

- 1) TxCMD = SendIdle
- 2) RxCMD = WaitIdle
- 3) LinkPhyStat = Down
- 4) Idle data and SKIP ordered-sets are transmitted on configured lanes only.
- 5) **SymbolErrorCounter** is enabled
- 6) For 64b/66b encoding the **ErrorDetectionPerLaneCounter** is enabled
- 7) If FEC_en is true, transmitter sends FEC encoded data, and receiver enables the FEC decode.
- 8) For EDR-capable ports, **FECPortCorrectedSymbolCounter** and **FECLaneCorrectedSymbolCounter** are disabled.
- 9) Set align_enable to false.
- 10) If RxStatus = RcvdTS1 or If DelayTimeOut(2 ms), the next state is Config.RcvrCfg.

11) If RxStatus = RcvIdle then next state is LinkUp

5.8.4.7 LINKUP STATE

The LinkUp state is the normal operational state of the port. Packets provided by the Link Layer are transmitted, and received packets are forwarded to the Link Layer. This is the only state in which the Link Physical logical status signal LinkPhyStat is assigned the value “Up”, indicating that the link is available to transfer packets. In the LinkUp state, Physical Link error handling is enabled. The Physical Link Error Handling is defined in [Section 5.9. “Link Physical Error Handling.” on page 235](#). If the FT_en is set to true a port may perform the Fine Tuning algorithm while in LinkUp state as described in [Section 5.13. “Fine Tuning.” on page 246](#). In the LinkUP state the following rules and commands shall be active:

- 1) TxCMD = Enable
- 2) RxCMD = Enable
- 3) LinkPhyStat = Up
- 4) The **SymbolErrorCounter** is enabled
- 5) For 64b/66b the **ErrorDetectionPerLaneCounter** is enabled
- 6) If FEC_en is true, transmitter sends FEC encoded data, and receiver enables the FEC decode.
- 7) For EDR-capable ports, FECPortCorrectedSymbolCounter and **FECLaneCorrectedSymbolCounter** are enabled.
- 8) If RxStatus = RxMajorError or LinkPhyRecover, the next state is Recovery.
- 9) If RxStatus = RxHeartbeatError,
 - a) The **LinkDownedCounter** is incremented
 - a) The **LinkRoundTripLatency** value is reset to FFFF FFFFh
 - b) The next state is the **LinkDownDefaultState** (Polling or Sleeping)

5.8.4.8 LINK ERROR RECOVERY STATES

The Link Error Recovery Super State controls port error recovery. The recovery process is triggered when an error is detected by the Link Layer or Link/Physical Layer. The Physical Layer error handling logic monitors minor errors and, when a rate threshold is reached or when a major error is detected, triggers recovery. (See [Section 5.9 on page 235](#))

The recovery process starts by sending a stream of TS1 ordered-sets to trigger error recovery at the remote port. When TS1 or TS2 is received by the local receiver, it will use the current configuration (width, lane reversal, serial data inversion, encoding, and speed) to retrain.

The second step of recovery starts when the receiver has completed retraining. The port starts sending a stream of TS2 ordered-sets. The port then waits for the remote port to

complete retraining. When the port is both sending and receiving TS2, both receivers have been retrained, and the second step of recovery is complete.

In the final step of link recovery, the port transmits an idle data stream and waits for idle data. When the port is both sending and receiving idle data, recovery is complete, and the link is up.

If the recovery process fails at any step, the Link/Physical Layer state machine returns to its link down default state.

The states of Recovery Super State are expanded in the state graph shown in [Figure 66](#).

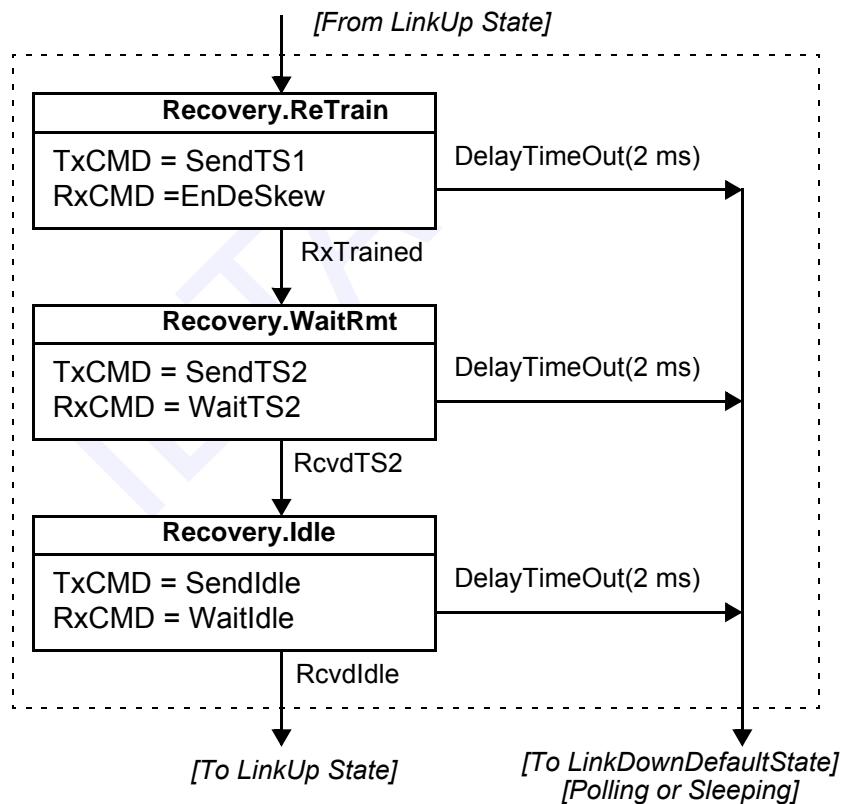


Figure 66 Recovery Super State (Expanded)

5.8.4.8.1 RECOVERY.RETRAIN STATE

In the Recovery Retrain (Recovery.Retrain) state, the transmitter is commanded to send a series of TS1 ordered-sets using the currently configured lanes. The receiver is enabled to reacquire symbol/block sync and then to de-skew the lanes. If FEC is enabled as indicated by FEC_en receiver is enabled to reacquire FEC block lock. In the Recovery.Retrain state the following rules and commands shall be active:

- 1) TxCMD = SendTS1 (minimum of 16 TS1s)
- 2) RxCMD = EnDeSkew
- 3) LinkPhyStat = Down
- 4) TS1 and SKIP ordered-sets are transmitted on configured lanes only.
- 5) The port **SymbolErrorCounter** is enabled.
- 6) For 64b/66b the **ErrorDetectionPerLaneCounter** is enabled
- 7) If FEC_en is true, transmitter sends FEC encoded data, and receiver enables the FEC decode.
- 8) For EDR-capable ports, FECPortCorrectedSymbolCounter and **FECLaneCorrectedSymbolCounter** are disabled.
- 9) Set align_enable to true.
- 10) If the RxStatus = RxTrained, the next state is Recovery.WaitRmt.
- 11) Else if DelayTimeOut(2 ms),
 - a) The **LinkDownedCounter** is incremented
 - b) The **LinkRoundTripLatency** value is reset to FFFF_FFFFh
 - c) The next state is the **LinkDownDefaultState** (Polling or Sleeping)

5.8.4.8.2 RECOVERY_WAITRMT STATE

In the Recovery Wait Remote state (Recovery.WaitRmt), the transmitter is commanded to send a series of TS2 ordered-sets on the configured lanes. The receiver monitors the receive streams on the configured lanes. In the Recovery.WaitRmt state the following rules and commands shall be active:

- 1) TxCMD = SendTS2 (minimum of 16 TS2s)
- 2) RxCMD = WaitTS2
- 3) LinkPhyStat = Down
- 4) TS2 and SKIP ordered-sets are transmitted on configured lanes only
- 5) The port **SymbolErrorCounter** is enabled.
- 6) For 64b/66b the **ErrorDetectionPerLaneCounter** is enabled
- 7) If FEC_en is true, transmitter sends FEC encoded data, and receiver enables the FEC decode.
- 8) For EDR-capable ports, FECPortCorrectedSymbolCounter and **FECLaneCorrectedSymbolCounter** are enabled.
- 9) Set align_enable to true.
- 10) If the RxStatus = RcvdTS2 and Deskew Complete, the next state is Recovery.Idle
- 11) Else if DelayTimeOut(2 ms)
 - a) The **LinkDownedCounter** is incremented

-
- b) The **LinkRoundTripLatency** value is reset to FFFF_FFFFh
 - c) The next state is the **LinkDownDefaultState** (Polling or Sleeping)

5.8.4.8.3 RECOVERY.IDLE STATE

In the Recovery.Idle state (Recovery.Idle), both ports have retrained, and the transmitter is commanded to send on the configured lanes. The receiver monitors the receive streams on the configured lanes waiting for idle data. In the Recovery.Idle state, the following rules and commands shall be active:

- 1) TxCMD = SendIdle (minimum of 16 symbol times)
- 2) RxCMD = WaitIdle
- 3) LinkPhyStat = Down
- 4) Idle Data and SKIP ordered-sets are transmitted on configured lanes only.
- 5) The port **SymbolErrorCounter** is enabled.
- 6) For 64b/66b the **ErrorDetectionPerLaneCounter** is enabled
- 7) If FEC_en is true, transmitter sends FEC encoded data, and receiver enables the FEC decode.
- 8) For EDR-capable ports, FECPortCorrectedSymbolCounter and **FECLaneCorrectedSymbolCounter** are enabled.
- 9) Set align_enable to false.
- 10) If RxStatus = RcvdIdle,
 - a) The **LinkErrorRecoveryCounter** is incremented
 - b) The next state is LinkUp.
- 11) Else If DelayTimeOut(2 ms),
 - a) The **LinkDownedCounter** is incremented
 - b) The **LinkRoundTripLatency** value is reset to FFFF_FFFFh
 - c) The next state is the **LinkDownDefaultState** (Polling or Sleeping)

5.8.5 STATE MACHINE DELAYS AND TIMEOUTS

Operations in certain states involve delays or time-out periods. These time-out periods start as the state is entered and are cleared when a transition to a different state is taken.

C5-15: All ports shall use the delays and tolerances defined in [Section 5.8.5, "State Machine Delays and Timeouts," on page 225](#).

The delay timeouts used by the link training state machine shall have the following tolerances:

- 1) DelayTimeOut (2 ms): 2 ms (-1 µs/ +126 µs)
- 2) DelayTimeOut (4 ms): 4 ms (-1 µs/ +126 µs)

3) DelayTimeOut (16 ms): 16 ms (-1 μ s/ +126 μ s)

4) DelayTimeOut (100 ms): 100 ms (-1 μ s/ +126 μ s)

5) DelayTimeOut (150 ms): 150 ms (-1 μ s/ +126 μ s)

6) DelayTimeOut (400 ms): 400 ms (-1 μ s/ +126 μ s)

7) DelayTimeOut (Test Time) 1 ms - 4096 ms (-1 μ s/+126 μ s)

Note: The delay tolerances specified above do not include plus/minus 100 ppm allocated to the clock oscillator. The two tolerances are cumulative (delay tolerance + oscillator tolerance).

5.8.6 TRANSMITTER INTERFACE AND BEHAVIOR

The link training state machine controls the transmitter behavior using a set of transmitter commands (TxCMD). The transmitter reports its status as TxStatus. The expected behavior of each command is specified in the following paragraphs.

C5-16: All ports shall implement the transmitter behaviors defined in [Section 5.8.6.1, "Transmitter Interface and Behavior," on page 226](#) but are not required to implement the optional lane reversal function.

o5-1: All ports that implement the optional lane reversal function shall implement the behavior as defined in [Section 5.8.6.6, "TxCMD = RevLanes," on page 228](#).

5.8.6.1 TxCMD = DISABLE

When transmit command is “disabled”, the transmitter output signals on all physical lanes shall be driven to a quiescent condition.

5.8.6.2 TxCMD = SENDTS1

The SendTS1 command instructs the transmitter to send a sequence of training sequence one ordered-sets (TS1). When commanded to send TS1 ordered-sets, the transmitter shall implement the following rules:

- 1) The first TS1 ordered-set shall be transmitted at the next available ordered-set or symbol/block boundary. The packet currently being transmitted **may** be terminated abnormally to send TS1 ordered-sets that trigger error recovery.
- 2) For 8b/10b encoding: The first TS1 ordered-set transmitted should force the current running disparity in all lanes to the same value (positive or negative)
- 3) The SKIP ordered-sets have the highest transmit priority. When scheduled, a SKIP ordered-set shall be transmitted at the next ordered-set boundary
- 4) Complete TS1 ordered-sets shall be transmitted back-to-back as long as the SendTS1 command is active.
- 5) At a minimum, sixteen (16) TS1 ordered-sets shall be transmitted.
- 6) In the Polling and Configuration Super States:
 - a) The SKIP ordered-sets shall be transmitted on all enabled physical lanes.

- b) The TS1 ordered-sets shall be transmitted on all enabled physical lanes.
- 7) In the Recovery Super State:
 - a) The SKIP ordered-sets shall be transmitted only on configured physical lanes.
 - b) The TS1 ordered-sets shall be transmitted only on configured physical lanes.
 - c) Lanes not included in the configured link width (unused) shall be forced to a quiescent condition.

5.8.6.3 TxCMD = SENDTS2

The SendTS2 command instructs the transmitter to send a sequence of training sequence two ordered-sets (TS2). When commanded to send TS2 ordered-sets, the transmitter will implement the following rules.

- 1) The first TS2 ordered-set shall be transmitted at the next available ordered-set boundary.
- 2) Complete TS2 ordered-sets shall be transmitted back-to-back as long as the SendTS2 command is active.
- 3) At a minimum, sixteen (16) TS2 ordered-sets shall be transmitted.
- 4) The SKIP ordered-sets have the highest transmit priority. When scheduled, a SKIP ordered-set shall be transmitted at the next packet or ordered-set boundary.
- 5) In the Config.WaitRmt and Config.TxRevLane States:
 - a) The SKIP ordered-sets shall be transmitted on all enabled physical lanes.
 - b) The TS2 ordered-sets shall be transmitted on all enabled physical lanes.
- 6) In the Recovery.WaitRmt, the Config.WaitRmtTest, and the Config.WaitCfgEnhanced State:
 - a) The SKIP ordered-sets shall be transmitted only on configured physical lanes.
 - b) The TS2 ordered-sets shall be transmitted only on configured physical lanes.
 - c) Lanes not included in the configured link width (unused) shall be forced to a quiescent condition.

5.8.6.4 TxCMD = SENDTS3

The SendTS3 command instructs the transmitter to send a sequence of training sequence three ordered sets (TS3). When commanded to send TS3 ordered-sets, the transmitter will implement the following rules:

- 1) The first TS3 ordered set shall be transmitted at the next available ordered-set boundary.
- 2) Complete TS3 ordered sets shall be transmitted back-to-back as long as the SendTS3 command is active.
- 3) At a minimum, sixteen (16) TS3 ordered sets shall be transmitted.

- 4) The SKIP ordered sets have the highest transmit priority. When scheduled a SKIP ordered set shall be transmitted at the next packet or ordered set boundary.
- 5) The SKIP ordered-sets shall be transmitted on all enabled physical lanes.
- 6) The TS3 ordered-sets shall be transmitted on all enabled physical lanes.

5.8.6.5 TxCMD = SENDIDLE

The SendIdle command instructs the transmitter to send the link idle data sequence. When commanded to send idle data, the transmitter shall implement the following rules.

- 1) The SKIP ordered-sets have the highest transmit priority. When scheduled, a SKIP ordered-set shall be transmitted as soon as possible by interrupting link idle data sequence transmission.
- 2) The SKIP ordered-sets shall be transmitted only on configured physical lanes.
- 3) The link idle data sequence/blocks shall be transmitted only on configured physical lanes.
- 4) Lanes not included in the configured link width (unused) shall be forced to a quiescent condition.
- 5) For 8b/10b encoding: at a minimum, sixteen (16) symbol times of link idle data shall be transmitted on each physical lane for Config.Idle and Recovery.Idle states. In LinkUp state, idle data symbols may be preempted by higher priority traffic, even if fewer than 16 Idle symbols have been transmitted.

For 64b/66b encoding with RS-FEC not enabled: at a minimum, sixteen (16) block times of link idle data shall be transmitted on each physical lane for Config.Idle and Recovery.Idle states. In LinkUp state, idle data symbols may be preempted by higher priority traffic, even if fewer than 16 Idle symbols have been transmitted.

For 64b/66b encoding with RS-FEC enabled: at a minimum, thirty-two (32) FEC codewords times of link idle data shall be transmitted on the port for Config.Idle and Recovery.Idle states. In LinkUp state, idle data symbols may be preempted by higher priority traffic, even if fewer than 32 Idle codewords have been transmitted.

- 6) For 64b/66b encoding - if Fine Tuning is enabled and the LinkPhy state is LinkUp or Config.Idle, then the transmitter is permitted to replace the Idle block with Fine Tuning blocks on all lanes.

5.8.6.6 TxCMD = REVLANES

The Reverse Lanes command is an optional modifier to the SendTS2 command. The RevLanes command instructs the transmitter to reverse the order of its transmit lanes. When commanded to reverse lanes, the transmitter shall implement the following rules.

- 1) The operation of the SendTS2 command shall not be interrupted.
- 2) The lane swap shall not cause the remote receiver to detect an error.
- 3) Four wide (4x) port shall reverse lanes causing the following lane swaps: 0 to 3, 1 to 2, 2 to 1, and 3 to 0.

- 4) Eight wide (8x) port shall reverse lanes causing the following lane swaps: 0 to 7, 1 to 6, 2 to 5, 3 to 4, 4 to 3, 5 to 2, 6 to 1, and 7 to 0.
- 5) Twelve wide (12x) port shall reverse lanes causing the following lane swaps: 0 to 11, 1 to 10, 2 to 9, 3 to 8, 4 to 7, 5 to 6, 6 to 5, 7 to 4, 8 to 3, 9 to 2, 10 to 1, and 11 to 0.

5.8.6.7 SPEED = MAXBOTHACTIVE

This command configures the transmitter to transmit data at the maximum rate commonly enabled on both ports of the link that is also supported by the link medium.

5.8.6.8 SPEED = MINENABLEDSPEED

This command configures the transmitter to transmit data at the minimum speed enabled on the port, generally the SDR rate.

5.8.6.9 TxCMD = ENABLE

This command enables the normal operation condition of the transmitter. When enabled, the transmitter will implement the following rules.

- 1) The transmission of SKIP ordered-sets, link heartbeats, packets, and link idle data shall be restricted to the configured link width.
- 2) Physical lanes not included in the configured link width (unused) shall be forced to a quiescent condition.
- 3) SKIP ordered-sets have the highest transmit priority. When scheduled, a SKIP ordered-set shall be transmitted as soon as possible at the next packet boundary, at the next ordered-set boundary, or by interrupting link idle transmission.
- 4) On ports that implement Rel. 1.2 Enhanced Signaling capability, Link Heartbeat Ordered-Sets have the second highest transmit priority. When a Link Heartbeat ordered-set is available for transmission, it shall be transmitted as soon as possible at the next packet boundary, at the next ordered-set boundary, or by interrupting link idle transmission.
- 5) Packets have the next highest transmit priority. When a packet is available for transmission, it shall be transmitted as soon as possible at the next packet boundary, at the next ordered-set boundary, or by interrupting link idle transmission. For speeds higher than QDR when MPR_en = false, a start of packet block must occur 64Bytes or more from the previous start of packet - see [Section 5.15, "Max Packet Rate," on page 255](#).
- 6) On links where both ports support Fine Tuning and Fine Tuning is enabled, Fine Tuning blocks have the second lowest priority.
- 7) When there are no SKIP ordered-sets, link heartbeats, packets, or fine tuning blocks to transmit, the transmitter shall transmit the link idle data pattern.

5.8.6.10 TxCMD = SENDPRBS23

The SendPRBS23 command instructs the transmitter to send the Pseudo Random PRBS23 pattern as defined in [Section 5.8.4.6.6.14](#). When commanded to send PRBS23 the transmitter shall implement the following rules:

- 1) PRBS23 shall be transmitted on all enabled physical lanes. The pattern transmitted
on each physical lane shall not be correlated with the pattern transmitted on any
other lane. When different polynomials mode is enabled (both sides set the TPDP bit
of TS3), each lane uses a different polynomial for the PRBS23, which guarantees
that there is no correlation between the lanes.
- 2) The PRBS23 pattern shall be transmitted back-to-back as long as the SendPRBS23
command is active.
- 3) The SKIP ordered sets will not be transmitted.

5.8.6.11 TxCMD = SENDPRBS11

The SendPRBS11 command instructs the transmitter to send the Pseudo Random
PRBS11 pattern as defined in [Section 5.8.4.6.6.14](#). When commanded to send PRBS11
the transmitter shall implement the following rules:

- 1) PRBS11 shall be transmitted on all enabled physical lanes. The pattern transmitted
on each physical lane shall not be correlated with the pattern transmitted on any
other lane. When different polynomials mode is enabled (both sides set the TPDP bit
of TS3), each lane uses a different polynomial for the PRBS11, which guarantees
that there is no correlation between the lanes.
- 2) The PRBS11 pattern shall be transmitted back-to-back as long as the SendPRBS11
command is active.
- 3) The SKIP ordered sets will not be transmitted.

Implementation Note:

The SENDPRBS23 and SENDPRBS11 TxCMD functions are required to transmit uncorrelated patterns
on all lanes. This is required, in order to achieve lane-to-lane crosstalk conditions necessary for optimal
equalization during the Config.Test state.

The means in which a device achieves the above requirement is implementation specific, it can be
achieved by implementing a different PRBS seed on each lane, implementing a delay between the lanes,
or by any other method which ensures uncorrelated patterns on each lane. Note that a device using
delay to create the uncorrelated data must delay by at least 132 UI to compensate for the maximum al-
lowed lane to lane skew.

5.8.7 RECEIVER INTERFACE AND BEHAVIOR

The link training state machine controls the receiver behavior using a set of receiver com-
mands (RxCMD). The receiver reports its status as RxStatus. The expected behavior of
the receiver in response to each command is described in the following paragraphs. Re-
ceiver status conditions are defined as part of receiver command definition.

C5-17: All ports shall implement the Receiver behaviors defined in [Section 5.8.7, “Re-
ceiver Interface and Behavior,” on page 230](#) but are not required to implement the op-
tional lane reversal function or correction of inverted serial data.

o5-2: All ports that implement the optional correction of inverted serial data shall implement the behavior as defined in [Section 5.8.7.3, “RxCMD = EnConfig,” on page 231](#) rule 2.

o5-3: All ports that implement the optional lane reversal function shall implement the behavior as defined in [Section 5.8.7.3, “RxCMD = EnConfig,” on page 231](#) rule 4.

5.8.7.1 RxCMD = DISABLE

When the receiver is disabled, the receiver shall implement the following rules:

- 1) The receiver shall not forward packets to the upper layer protocols.
- 2) The receiver shall not update the port error counters. (See [Section 5.6.3 on page 173](#))

5.8.7.2 RxCMD = WAITTS1

This command instructs the receiver to look for TS1 on all physical lanes independently. When commanded to WaitTS1, the receiver shall implement the following rules.

- 1) For 8b/10b encoding - Symbol synchronization shall be enabled on all lanes.
For 64b/66b encoding - Block synchronization shall be enabled.
- 2) The receiver status shall be RcvdTS1 when at least one complete and error free TS1 or TS2 ordered-set is detected in one or more physical lanes.
- 3) The training sequence ordered-set lane number shall not be checked.

5.8.7.3 RxCMD = ENCONFIG

When the receiver is commanded to enable auto-configure, the receiver will attempt to configure the port as restricted by port capability and by capabilities enabled by management commands (See [Section 5.6.1 on page 165](#)). When commanded to enable auto-configure, the receiver will implement the following rules:

- 1) The receiver shall only attempt to configure the link to speeds enabled by the Link-SpeedEnabled and LinkSpeedExtEnabled variables. (Auto-configuration of link speed is supported under Rel. 1.2 Enhanced Signaling, but not supported by legacy devices).
- 2) The receiver may optionally correct inverted receiver data.
- 3) The receiver shall verify proper lane polarity using the training sequence data (symbols 3 through 16 of both TS1 and TS2) in the received training sequence. If the lane polarity is not correct (or cannot be corrected), the receiver shall not report RxTrained.
- 4) The receiver may optionally correct reversed lanes.
- 5) The receiver shall verify proper lane order using the lane number symbol (the second symbol in both TS1 and TS2) in the received training sequence. If proper lane ordering is not present (or cannot be corrected), the port shall not report Rx-Trained.

- 6) The receiver shall only attempt to configure the link to widths enabled by the LinkWidthEnabled variable.
 - a) When the 12x width is enabled, the receiver shall attempt symbol synchronization on twelve physical lanes (11-0) and shall verify that all twelve lanes are receiving TS1 or TS2 ordered-sets.
 - b) When the 8x width is enabled, the receiver shall attempt symbol synchronization on eight physical lanes (7-0) and shall verify that all eight lanes are receiving TS1 or TS2 ordered-sets.
 - c) When the 4x width is enabled, the receiver shall attempt symbol synchronization on four physical lanes (3-0) and shall verify that all four lanes are receiving TS1 or TS2 ordered-sets.
 - d) When the 1x width is enabled, the receiver shall attempt symbol synchronization on physical lane (0) and shall verify that it is receiving TS1 or TS2 ordered-sets.
- 7) The receiver shall use widest enabled and verified link width for completion of link training and report that width as the configured link width.
- 8) The receiver shall use the sixteen symbol long TS1 or TS2 as reference for link de-skew operations.
- 9) The receiver shall be capable of de-skewing a minimum of six symbol times of total link skew.
- 10) After successful link de-skew, the receiver shall receive eight (8) consecutive error free TS1 or TS2 (all TS1, all TS2 or TS1 followed by TS2) ordered-sets or blocks simultaneously on all configured lanes before reporting RxTrained status.

5.8.7.4 RxCMD = ENDESKW

When the receiver is commanded to enable de-skew, it will implement the following rules:

- 1) For 8b/10b encoding- The receiver shall use the sixteen symbol long TS1 or TS2 as reference for link de-skew operations.

For 64b/66b encoding when RS-FEC is not enabled - The receiver shall use the 66 bit long SKP block as reference for link de-skew operations. A receiver may optionally use the 66 bit TS1 or TS2 blocks as reference for link de-skew operations.

For 64b/66b encoding when RS-FEC is enabled - The receiver shall enable the lane alignment lock and FEC codeword lock state machines (set the align_enable to true), use the alignment sequence to acquire lane lock, de-skew, and FEC codeword lock.
- 2) For 8b/10b encoding the receiver shall be capable of de-skewing a minimum of six symbol times of total link skew.

For 64b/66b encoding when RS-FEC is not enabled, the receiver shall be capable of de-skewing a minimum of two block times of total skew.

For 64b/66b encoding when RS-FEC is enabled, the receiver shall be capable of de-skewing a minimum of 132 bit time of total skew.
- 3) For 64b/66b encoding if FEC_en is true the receiver shall acquire FEC block lock.

- 4) After successful link de-skew, the receiver shall receive eight (8) consecutive error free TS1 or TS2 (all TS1, all TS2 or TS1 followed by TS2) ordered-sets simultaneously on all configured lanes before reporting RxTrained status.

NOTE: For 64b/66b encoding when Fire-Code FEC is enabled, as indicated by FEC_en, the receiver must first acquire FEC block lock before de-skewing the lanes.

5.8.7.5 RxCMD = WAITTS2

This command instructs the receiver to confirm configuration of the remote port. When the remote port configures its receiver, it starts to transmit TS2 ordered-sets. When commanded to wait for TS2, the receiver shall implement the following rules:

- 1) The receiver shall ignore data on the lanes that are not included in the configured link width (unused).
- 2) The receiver shall ignore properly formatted TS1 and SKIP ordered-sets.
- 3) The receiver shall receive eight (8) consecutive identical error free and properly lane ordered TS2 ordered-sets simultaneously on all configured lanes before reporting RcvdTS2 status.

5.8.7.6 RxCMD = WAITTS3

The WaitTS3 command instructs the receiver to determine the advertised speed of the remote port. When the far port has finished sending and receiving idle data it starts to transmit TS3 ordered-sets. When commanded to wait for TS3, the receiver shall implement the following rules:

- 1) The receiver shall ignore data on lanes that are not included in the configured link width (unused).
- 2) The receiver shall receive eight (8) consecutive identical error free TS3 ordered sets simultaneously on all configured lanes before reporting RcvdTS3 status.
- 3) The receiver shall ignore error free idle data and SKIP ordered-sets
- 4) The Receiver shall compare the link speed indicator from the received TS3 ordered-sets with the link speed indicator from the links transmitted TS3 ordered-sets. The highest common speed will be saved and passed to the transmitter for use in the transmission of further data.

5.8.7.7 RxCMD = WAITIDLE

This command instructs the receiver to confirm reception of the final handshake necessary to complete the link training process. When the far port confirms proper configuration of the near port, it starts to transmit the link idle data pattern. When commanded to wait for idle data, the receiver shall implement the following rules:

- 1) The receiver shall ignore properly formatted TS2 and SKIP ordered-sets.
- 2) For 8b/10b encoding the receiver shall assert RcvdIdle after reception of at least eight (8) error free symbol times of link idle data sequence on all configured lanes (i.e. eight (8) per lane).

For 64b/66b encoding the receiver shall assert RcvdIdle after reception of at least eight (8) error free idle or FTB blocks on all configured lanes.

5.8.7.8 RxCMD = WAITPRBS23

This command instructs the receiver to test whether the incoming bit stream is a valid PRBS23 as described in [Section 5.8.4.6](#). When commanded to wait for PRBS23 the receiver shall implement the following rules:

- 1) If the bit stream is a valid PRBS23 the receiver shall assert the RcvdPRBS status.
- 2) After each PRBS23 error the receiver shall de-assert the RcvdPRBS status for 1024 UI.

5.8.7.9 RxCMD = WAITPRBS11

This command instructs the receiver to test whether the incoming bit stream is a valid PRBS11 as described in [Section 5.8.4.6](#). When commanded to wait for PRBS11 the receiver shall implement the following rules:

- 1) If the bit stream is a valid PRBS11 the receiver shall assert the RcvdPRBS status.
- 2) After each PRBS11 error the receiver shall de-assert the RcvdPRBS status for 1024 UI.

5.8.7.10 RxCMD = ENABLE

This command enables the normal receiver operation. The receiver will implement the following rules:

- 1) Received packets shall be transferred to the upper layers of the protocol (link layer).
- 2) Link error detection shall be enabled. (See [Section 5.9 on page 235](#))
- 3) If Rel. 1.2 Enhanced Signaling capabilities are implemented,
 - a) detection of Heartbeat SND Ordered Sets shall be enabled, and
 - b) detection of Heartbeat Errors shall be enabled.

5.8.7.11 RxCMD = WAITTS-T

This command enables the receiver to detect the TS-T ordered set. The receiver will implement the following rules:

- 1) The receiver status shall be RcvdTS-T when at least one complete and error free TS-T ordered-set is detected in one or more physical lanes.

5.8.7.12 RxSTATUS = RcvdTS1

The received TS1 status is valid at any time TS1 is received. The receiver status RcvdTS1 is defined in [Section 5.8.7.2, "RxCMD = WaitTS1," on page 231](#).

5.8.7.13 RxSTATUS = RxTRAINED

The receiver trained status is valid only when the RxCMD is EnConfig or EnDeSkew. The receiver status RxTrained is defined in [Section 5.8.7.4, "RxCMD = EnDeSkew," on page 232](#).

5.8.7.14 RxSTATUS = RcvdTS2

The received TS2 status is valid only when the RxCMD is WaitTS2. The receiver status RcvdTS2 is defined in [Section 5.8.7.5, “RxCMD = WaitTS2,” on page 233](#).

5.8.7.15 RxSTATUS = RcvdTS3

The received TS3 status is valid only when the RxCMD is WaitTS3. The receiver status RcvdTS3 is defined in [Section 5.8.7.6, “RxCMD = WaitTS3,” on page 233](#).

5.8.7.16 RxSTATUS = RcvdTS-T

The received TS-T status is valid only when the RxCMD is WaitTS-T. The receiver status RcvdTS-T is defined in [Section 5.8.7.11, “RxCMD = WaitTS-T,” on page 234](#).

5.8.7.17 RxSTATUS = RcvdIdle

The received idle status is valid when the RxCMD is WaitIdle and in states where a received Idle causes a state transition. The receiver status RcvdIdle is defined in [Section 5.8.7.7, “RxCMD = WaitIdle,” on page 233](#).

5.8.7.18 RxSTATUS = RcvdPRBS

The received PRBS status is valid when the RxCMD is either WaitPRBS23 or WaitPRBS11. The receiver status RcvdPRBS is defined in [Section 5.8.7.8, “RxCMD = WaitPRBS23,” on page 234](#) and [Section 5.8.7.9, “RxCMD = WaitPRBS11,” on page 234](#).

5.8.7.19 RxSTATUS = RxMajorError

Receiver Major Error status can be asserted by multiple sources. The receiver status RxMajorError is defined in [Section 5.9.4, “Major Link Physical Errors Events,” on page 238](#).

5.8.7.20 RxSTATUS = RxHeartbeatError

Receiver Heartbeat Error status can be asserted by multiple sources. The receiver status RxMajorError is defined in [Section 5.14.2, “Heartbeat Error Handling,” on page 254](#).

5.9 LINK PHYSICAL ERROR HANDLING

This section describes link error detection and link error recovery implemented in the Link/Physical layer.

The Link/Physical layer does not interpret packet payloads, packet framing, or CRCs. These errors are detected at the Link Layer Protocol described in Volume 1.

C5-18: All ports shall implement link physical error handling as defined in [Section 5.9, “Link Physical Error Handling,” on page 235](#).

5.9.1 LINK PHYSICAL ERRORS EVENTS

Link/Physical errors stem from two fundamental sources: link bit errors and protocol violations. Bit errors may appear as 8b/10b coding violations, running disparity violations, incorrect but valid 8b/10b code groups, 64b/66b sync header violation, 64b/66b control block violation, and FEC correctable and uncorrectable codewords. Bit errors which re-

sult in incorrect but valid code groups may be detected as protocol errors or as CRC errors when checked by the upper layers of the protocol. Burst errors may severely corrupt multiple code groups on one or more lanes. These major error events may result in multiple coding violations, protocol errors, loss of lane to lane de-skew, or loss of symbol/block synchronization. Protocol errors may be the result of simple bit errors, or they may be the result of some other event. Receiver-detected errors are handled in several basic ways:

- 1) Minor error events which do not significantly impact link/physical layer processing shall be marked and forwarded to the upper layers.
 - 2) When minor error events occur simultaneously on multiple lanes, they shall be treated as a single minor error event.
- 3) Major error events shall result in a link error event which triggers link/physical error recovery.

The error threshold logic described in [Section 5.9.3 on page 237](#) monitors minor error events and may trigger a major event if the rate is too high.

5.9.2 MINOR LINK PHYSICAL ERRORS EVENTS

Minor error events are input to the error threshold logic described in [Section 5.9.3 on page 237](#) and to the **SymbolErrorCounter** described in [Section 5.6.3](#). The following errors are counted as minor error events:

- 1) Invalid 8b/10b codes groups and running disparity errors shall be counted as minor error events. (See [Section 5.2](#)).
- 2) Unsupported or disabled valid code groups shall be counted as minor error events. (See [Section 5.5.1.2](#) and [Section 5.5.1.2.1](#)).
- 3) Start or End Packet Delimiter in the wrong lane of a multi-lane link shall be counted as a minor error event. (See [Section 5.7.4](#), [Section 5.7.5](#), and [Section 5.7.6](#)).
- 4) Any control symbol/block within the boundaries of a packet shall be counted as a minor error event.
- 5) For 8b/10b encoding: PAD symbols on a non-12x and non-8x link shall be counted as minor error events. (See [Section 5.7.5](#) and [Section 5.7.6](#)).
- 6) For 8b/10b encoding: On an 8x or 12x link, PAD symbols not preceded by a End of Packet Delimiter shall be counted as minor error events. PAD symbols in the wrong lane shall also be counted as minor error events. (See [Section 5.7.5](#), [Section 5.7.6](#) and [Section 5.12.3](#))
- 7) Invalid 64b/66b sync header shall be counted as minor error events. (See [Section 5.3.2, "Checking the validity of a received block," on page 100](#))
- 8) Unsupported or unknown control block types shall be counted as minor error events. (See [Section 5.3.2, "Checking the validity of a received block," on page 100](#))
- 9) For 64b/66b encoding any control block within the boundaries of a packet shall be counted as a minor error event.

- 10) For 64b/66b encoding end of packet block outside the boundary of a packet shall be
1 counted as a minor error event.

2 Note that for 64b/66b encoding with EDPL (see [Section 5.3.4, “Error Detection Per](#)
3 [Lane,” on page 104](#)), EDPL errors are not counted as minor link physical errors because
4 any errors that do occur will be caught and counted using other mechanisms, and be-
5 cause EDPL bytes are only transmitted infrequently, in SKIP blocks, so they don’t provide
6 an accurate count of link physical transmission errors.

5.9.3 LINK PHYSICAL ERROR THRESHOLD ALGORITHM

To detect an excessive number of minor errors (see [Section 5.9.2](#) above), an error threshold algorithm is implemented. The threshold shall be set to detect an error ratio of four (4) or more minor errors within sixteen (16) symbol/block times. Both the “leaky bucket algorithm” and “sliding window algorithm” implementations are sufficient. The error threshold function will implement the following rules:

- 1) The error threshold shall be enabled only when the link training state machine is in
the LinkUp state.
- 2) The error threshold shall be disabled and cleared when the link training state ma-
chine is not in the LinkUp state.
- 3) The error threshold shall be disabled and cleared when RS-FEC is enabled. When
RS-FEC is enabled the detection of an excessive number of errors is provided by
the RS-FEC symbol error rate threshold. See [Section 5.9.4](#).

The following implementation notes describe leaky bucket and sliding window error threshold algorithms.

Implementation Note:

The error threshold counter which employs the “leaky bucket algorithm” should implement with the following rules:

- 1) When a minor error is detected, the error threshold counter is incremented.
- 2) The error threshold counter is decremented every sixteen (16) symbol/block times.
- 3) The decrement event may be a free-running symbol time counter that is not synchronized to any specific link state or error condition.

An error threshold event will be reported when the error threshold counter is value is four (4) or greater.

Implementation Note:

The error threshold counter which employs the “sliding window algorithm” should implement with the following rules:

- 1) Logic tracks the minor error status of the most recent 16 symbol/block times.
- 2) An error threshold event will be reported when four (4) or more minor errors are present in the most recent 16 symbol/block time of history

5.9.4 MAJOR LINK PHYSICAL ERRORS EVENTS

Major error events trigger the link error recovery process of the Link Training State Machine. These error events are not counted directly. However, successful error recovery attempts are counted by the ***LinkErrorRecoveryCounter*** and failed error recovery attempts are counted by the ***LinkDownedCounter***. Both counters are defined in [Section 5.6.3, “Port Performance Counters,” on page 173](#). The following errors events will start the link error recovery process:

- 1) Training Sequence one or two (TS1 or TS2) received in the linkup state shall trigger link error recovery. (TS1 indicates that the remote port has initiated the link error recovery process)
- 2) The assertion of LinkPhyRecover from the Link Layer shall trigger link error recovery.
- 3) Loss of symbol(s)/block(s) caused by elastic buffer overflow or underflow shall trigger link error recovery.
- 4) Loss of block lock on any lane shall trigger link error recovery.
- 5) The absence of 4 expected SKIP ordered-sets should trigger link error recovery.
- 6) Clear loss of lane-to-lane de-skew should trigger link error recovery.
- 7) A minor error threshold event should trigger link error recovery.
- 8) Loss of RS-FEC lock due to a detection of a 2 uncorrectable RS-FEC codewords out of 256 RS-FEC codewords should trigger link error recovery. Both the “leaky bucket algorithm” and “sliding window algorithm” implementations are sufficient for detection of the RS-FEC lock loss.

5.9.5 HEARTBEAT ERROR

A Heartbeat Error error event triggers a transition to the default state in the Link Training State Machine. These error events are counted by the ***LinkDownedCounter***. The following error event is a Heartbeat Error event:

- 4 HEARTBEAT transmission periods (400 ms) elapse without receiving valid HEARTBEAT ACK ordered-sets simultaneously on all active lanes.

The characteristics of a valid HEARTBEAT ordered-set are described in [Section 5.5.2.5, “Link Heartbeat Ordered-Set \(HRTBT\),” on page 155](#) and [Section 5.5.3.10, “Link Heartbeat Ordered-Set \(HRTBT\),” on page 163](#). A valid Link Heartbeat ordered-set has the fol-

lowing characteristics:

- For 8b/10b encoding: Error-free reception of 8b/10b symbols on all configured lanes.
- For 64b/66b encoding: Reception of both consecutive error free HRTBT blocks.
- For 8b/10b encoding: Lane IDs match the lanes on which they were received.
- Opcode of either 5Dh (SND) or ACh (ACK).
- Received GUID on a SND HRTBT that doesn't match the receive port's own GUID. Matching SND GUID is interpreted as transmitter/receiver crosstalk on SND transmission from a port transmitter on the same device.
- Received GUID and PortNum on an ACK HRTBT that matches the receive port's own GUID and PortNum. Mismatched ACK GUID and Portnum is interpreted as transmitter/receiver crosstalk on ACK transmission.

5.10 INTERNAL SERIAL LOOPBACK

The Internal Serial Loopback is optional and is not required on all ports. Internal serial loopback allows a port to receive its transmitted data stream without the need for external support. The internal serial loopback function is intended for self-test and fault isolation use only.

o5-4: All ports that implement internal serial loopback option shall implement it as defined in [Section 5.10, “Internal Serial Loopback,” on page 239](#).

The following rules apply to the internal serial loopback:

- 1) When disabled, the internal serial loopback shall have no effect on port operation.
- 2) When enabled, internal serial loopback shall connect the transmit serial data output to the receive serial data input as close to the chip I/O cells as is practical.
- 3) When internal serial loopback is enabled, the receive data from the receiver input pins shall be blocked.
- 4) When internal serial loopback is enabled, the chip's transmitter output pins shall be forced to a quiescent condition.

5.11 CLOCK TOLERANCE COMPENSATION

Each end of a 1x link (or each physical lane on 4x, 8x and 12x links) utilizes a transmitter and a receiver. The transmitter logic operates using a tightly controlled reference clock called the “transmit clock”. Likewise, the receiver logic operates using a clock recovered from the incoming bit stream called the “receive clock”. Once the link is trained, the recovered receive clock operates at the same frequency as the transmit clock at the other end of the link.

The UI_D parameter in *InfiniBand Architecture Specification, Volume 2, Table 48 Host Driver Characteristics for 2.5 Gb/s (SDR)* on page 285 specifies the transmit clock accu-

racy as +/- 100 ppm (parts per million). The worst-case frequency difference between the transmit and receive clocks of a link occurs when one of the transmitters is at the +100 ppm and the other one is at the -100 ppm tolerance, resulting in a 200 ppm difference. In other words, the transmit and receive clocks can shift by as much as one clock period every 5000 clocks.

A common design practice is to clock most of the receive path logic in the transmit clock domain. This is accomplished by using an “elastic buffer” in the very early stages of the receive path. The elastic buffer compensates for the differences between the transmit and receive clock domains by dropping or inserting symbols. The input side of the elastic buffer operates in the receive clock domain, and the output side operates in the transmit clock domain.

To perform the compensation function, the elastic buffer needs to identify safe zones in the incoming symbol sequence to insert or drop a symbol. This zone is provided by the “SKIP ordered-set”.

C5-19: All ports shall implement clock tolerance compensation as defined in [Section 5.11, “Clock Tolerance Compensation,” on page 239](#).

5.11.1 TRANSMITTER “SKIP” REQUIREMENTS

The transmitters are required to transmit SKIP ordered-sets periodically, complying with the following rules:

For 8b/10b encoding:

- 1) The SKIP ordered-set shall be scheduled for insertion at least once every 4608 symbol times.
- 2) The SKIP ordered-set shall be scheduled for insertion at most once every 4352 symbol times.
- 3) A scheduled SKIP ordered-set shall be inserted at the next packet or ordered-set boundary.

For 64b/66b encoding:

- 1) The SKIP ordered-set shall be scheduled for insertion at least once every 4608 block times.
- 2) The SKIP ordered-set shall be scheduled for insertion at most once every 4352 block times.
- 3) A scheduled SKIP ordered-set shall be inserted at the next packet or ordered-set boundary.

Note: a SKIP ordered-set in 64b/66b encoding as defined in - [Section 5.5.3.6, “Skip Block \(SKP\),” on page 162](#) - is 1-5 SKP consecutive blocks. Transmitter must transmit a SKIP ordered-set composed of 3 consecutive SKP blocks.

5.11.2 RECEIVER “SKIP” REQUIREMENTS

The receivers are required to receive and process SKIP ordered-sets periodically, complying with the following rules:

For 8b/10b encoding:

- 1) The receivers shall recognize received SKIP ordered-sets that are comprised of one (1) comma (COM) symbol followed by one (1) to five (5) skip (SKP) symbols.
- 2) The receivers shall be tolerant to receive and process SKIP ordered-sets at an average rate of once in every 4352 to 4608 symbol times.
- 3) The receivers shall be tolerant to receive and process SKIP ordered-sets separated from each other at least 128 symbol times -- measured as the distance between the leading comma (COM) symbols.
- 4) The receivers shall be tolerant to receive and process SKIP ordered-sets separated from each other at most 8832 symbol times -- measured as the distance between the leading comma (COM) symbols.

For 64b/66b encoding:

- 1) The receivers shall recognize received SKIP ordered-sets that are comprised of one (1) to five (5) skip (SKP) blocks.
- 2) The receivers shall be tolerant to receive and process SKIP ordered-sets at an average rate of once in every 4352 to 4608 block times.
- 3) The receivers shall be tolerant to receive and process SKIP ordered-sets separated from each other at least 128 block times, measured as the distance between the first SKP blocks.
- 4) The receivers shall be tolerant to receive and process SKIP ordered-sets separated from each other at most 8832 block times, measured as the distance between the first SKP blocks.

5.12 RETIMING REPEATERS

The InfiniBand™ Architecture allows for the use of “retiming repeaters” to recover from potentially weakened signal strength and built-up jitter between two end nodes of a link.

There are two types of retiming repeater: “SKIP ordered-set dependent” retiming repeaters, which use the SKIP ordered-set to compensate for frequency difference, and “transparent” retiming repeaters, which do not depend on or use the SKIP ordered-set, but operate at a single common frequency for both transmission and reception.

This section describes, in general, the operation of SKIP ordered-set dependent retiming repeaters, except where transparent retiming repeaters are specifically mentioned. Both types of retiming repeater reset jitter and meet the signaling requirement specified in (*In-*

finiBand Architecture Specification, [Volume 2, Chapter 6: High Speed Electrical Interfaces](#).

Implementation Note:

The major difference between the two types of retiming repeaters is in whether the retiming repeater uses an internal clock, or uses a clock recovered from the received data stream, for transmitting the repeated data. A SKIP ordered-set dependent retiming repeater will typically have its own internal clock, with the usual +/- 100 ppm frequency tolerance, and may need to insert or remove SKIP symbols to compensate for frequency differences with the originating port. A transparent retiming repeater will typically transmit the repeated data using a clock recovered from the received data.

Transparent retiming repeaters will typically be more generally useful, since they are not dependent on the specific InfiniBand Link/Phy protocol's use of the SKIP ordered-set.

C5-20: All retiming repeaters shall implement functions as defined in [5.12 Retiming Repeaters](#).

The following general rules apply to retiming repeaters:

- 1) Retiming repeaters shall reset the jitter budget.
- 2) Not more than two SKIP ordered set-dependent retiming repeaters shall be allowed between two protocol-aware ports.
- 3) Retiming repeaters **may** join dissimilar physical media such as copper-to-fiber optic links.
- 4) Retiming repeaters are not in-band-addressable devices. Hence they cannot be managed through in-band management messages.

5.12.1 RETIMING REPEATER FUNCTIONS

[Figure 67](#) depicts a block diagram of a conceptual retiming repeater. This figure is provided as a visual aid to help explain the fundamental functions of retiming repeaters.

Retiming repeaters are fairly simple devices. A 1x retiming repeater consists of two ports. It simply transmits every symbol/block received on one port to the other port. The receiver circuitry on each port operates from its own recovered receive clock (Rx Clock) domain. However, the transmit circuitry operates on a locally generated transmit clock (Tx Clock) domain, which may have a different frequency than the frequency of the incoming data. In order to compensate for the differences in the receive and transmit clock domains, elastic buffers are used in each direction. Transparent retiming repeaters transmit the data using the clock recovered from the received data, and have less requirement for elastic buffers.

Retiming repeaters are not link-protocol-aware devices. In other words, they do not recognize link and data packets. However, SKIP ordered-set dependent retiming repeaters do recognize two kinds of ordered-sets in the incoming symbol stream:

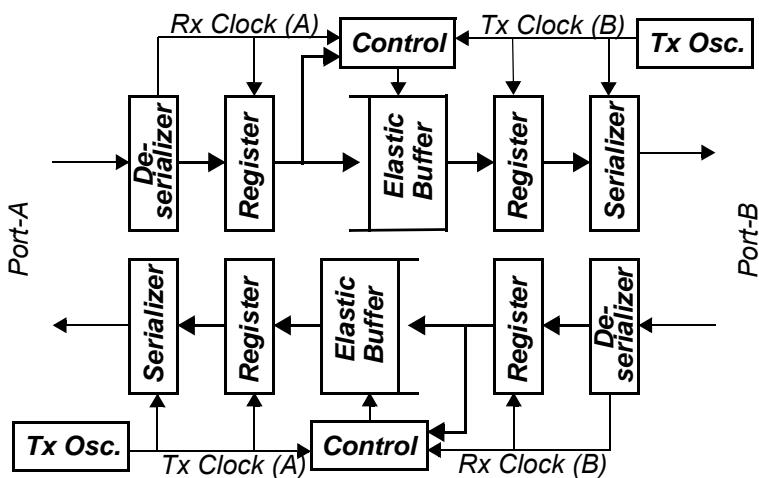


Figure 67 A Conceptual Retiming Repeater Block Diagram

- SKIP ordered-set
- TS1 and TS2 Training Sequence ordered-sets at 8b/10b encoded. The SKIP ordered-set dependent retiming repeater may also optionally recognize incoming 64b/66b encoding TS1 and TS2

SKIP ordered-sets are both recognized in the incoming symbol stream and used in the elasticity operations performed in each direction. Elasticity operations performed by the retiming repeaters are specified in [Section 5.12.2](#) later in this chapter.

During Link Training, repeated training sequence ordered-sets are transmitted between two end points. In 8b/10b encoding repeaters use the periodic comma (COM) symbols in this stream to detect symbol boundary misalignment, in 64b/66b encoding repeaters may acquire block lock, PRBS23 lock, or PRBS11 lock.

5.12.2 CLOCK TOLERANCE COMPENSATION

SKIP ordered-set dependent retiming repeaters perform clock tolerance compensation during SKIP sequences. The 8b/10b encoded SKIP sequence -- as observed by repeaters -- is comprised of one comma (COM) symbol followed by two to four skip (SKP) symbols, the 64B66B encoded SKIP sequence -- as observed by repeaters -- is comprised two to four skip (SKP) blocks. Repeaters are allowed to insert or delete one skip symbol/block in this ordered-set in order to compensate for the differences between the receive and transmit clock frequencies.

SKIP ordered-set dependent retiming repeaters shall use the following rules to insert or delete a skip symbol/block in a SKIP ordered-set:

- 1) For 8b/10b encoding when necessary, the retiming repeaters shall insert a skip (SKP) symbol after the comma (COM) symbol within the current SKIP ordered-set. For 64b/66b encoding when necessary, the retiming repeater shall insert a skip

(SKP) block immediately after the first skip block of the ordered-set. The inserted skip shall have the same EDPL value as the other SKP block in the ordered set.

- 2) For 8b/10b when necessary, the retiming repeaters shall delete any skip (SKP) symbol after the comma (COM) symbol in the current SKIP ordered-set. For 64b/66b when necessary, the retiming repeaters shall delete any skip (SKP) block from the current SKIP ordered-set.

Implementation Note:

Skip ordered-set dependent retimers when working at 64b/66b encoding will need to be able to either implement FEC encoder and decoder or to disable FEC on a system level.

In order to compensate for clock tolerance SKP blocks need to be added or deleted by the retimer, when FEC is enabled the retimer must be able to lock on the FEC blocks, decode the FEC block, detect the skip ordered-set, compensate for the clock differences by adding/deleting SKP blocks, encode the new FEC block and transmit it.

5.12.3 ERROR HANDLING CONSIDERATIONS

Retiming repeaters shall not check for code violations (i.e. decode and disparity errors). Symbols received on one port with or without transmission errors are simply transmitted by the other port.

Retiming repeaters shall not check for block violation (i.e. sync header of 00 or 11 or unknown block type). Blocks received on one port with or without transmission errors are simply transmitted by the other port.

The elastic buffers are not expected to underflow or overflow. However, when these conditions are detected, these rules apply:

- 1) When an underflow condition is detected, the retiming repeaters shall insert a pad (PAD) symbol at 8b/10b encoding or idle block at 64b/66b encoding.
- 2) When an overflow condition is detected, the retiming repeaters shall replace two arbitrary but consecutive symbols with a pad (PAD) symbol at 8b/10b encoding, and shall replace any arbitrary block with an Idle block.

5.12.4 SYMBOL BOUNDARY ALIGNMENT

When protocol-aware nodes detect link errors (including loss-of-symbol synchronization), they initiate error recovery (see [Section 5.9 on page 235](#)) by sending training sequences that contain the commas needed for symbol synchronization. Retiming repeaters cannot initiate training sequences on their own. Instead, for 8b/10b encoding, they detect periodic unaligned comma (COM) symbols within this sequence to determine the loss-of-symbol boundary alignment, and for 64b/66b encoding, they detect a loss of block lock.

The following rules define the symbol boundary alignment process used by retiming repeaters:

- 1) Periodic comma symbols (COM) in TS1, TS2 or SKIP ordered-sets shall be used to
2 acquire symbol boundary synchronization.
3
- 4) SKIP ordered-set dependent retiming repeaters shall reacquire symbol boundary
5 synchronization when three consecutive misaligned comma symbols are detected.
6 A misaligned comma symbol is the 0011111 or 1100000 symbol pattern which does
7 not begin at the current symbol boundary.
8
 - 9 • SKIP ordered-set dependent retiming repeaters **may** use the 7-bit comma bit
10 pattern 1100000 or 0011111 to acquire alignment, rather than the full 10-bit com-
11 ma symbol pattern.
- 12) Retiming repeaters shall disable their symbol boundary synchronization circuitry
13 when five consecutive instances of aligned comma symbols (COM) are detected.

Retiming devices acquire block lock in the same way that a end port acquire block lock -
see [Section 5.3.3, "Block synchronization," on page 101](#)

5.12.5 MULTI-LANE REPEATER CONSIDERATIONS

Multi-lane retiming repeaters are used by multi-lane (4x, 8x and 12x) links. Like protocol aware ports, retiming repeaters are required to use a common clock source for all lanes in the link. The logic within a retiming repeater is not required to synchronize lane to lane operation. Multi-lane retiming repeaters can be shared by multiple links that use a subset of the retiming repeater's lanes. For example, a 4x retiming repeater can be used by up to 4 1x links. The 12x retiming repeater can be used by multiple combinations of 1x, 4x, and 8x links.

The following rules apply to operation of multi-lane retiming repeaters:

- 1) For each direction all lanes of multi-lane retiming repeaters shall have a common
2 transmit clock source.
3
- 4) The lanes in opposed directions **may** have independent transmit clock sources.
5
- 6) Each lane of multi-lane retiming repeaters shall operate independently of the other
7 lanes.
8
- 9) Multi-lane retiming repeaters **may** implement lane to lane de-skew.
10

5.12.6 POWER STATE CONSIDERATIONS

The retiming repeaters are not protocol-aware devices. Hence they cannot be managed through in-band management packets. However, they are expected to be managed by either:

- 1) the Management Entity of the chassis of which they are part, or
2) the Management Entity of the InfiniBand module of which they are part.

The retiming repeaters that are managed using one of these schemes are also said to be “proxy-managed.”

During normal operation, a neighboring node may transition to $X_{Standby}$ or X_{Sleep} state. In these states, the node drives the link to “quiescent” state. Similarly, when the neighboring node transitions to the $X_{Polling}$ state, the link will be driven to its “quiescent” state periodically. Retiming repeaters are expected to detect the link going to “quiescent” state at its receiver on either side and to propagate this link “quiescent” state on the transmitter of the other side.

When the neighboring node transitions to X_{On} state, the link goes from “quiescent” to “active” state. Similarly, when the neighboring node transitions to the $X_{Polling}$ state, the link will transition from “quiescent” to “active” state periodically. The retiming repeaters are expected to detect this link state change on its receiver at either side and to propagate it to the transmitter on the other side.

Proxy managed retiming repeaters shall comply with power states and behavior defined in *InfiniBand Architecture Specification, Volume 2-DEPR, Section 6.3, “Port Power Management States”*. Additionally, operation of the retiming repeaters when propagating “quiescent” and “active” link activity are governed by the following rules:

- 1) Retiming repeaters shall detect “quiescent” to “active” transition on one side and propagate that state to the other side within 100 microseconds.
- 2) Retiming repeaters shall detect “active” to “quiescent” transition on one side and propagate that state to the other side within 100 microseconds.

5.13 FINE TUNING

For FDR and higher data rate devices, a fine tuning algorithm is defined to allow for minor transmitter equalization adjustments while the link is up, to account for temperature variation or improve on very high bit error ratios.

Fine tuning is allowed during LinkUp as a background process while normal data and link packets are flowing. Either link partner is allowed to initiate a sequence of commands to adjust their link partner’s transmitter coefficients. The incremental change of each fine tuning request is intended to be small enough that link errors are avoided and traffic flow is not affected.

This section covers the parametric exchange protocol but leaves the management of the coefficient changes and receiver evaluation to vendor-specific implementations.

o5-5: All ports claiming compliance with InfiniBand Rel. 1.3 shall implement the Fine Tuning capability as defined in [Section 5.13, “Fine Tuning,” on page 246](#).

5.13.1 FINE TUNING BLOCK

A unique 64b/66b control block was created to support the fine tuning algorithm. This control block is used to exchange tuning adjustment requests from the receiver to its link

partner's transmitter, and then communicate back the acceptance or rejection of the requested tuning change. The fields below are part of the Fine Tuning block, described in [Section 5.5.3.9](#).

For FIR tap definition see [Section 5.16, "FDR and Higher Rate Transmitter Equalization," on page 256](#), and [Section 6.6.5.1, "FDR Transmitter Equalizer Implementation," on page 296](#).

FTB CMD (Byte 2)

- [1:0] - C₊₁ FIR Tap
- [3:2] - C₀ FIR Tap
- [5:4] - C₋₁ FIR Tap
- [7:6] - Reserved

Each 2 bit field:

- 00 - No Change
- 01 - Increase
- 10 - Decrease
- 11 - Reserved

FTB STAT (Byte 3)

- [1:0] - C₊₁ FIR Tap
- [3:2] - C₀ FIR Tap
- [5:4] - C₋₁ FIR Tap
- [7:6] - Reserved

Each 2 bit field:

- 00 - Ready
- 01 - Change Accept
- 10 - Change Reject
- 11 - Reserved

FTB Lane Req (Byte 5)

[7:0] - Lane Identifier - Uses the Lane ID mapping described in [Table 24 on page 143](#).

5.13.2 FINE TUNING RULES

The following rules should be employed when implementing the fine tuning algorithm:

- Fine tuning must be enabled in the TS3 - see [Section 5.5.2.4, "Training Sequence Three Ordered-Set \(TS3\)," on page 143](#).

- Fine tuning blocks have the same transmit priority as IDLE blocks. If a fine tuning block is scheduled for transmission, it should be sent in place of IDLE blocks.
- Requests for tuning changes apply only to the direction of the requesting receiver and responding transmitter. Although both directions of the link can be performing fine tuning simultaneously, the tuning adjustments are independent.

An increment or decrement request shall result in a transmitter equalization adjustment in the C_{-1} , C_0 , or C_{+1} tap weight of between 0.05 and 0.0083, as specified in [Table 53 on page 294](#), when measured using the algorithm defined in [Section 5.16, "FDR and Higher Rate Transmitter Equalization," on page 256](#).

5.13.3 FINE TUNING STATE MACHINE

Each lane on each port has both Initiator and Responder State Machines. These state machines handle the interlocked handshake between link partners passing CMD and STAT parameters for the tuning sequence. The Initiator provides the CMD and Lane Req portion of the FTB and starts and ends the FTB interval. The Responder provides the requested STAT for each FTB CMD indicating whether the requested change was accepted or rejected. Since each link partner could be requesting tuning independently, this process is bidirectional; each node's Initiator and Responder state machines could be busy simultaneously.

State transitions are governed by the following rules:

- Transmitter must send a minimum of 16 FTBs with identical Status or identical CMD and lane request in a given state.
- Receiver must receive a minimum of 8 FTBs with identical Status or identical CMD and identical lane request to advance states.

Fine tuning blocks should be scheduled for transmission whenever FTB_ACTIVE is asserted, which can be due to activity in either the Initiator and/or Responder state machines. If FTB_ACTIVE is asserted because the Responder state machine has been activated, but that port does not require any tuning changes, the Initiator state machine remains in the INACTIVE state, and CMD = 'No Change' is placed in the FTB. Similarly, if FTB_ACTIVE is asserted because the Initiator state machine has been activated, but that port has not received tuning change requests from its link partner, the Responder state machine remains in the INACTIVE state and STAT='Ready' is placed in the FTB.

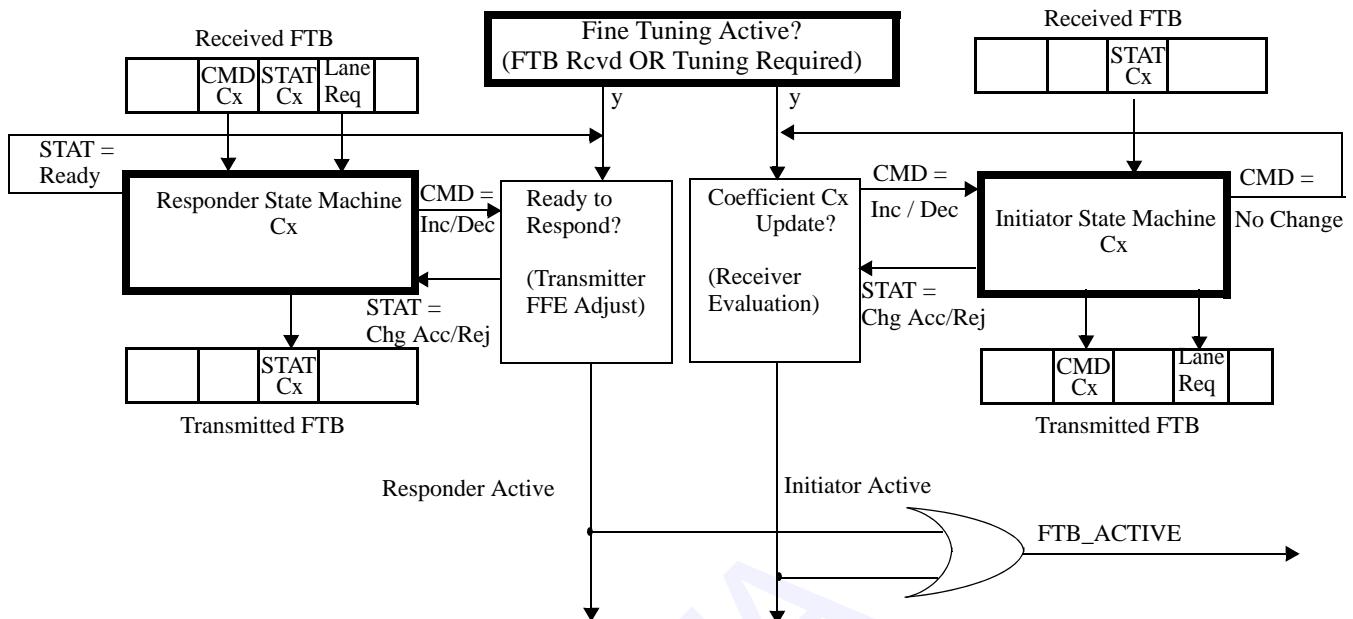


Figure 68 Fine Tuning Flow Diagram

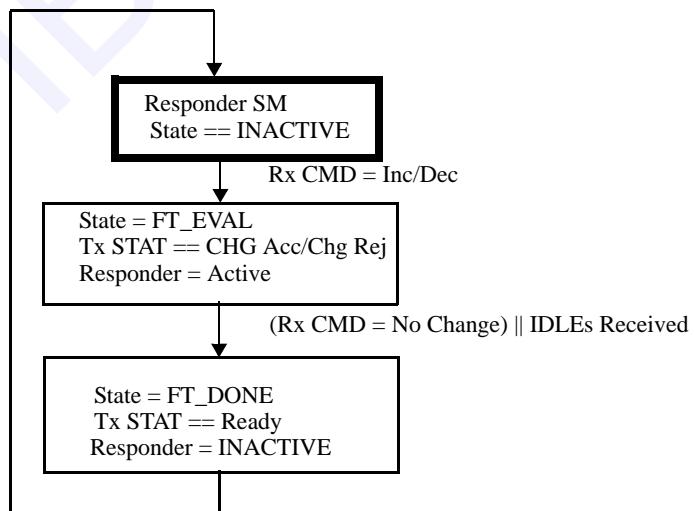
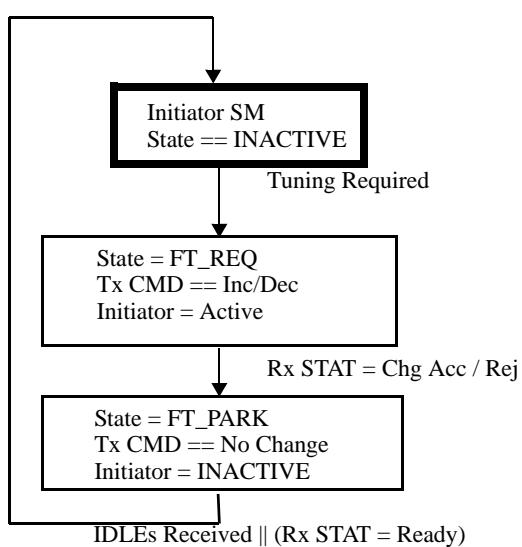


Figure 69 Responder State Machine

**Figure 70 Initiator State Machine**

5.13.4 FINE TUNING OPERATION

Fine Tuning is negotiated during the exchange of TS3 ordered sets and if enabled shall be allowed after the Training State Machine reaches Config.Idle or LinkUp. The Initiator and Responder state machines operate in parallel and as shown in [Figure 69](#) and [Figure 70](#) above, operate on independent parameters with the received and transmitted FTB. The figures above illustrate ' C_x ' and its associated Initiator/Responder pair used for any of the three FIR coefficients. This method does not prevent an implementation that provides support for all three coefficients simultaneously using three pair of Initiator/Responder state machines.

5.13.4.1 INITIATOR STATE MACHINE OPERATION:

If Fine Tuning is not enabled the Initiator state machine is disabled.

Note: Possible implementation of the above requirement is to disable the transition from INACTIVE State to FT_REQ when Fine Tuning is not enabled.

The Initiator state machine reset state is INACTIVE State.

The Initiator state machine is illustrated in [Figure 70](#).

INACTIVE State:

Initiator:

INACTIVE - initiator sends IDLE blocks when there are no packets scheduled for transmission.

Next state:

If Fine Tuning is enabled and coefficient update is required then next state is **FT_REQ**. Else next state is **INACTIVE** State.

FT_REQ:

Initiator:

ACTIVE- initiator sends Fine Tuning blocks when there are no packets scheduled for transmission.

Tx CMD - Increment/Decrement in the command byte in the FTB is set to the requested coefficient request.

Next state:

If the Rx Status in the received FTBs accepts or rejects the coefficient change then next state is **FT_PARK**. Else next state is **FT_REQ**.

FT_PARK:

Initiator: May have two sub-states:

INACTIVE - initiator sends IDLE blocks when there are no packets scheduled for transmission.

Tx CMD - no change

Next state:

If the Rx Status in the received FTBs is Ready or the initiator receives IDLE blocks next state is **INACTIVE**. Else next state is **FT_PARK**.

5.13.4.2 RESPONDER STATE MACHINE OPERATION:

If Fine Tuning is not enabled the Responder state machine is disabled.

Note: Possible implementation of the above requirement is to disable the transition from INACTIVE State to FT_EVAL when Fine Tuning is not enabled.

The Responder state machine reset state is INACTIVE State.

The Responder state machine is illustrated in [Figure 69](#).

INACTIVE State:

Responder:

INACTIVE - Responder sends IDLE blocks when there are no packets scheduled for transmission.

Next state:

If Fine Tuning is enabled and received Rx CMD is equal to increment or decrement then next state is **FT_EVAL**. Else next state is **INACTIVE** State.

FT_EVAL:

Responder:

ACTIVE - initiator sends FTBs when there are no packets scheduled for transmission.

Tx Status - The responder evaluates the coefficient request and updates the Tx status to Accept or reject in the status byte of the FTB.

Next state:

If the Rx CMD in the received FTBs is equal to No Change, or if IDLEs are received, then next state is **FT_DONE**. Else next state is **FT_EVAL**.

FT_DONE:

Responder:

ACTIVE - Responder sends FTBs when there are no packets scheduled for transmission.

Tx Status - Ready.

Next state:

Next state is **INACTIVE**. Note that the Responder state machine can transition to that state after sending the minimal 16 FTBs with Status Ready.

5.14 LINK HEARTBEAT

At the DDR, QDR, FDR and EDR link bit rates described for Rel. 1.2 Enhanced Signaling and later releases, the lower peak-to-peak received voltage and the consequent lack of a signal detect at the receiver results in higher vulnerability to crosstalk in the receiver logic. It is possible that if a link is disconnected while a port is in the LinkUp state a receiver may interpret crosstalk from its own transmitter as valid data and incorrectly maintain LinkUp status. The link heartbeat function ensures that received data actually originates from the opposite end of a link, rather than from crosstalk from the port's own transmitter.

Link Heartbeats are also used to determine link round-trip latency. Since InfiniBand™ links use a variety of copper and optical physical layer transmission media, a link may be between a few inches and many kilometers long, causing a wide range of link latency values. Prior releases of the InfiniBand™ specification allowed no straightforward way for the link layer logic or for a subnet manager to determine the round-trip latency across a link. Link Heartbeats allow the determination of link round-trip latency, for helping to improve link management and bandwidth allocation algorithms, which is useful for SDR as well as for DDR, QDR, FDR and EDR link rates.

Implementation Note:

Knowledge of link lengths can be useful for link management functions, e.g., to tell whether optical transceivers or retiming repeaters are likely to be present. A 1 km link, for example, may be assumed to have an optical transceiver inserted, even though, since they are simple devices, and are not in-band-addressable, this information can't be determined using in-band management messages.

Knowledge of link lengths is also useful for congestion control and bandwidth maximization mechanisms, since a long link with high link latency may be throttled in bandwidth due to credit starvation if it is not allocated extra buffering and extra credits. A switch chip, for example, with a combination of long and short links on different ports and with a flexible buffering allocation capability, can move buffer space and credit allocation from a short link to a long link, maximizing usable bandwidth on both.

C5-20.2.1: All ports claiming compliance with InfiniBand Rel. 1.2 Enhanced Signaling shall implement link heartbeat functionality as defined by [Section 5.14, “Link Heartbeat,” on page 253](#), at all link speeds.

5.14.1 OPERATION OF LINK HEARTBEATS

Each port capable of Rel. 1.2.1 Enhanced Signaling or 1.3.1 operation will create regularly transmitted HEARTBEAT ordered-sets when connected to another capable port, at any link speed. A HRTBT ordered-set, described in [Section 5.5.2.5, “Link Heartbeat Ordered-Set \(HRTBT\),” on page 155](#) and [Section 5.5.3.10, “Link Heartbeat Ordered-Set \(HRTBT\),” on page 163](#) contains an OpCode identifying it as a SND HEARTBEAT or an ACK HEARTBEAT, along with information identifying its source.

A port schedules for transmission a Link Heartbeat with OpCode=5Dh, indicating SND, once every 100 ms while the port is in the LinkUp state. The SND HEARTBEAT contains the sending port's base GUID and, if applicable, the sending port's switch port number. In the SND HEARTBEAT, the OpCode byte is set to 5Dh.

Upon reception of a valid SND HEARTBEAT the port schedules an ACK (Acknowledge) HEARTBEAT for transmission at the next possible transmission time. The ACK heartbeat is sent with the GUID and switch port number contained in the SND heartbeat, i.e., these values are reflected back in the ACK HEARTBEAT ordered-set. In the ACK HEARTBEAT, the OpCode byte is set to ACh.

Upon reception of a valid ACK (acknowledge) HEARTBEAT, the port finds the difference in time between its original SND transmission and the ACK reception. This difference correlates to the link round-trip latency, which correlates to link length.

For 8b/10b encoding, both SND HEARTBEAT and ACK HEARTBEAT ordered-sets are transmitted on every active lane simultaneously (like SKP ordered-sets).

For 64b/66b encoding, both SND HEARTBEAT and ACK HEARTBEAT ordered-sets are transmitted on 2 consecutive blocks.

SND HEARTBEAT ordered-sets are scheduled once every 100 ms in LinkUp State, to match the minimum DelayTimeOut period. Transmitting HEARTBEAT ordered-sets or blocks incurs negligible bandwidth loss and provides prompt notification of changes in link integrity.

Transmission of Link Heartbeat ordered-sets interspersed with other packets and ordered-sets uses the priority scheme described in [Section 5.8.6.9, “TxCMD = Enable,” on page 229](#).

5.14.2 HEARTBEAT ERROR HANDLING

Heartbeat errors are detected under the conditions described in [Section 5.9.5, “Heartbeat Error,” on page 238](#).

Heartbeat errors indicate that a link has been disconnected, and should result in the Link Training State Machine returning to **LinkDownDefaultState** (Polling or Sleeping) with the **LinkDownedCounter** being incremented.

5.14.3 HEARTBEAT LATENCY CALCULATION AND REPORTING

Round-trip latency is measured as the time between the end of Heartbeat SND transmission and the end of reception of the corresponding Heartbeat ACK. The longest possible round-trip latency, over a 10 km single-mode optical link, with roughly 5 ns/meter propagation speed, will be roughly 100 microseconds, and the shortest possible round-trip latency will be less than 100 nanoseconds.

Implementation Note:

Since the fundamental granularity is limited by clock cycle and HEARTBEAT ordered-set lengths to at least 4 ns, and the maximum round-trip link latency is roughly 100,000 ns across a 10 km link, a 2-byte counter (e.g., counting to 64K cycles with 4 ns granularity) should be sufficient. A 3-byte counter, with 4 ns. granularity, would allow accurate round-trip latency measurement of a link roughly 6,700 km long.

Use of the Heartbeat mechanism allows for a PortInfo value, *SM.PortInfo(LinkRoundTripLatency)*, which is accessible through the subnet management MAD Get mechanism, as described in [Section 5.6.2, “Status Outputs \(MAD get\),” on page 169](#). This is a 32 bit value representing the round-trip latency of the link, measured in 4 nanosecond intervals. This value is reset to FFFF_FFFFh upon entry to **LinkDownDefaultState**. When a SND heartbeat is transmitted, the current time (most likely via a free-running counter) is latched. Upon reception of an ACK heartbeat, the current time is compared against the latched time, and the difference, converted to nanoseconds, is reported as the LinkRoundTripLatency.

Due to delays in scheduling, an ACK may be sent significantly later (by as much as roughly 8 microsec.) than the time that the SND was received, resulting in an erroneously high link latency. Therefore, only the lowest value seen across a set of link latency measurements will indicate the most representative value, and only the lowest measured value should be reported as *SM.PortInfo(LinkRoundTripLatency)*.

The precision of reporting is dependent on how well the link round trip latency has been measured, but is expected to be measured to precision of better than 100 nanoseconds. When a port is brought to the Sleeping, Polling, or Disabled State, the **LinkRoundTripLatency** is re-set to FFFF_FFFFh.

5.15 MAX PACKET RATE

The maximum packet rate defines the minimal number of symbol times allowed between start of data packet delimiters (SDP). This requirement should allow a port to have an upper bound to the number of data packets it will need to process per second.

At FDR or higher data rates a transmitter shall not send a SDP within less than 64 Bytes (8 blocks) from the beginning of the previously transmitted SDP block.

Transmitter must add idle data (idle blocks) to the Packet Byte Stream to enforce this rule.

Note: Skip Blocks shall not be accumulated to achieve the 64 Byte between SDPs. The insertion of skip blocks may increase the time between SDPs. Removal of skip blocks shall not create a violation to the Max Packet Rate rule.

A link may negotiate to disable this feature and increase the max packet rate by setting the MPR bit to 1 in the TS3 - see [Section 5.5.2.4, “Training Sequence Three Ordered-Set](#)

[\(TS3\).” on page 143.](#)

A receiver may drop any packet if the start of packet delimiter (SDP) block is received less than 64 symbol times from the previous start of packet delimiter (SDP) block while Max Packet Rate is enabled (MPR_en is true).

o5-6: All ports claiming compliance with InfiniBand Rel. 1.3 shall implement the Max Packet Rate capability as defined in [Section 5.15, “Max Packet Rate,” on page 255](#), for FDR and higher data rates.

5.16 FDR AND HIGHER RATE TRANSMITTER EQUALIZATION

An transmitter capable of FDR or higher data rates shall implement a coefficient-based equalization mode in order to support fine-grained control over Tx equalization resolution. Additionally, a transmitter shall support a specified number of presets that give a coarser control over Tx equalization resolution. Both coefficient space and preset space are controllable via the equalization process in Config.CfgEnhanced, as described in [Section 5.8.4.6.5, “Config.CfgEnhanced State - Release 1.3.1 Enhanced Signaling,” on page 211](#). The coefficient space is also controllable via the fine tuning algorithm described in [Section 5.13, “Fine Tuning,” on page 246](#).

All transmitters must implement support for the equalization procedure, whereas receivers may optionally implement it.

Implementation of transmitter equalization is described in [Section 6.6.5, “Host Driver Output Characteristics for FDR,” on page 293](#) and [Section 6.6.6, “Host Driver Output Characteristics for EDR,” on page 302](#).

o5-7: All ports claiming compliance with InfiniBand Rel. 1.3 shall implement the FDR and Higher Rate Transmitter Equalization capability as defined in [Section 5.16, “FDR and Higher Rate Transmitter Equalization,” on page 256](#), for FDR and higher data rates.

5.17 PHYSICAL LAYER COMPLIANCE TESTING

The following facilities and procedures are intended to simplify physical layer device characterization and compliance testing to specifications described in [Chapter 6: High Speed Electrical Interfaces](#), while minimizing the amount of circuitry and complexity required in the InfiniBand devices.

C5-20.2.1: All ports claiming compliance with InfiniBand Rel. 1.2 Enhanced Signaling shall implement physical layer compliance testing as defined by [Section 5.17, “Physical Layer Compliance Testing,” on page 256](#), at all link speeds.

5.17.1 COMPLIANCE TESTING OVERVIEW

Device testing is accomplished by attaching the device to test equipment over a link, with the receiver attached to a signal generator such as an arbitrary pattern generator, and the transmitter attached to a signal monitor such as an oscilloscope. Testing at the physical layer does not require intervention by a Subnet Manager. New facilities are included

in the port to simplify compliance testing for the high-speed devices, including modification to the Link Training State Machine from the structure shown in [Figure 55](#) to the structure shown in [Figure 56](#), and definition of an ordered-set, Training Sequence for Test (TS-T), which is used exclusively for compliance testing. Multiple testing modes are possible, as determined by an Opcode in the TS-T ordered-set.

When the port is placed into the Phy Test state by reception of a TS-T ordered-set from attached test equipment on any lane, it disables all link-level protocol other than 8b/10b code/decode and TS1 detection, and only enables the physical layer circuitry, including clock and data recovery, link transmitter, driver de-emphasis, PRBS generator and checker and receiver equalization circuitry.

Transmitter testing is accomplished by using the TS-T or other equivalent means to command the transmitter to transmit particular test patterns, and measuring the characteristics (jitter, amplitude, rise/fall times, etc.) of the transmitted patterns.

For receiver testing, the test equipment uses the TS-T or other equivalent means to command the transmitter to send particular symbols depending on the validity (signal strength, 8b/10b decode correctness as to bit errors, disparity, or PRBS checks error etc.) of the received data, and then feeding the receiver input signals with various characteristics (signal strength, deterministic jitter, data dependent jitter, etc.) and monitoring the symbols the transmitter sends. The receiver testing uses specific test patterns (PRBS31 or PRBS9) rather than 8b/10b or 64b/66b encoded data.

Since there is no link-level logic enabled during this procedure (aside from optional vendor-dependent capabilities which are beyond the scope of this specification), no link-level functions (flow control, CRC, packet framing, etc.) are tested with these procedures. These facilities are intended to test compliance with transmitter and receiver specifications described in Chapter 6: High Speed Electrical Interfaces .

5.17.2 COMPLIANCE TESTING FACILITIES

Facilities for physical layer compliance testing include changes to the Link Training State Machine (LTSM), addition of a new Training Sequence for Test (TS-T) ordered-set, and capability for the link to generate patterns specifically used for compliance testing.

5.17.2.1 LINK TRAINING STATE MACHINE MODIFICATIONS FOR COMPLIANCE TESTING

The differences from the legacy Link Training State Machine include the following.

- 1) A new state is defined, "Phy Test", which allows physical layer compliance testing.
- 2) A new training sequence, Training Sequence-Test (TS-T) is added, which allows transitions from the LinkDownDefault State (either Polling or Sleeping) to the Phy Test state.
- 3) Two new transition arcs are added, from Polling and from Sleeping, to the Phy Test state, on reception of TS-T.
- 4) Two new transition arcs are added, from the Phy Test state to the LinkDownDefault state (either Polling or Sleeping), on RcvdTS1. These transitions allow exit from

Testing mode, without a power-off of the device. Testing procedures must ensure that no test pattern generates 8 contiguous TS1 patterns before testing is completed.

- 5) Within the Phy Test state, there are eight (8) different required modes of operation, as described in [Section 5.17.2.2, “Use of the Training Sequence for Test Ordered-Set \(TS-T\),” on page 258](#), including
 - (a) a “SKIP-less Idles” mode, of transmitting Idle data without SKIP ordered sets for clock compensation,
 - (b) a “SKIP-less TS1” mode, of transmitting back-to-back TS1 without SKIPS,
 - (c) a “Receiver Test” mode, of transmitting simple patterns indicating whether the receiver is detecting good data, or is detecting logical bit errors, or is detecting a loss of signal,
 - (d) a transmitter testing mode using high frequency pattern - 01010101,
 - (e) a transmitter testing mode using PRBS31,
 - (f) a transmitter testing mode using PRBS11,
 - (g) a transmitter testing mode using PRBS9, and
 - (h) a “Back to back TS1” mode, of transmitting back-to-back TS1s with SKIPS

Note: test modes (d)-(g) above are applicable only for devices that support data rates higher than QDR, test mode (h) is only applicable for devices supporting Rel 1.3.

- 6) The Phy Test state may optionally also include other test modes, and the TS-T ordered-set allows the transmitter and receiver to optionally be placed in multiple configurations, for optional compliance testing beyond the required modes described above.

The general operation of physical layer compliance testing is described below.

5.17.2.2 USE OF THE TRAINING SEQUENCE FOR TEST ORDERED-SET (TS-T)

The Training Sequence for Test (TS-T) is only generated by test equipment. There is no need for an InfiniBand device to be able to generate this ordered-set - only the need to recognize it and behave appropriately when one complete and error free TS-T ordered-set arrives at the receiver port on one or more lanes.

A TS-T may be received on only a single lane, or on multiple lanes. If a multi-lane TS-T ordered-set with different values on different lanes in symbols 8, 9 and 12-15, the port may use any of the received TS-T values.

Symbol 1 of a TS-T (the Lane ID in other ordered-sets) is reserved, so that the test equipment may insert a TS-T on any lane. Symbols 2-7 contain the TS-T unique data symbol, D17.2 (or 51h), whose 10-bit encoded value is the pattern (100011 0101) for both the positive and negative running disparity. Symbols 2-7 may be used by the port to determine the polarity of symbols 8-15, since the test equipment may insert a TS-T of either polarity.

Symbol 8 of a TS-T, the link speed identifier, is used to identify the speed at which the test equipment requests the IB device to operate. At least one bit of this bit map must be asserted to 1. If multiple bits are asserted, the test will be conducted at the highest **Link-SpeedEnabled** and **LinkSpeedExtEnabled** speed. This allows testing at the highest enabled speed by simply asserting all bits in this symbol.

Symbol 9 of a TS-T allows the test equipment to determine which testing mode the port will be placed in. The following values are defined.

0: SKIP-less Idle Data

For QDR or lower bit rates - each transmitter lane transmits a pseudo-random sequence of data symbols, generated by the 11th order LFSR = $X^{11} + X^9 + 1$ with no insertion of SKIP ordered-sets.

For FDR and EDR bit rates - each transmitter lane transmits a stream of scrambled Idle blocks, with no insertion of SKIP ordered-sets.

1: SKIP-less back-to-back TS1s

For QDR and lower bit rates each transmitter lane transmits an unbroken string of TS1 ordered-sets, with no insertion of SKIP ordered-sets.

For FDR and EDR bit rates each transmitter lane transmits an unbroken string of TS1 ordered-set blocks, with no insertion of SKIP ordered-sets.

2: Receiver test.

For QDR and lower bit rates:

On each lane, the transmitter sends an indication of the validity of the data received on the corresponding receiver lane.

VALID DATA: For each received symbol on a lane which decodes to a valid 8b/10b code point with good running disparity, the corresponding transmitter lane transmits a D10.2 (010101 0101) character.

LOGICAL ERROR: For each received symbol on a lane which decodes with a logical error (e.g., bit error, or running disparity error), the corresponding transmitter lane transmits a K28.5 D00.0 pair of symbols.

LOSS OF SIGNAL: For each received symbol on a lane which indicates an inadequate signal (e.g., inadequate signal swing, all 0s, all 1s, or noise), the corresponding transmitter lane transmits a K28.5 D01.0 pair of symbols.

For FDR and EDR bit rates:

On each lane the transmitter sends a PRBS31 corresponding to the validity of the received PRBS31 on the corresponding receiver lane. When a PRBS31 error was detected on the receiver lane, the transmitter produces a PRBS31 error on its transmitted stream.

Implementation Note:

A device may chose to implement the above receiver test by looping back the received PRBS31 stream to the transmitted stream hence propagating the received PRBS31 error to the transmitted stream

3: High frequency pattern.

Each transmitter lane transmits a high frequency pattern at the highest possible bit transition rate - 01010101 repeatedly.

4: PRBS31

Each transmitter lane transmits the PRBS31 pseudo-random sequence.

5: PRBS11

Each transmitter lane transmits the PRBS11 pseudo-random sequence.

6: PRBS9

Each transmitter lane transmits the PRBS9 pseudo-random sequence.

7: Back-to-Back TS1s

All lanes shall transmit back-to-back TS1s with Skip ordered sets. The Skip ordered set insertion shall follow the rules defined in [Section 5.8.1, “Link De-skew Training Sequence and SKIP ordered sets.” on page 188](#). This mode may be used to measure the lane-to-lane skew.

8-255: Optional Vendor-dependent opcodes to allow other testing modes.

A device in Phy Test state working in any mode 4-6 shall transmit the PRBS pattern in a way such that the pattern transmitted on each lane is not correlated with the pattern transmitted on any other lane, as described in [Section 5.5.2.6 on page 155](#). The method to achieve the above requirement is implementation specific but can be done by using a different PRBS seed or delay for each lane. Note that, if a delay is used, the amount of delay must be at least 132 UI to compensate for the maximum allowed lane to lane skew.

For QDR or lower data rates:

Symbols 12 and 13 allow the test equipment to optionally configure the transmitter in one of 65,536 vendor-dependent states, and symbols 14 and 15 allow the test equipment to configure the receiver in one of 65,536 vendor-dependent states. The values 0000h are identified for “normal” or “default” operation, so that test equipment can be expected to get good and correct operation when these fields are set to 0. Values other than 0 are optional and vendor-dependent. Typically only a very small subset of these states will be valid and will provide specific unique behaviors.

For FDR or higher data rates:

Symbols 12 and 13 allow the test equipment to optionally configure the transmitter in one of the 16 Transmitter settings as defined in [Table 54 Tx FIR Filter Coefficients](#)

and Amplitudes for 14.0625 Gb/s (FDR) on page 299, using the 4 least significant bits of Byte 12. Byte 12, bit number 4 in byte 12 shall represent the requested amplitude bit. Preset 0 shall represent the transmitter default preset used for active limiting copper or optical cables, and must meet the host transmitter requirements as defined in [Table 55 FDR host output specifications at Preset 0, for Limiting Active Cables on page 301](#). Other bits in Byte 12 and Byte 13 are allocated for other vendor-defined transmitter equalization.

Symbols 14 and 15 allow the test equipment to configure the receiver in one of 65,536 vendor-dependent states. The values 0000h are identified for “normal” or “default” operation, so that test equipment can be expected to get good and correct operation when these fields are set to 0. Values other than 0 are optional and vendor-dependent. Typically only a very small subset of these states will be valid and will provide specific unique behaviors.

The PRBS31 pattern generator shall produce the same result as the implementation shown in [Figure 71](#). This implements the inverted version of the bit stream produced by the polynomial in [Figure 71](#). The PRBS31 pattern checker shall produce the same result as the implementation shown in [Figure 72](#).

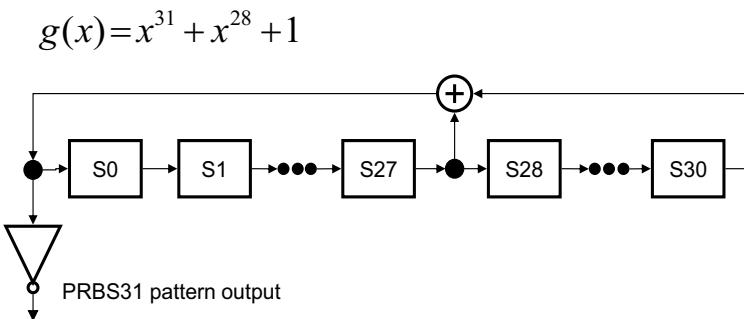


Figure 71 PRBS31 Generator

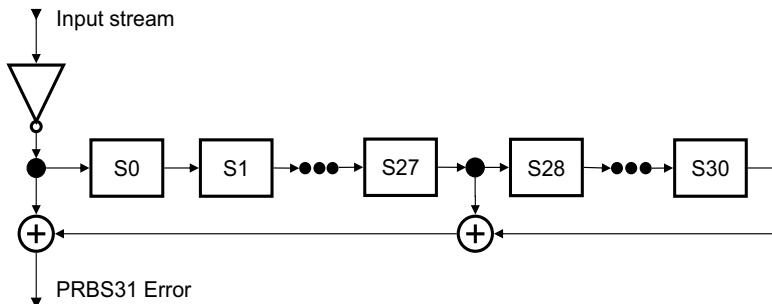


Figure 72 PRBS31 checker

The PRBS11 pattern generator shall produce the same result as the implementation shown in [Figure 64 on page 218](#). This implements the inverted version of the bit stream produced by the polynomial in [Figure 64](#).

The PRBS9 pattern generator shall produce the same result as the implementation shown in [Figure 73 on page 262](#). This implements the inverted version of the bit stream produced by the polynomial in [Figure 73](#).

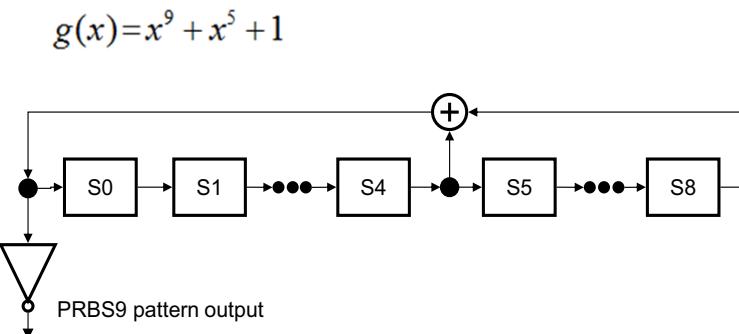


Figure 73 PRBS9 Generator

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42**Implementation Note:**

One particular use of this transmitter and receiver configuration facility may be for measuring transmitter equalization. As described above in [Section 5.5.2.4](#) and in [Section 6.6.5.3, “Transmit Equalization Presets.” on page 298](#), this facility allows transmitters and receivers to negotiate for use of up to 16 different transmitter equalization setting (TES) configurations (Rel. 1.2.1 - QDR, DDR, or SDR) and 16 Presets with 2 amplitude settings (Rel. 1.3/1.3.1 - FDR & EDR - a total of 32 independent transmitter settings) which may be adapted to the characteristics of the link. Testing of this transmitter equalization facility over various physical links will be an important function. This facility in the TS-T ordered-set can allow the test equipment to configure the transmitter to use a particular TES configuration as it is transmitting the data, in order to see how the different TES settings operate over various media.

Similarly, the facility allows the test equipment to test out various configurations of the receiver circuitry, such as using different receiver equalization setting (RES) configurations.

5.17.3 EXAMPLE COMPLIANCE TEST PROCEDURE

Efficient and fast compliance testing requires a simple and efficient procedure to measure transmitter and receiver characteristics. The proposed procedure described below allows simple testing of the most critical elements of the transmitter and receiver circuitry. More detailed procedures will be used in practice, as this procedure is intended to illustrate use of the facilities described above.

- 1) The Device Under Test (DUT) is powered up, and goes into Polling state, transmitting repeated TS1(s) for 2 ms in Polling.Active followed by 100 ms of a quiescent output in Polling.Quiet.
- 2) The Test Equipment (TE) (which may be a combination of a sophisticated arbitrary pattern generator and oscilloscope, or a simpler and more application-specific box with FPGAs, simple microprocessors and SERDES circuits) is attached to the DUT by a cable.
- 3) The TE transmits a TS-T to the DUT, with flags set to test a particular device capability (e.g., TS-T with fields set to Speed=QDR, Opcode=0, TxCfg=0000h and RxCfg=0000h).
- 4) The DUT detects the TS-T, transitions to the Phy Test state, and uses the values in the TS-T fields to determine the specific test mode.
 - If the Opcode in the TS-T was set to 0, indicating “SKIP-less Idle Data,” the DUT transmits Idle data. The TE measures parameters such as Rise/Fall Time, Driver Jitter, Transmitter Peak-Peak voltage, etc., using the wide-spectrum of data contained in the Idle data pattern.
 - If the Opcode in the TS-T was set to 1, indicating “SKIP-less back-to-back TS1s”, the DUT transmits TS1 patterns, allowing the TE to make similar measurements on a much shorter data pattern with a combination of high-frequency (1010...) and low-transition-density (K28.5) elements.

- If the Opcode in the TS-T was set to 2, indicating “Receiver Test”, the DUT transmits either D10.2, K28.5 D0.0, or K28.5 D01.0 patterns, depending on the validity of the received signal. The TE can test the sensitivity of the receiver by varying the signal given to the receiver and monitoring the transmitted pattern. Similarly, the TE can test the receiver’s sensitivity to input jitter the same way, to do a full four-corners test of receiver sensitivity to signal strength and received jitter.
- 5) The TE can move the device out of Phy Test state and back to the **LinkDownDefaultState** (Polling or Sleeping) at any time by sending a stream of 8 contiguous TS1 ordered-sets on a receiver lane.
 - 6) Once the device is back in the Polling state, the TE can test the transmitter under different conditions by sending in a new TS-T ordered-set (e.g., with speed set to DDR instead of QDR, or with TxCfg set to 1 instead of 0).
 - 7) At finish of testing, the TE would send the DUT a series of 8 TS1 ordered-sets, placing it back in Polling mode, and then the cable can be removed. The device is then ready to be connected to other IB devices.

During the Phy Test state, all lanes shall be operating at the same speed. All lanes of a port shall move in and out of the Phy Test state together, and all lanes shall be transmitting the same test pattern. This allows the test procedure described here to be implemented either with serial test equipment, or with more complex parallel test equipment running multiple lanes at once.

As described above, the format of the TS-T ordered-set and the function of the Phy Test state allows various other capabilities to be defined and used in a vendor-dependent manner, limited by the device and the testing equipment capabilities.

CHAPTER 6: HIGH SPEED ELECTRICAL INTERFACES

6.1 INTRODUCTION

This chapter describes the high speed interfaces for use with InfiniBand™ links. The signaling rates are for encoded data on the media, and correspond to the encoding for that signaling rate as specified below. The supported data rates are listed in [Table 45](#)¹.

Table 45 InfiniBand Link Data Rates

InfiniBand rate designator	Per-lane signaling rate, GBd	Unit Interval (UI) or bit period, ps	Codec	Aggregate full duplex throughput, GB/s (GBytes/sec)			
				Link Designator		4X interface	
SDR	2.5	400	8b/10b	(1+1) GB/s	10G-IB-SDR	(3+3) GB/s	30G-IB-SDR
DDR	5.0	200	8b/10b	(2+2) GB/s	20G-IB-DDR	(6+6) GB/s	60G-IB-DDR
QDR	10.0	100	8b/10b	(4+4) GB/s	40G-IB-QDR	(12+12) GB/s	120G-IB-QDR
FDR	14.0625	71. <u>11</u>	64b/66b	(6.8+6.8) GB/s	56G-IB-FDR	(20.4+20.4) GB/s	168G-IB-FDR
EDR	25.78125	38. <u>78</u>	64b/66b	(12.5+12.5) GB/s	104G-IB-EDR	(37.5+37.5) GB/s	312G-IB-EDR

This release of the specification extends and enhances and in some cases supersedes prior releases. Products built using this release shall interoperate with legacy products whose maximum data rate is two speed generations slower than their maximum data rate. For example, FDR devices are required to interoperate with DDR, QDR, and FDR devices. DDR, QDR, FDR, and EDR devices shall be capable of operating at the SDR data rate to allow for speed and link training negotiation, and may optionally operate at the intermediate data rates. Details of the speed negotiation and configuration are specified in [Section 5.8, “Link Initialization and Training,” on page 187](#).

6.1.1 BACKGROUND AND REFERENCE MATERIAL

While this specification endeavors to be complete, there is a great deal of background material which is helpful in understanding and correctly implementing this interface that is not included.

Much useful information can be found in books such as “High-Speed Digital Design, A Handbook of Black Magic” by Howard Johnson and Martin Graham and “Digital Systems Engineering” by William J. Dally and John W. Poulton. The web site <http://www.t11.org> has a great deal of useful information as does <http://ieee802.org/3>.

1. Aggregate full duplex throughput rates include all link-level overhead (e.g., packet headers and delimiters, link packets, SKIP ordered sets, etc.) but don’t include 8b/10b or 64b/66b encoding overhead. Note that all of the per-lane signaling rates are integral multiples of 156.25 Mb/s.

The Common Electrical I/O (CEI) Interoperability agreement for 6+ and 11+ Gbps I/O and CEI-25G-LR proposal from the Optical Internetworking Forum (OIF) (see reference [42]) were consulted in the preparation of this chapter.

Additional information is available from the IEEE (I&M) Subcommittee on Pulse Measurement Techniques (SCOPT). Methods of performing pulse amplitude and parametric measurements for SDR signaling should be conducted in accordance to IEEE Std.181-2011, where referenced in this chapter.

Note that the differential amplitude represents the value of the voltage between the true and complement signals. This may be expressed as RMS, peak, or peak-peak. The Peak-Peak value is twice the Peak value. This document uses the IEEE Std.181-2003 terminology in which the value is commonly referred to as “peak-peak” is rather called “unsigned amplitude”.

6.2 ELECTRICAL TOPOLOGIES

This chapter of the specification focuses on defining the electrical requirements for the interconnect between two Link End Nodes. An example application is a switch connected through a passive copper cable to an Host Channel Adapter (HCA) housed in a server. A Link End Node is any device that fully implements an InfiniBand interface, as defined in this InfiniBand specification. Some examples of Link End Node implementations are:

- Host Channel Adapter (HCA)
- Switch
- Router

This section outlines the channel topologies accommodated within the InfiniBand specification. While InfiniBand is a chip-to-chip standard in the broadest sense, this document focuses on topologies where interoperability between different vendors' devices is required. As a result, on-board and backplane topologies have been deprecated. Information on those obsolete topologies can be found in Rel. 1.2.1 of the Vol. 2 InfiniBand specification.

A physical InfiniBand link typically consists of two host boards interconnected with cabling, as illustrated in the high-level block diagram in [Figure 74 on page 267](#). The cabling is connected to the host board using a separable electrical connector. The connector technology used is dependent on the maximum supported data rate of the host, as well as the number of electrical lanes implemented for the link. Supported InfiniBand link widths are 1X, 4X, 8X, and 12X. Refer to [Chapter 7: Electrical Connectors for Modules and Cables](#) for supported InfiniBand cable connector types.

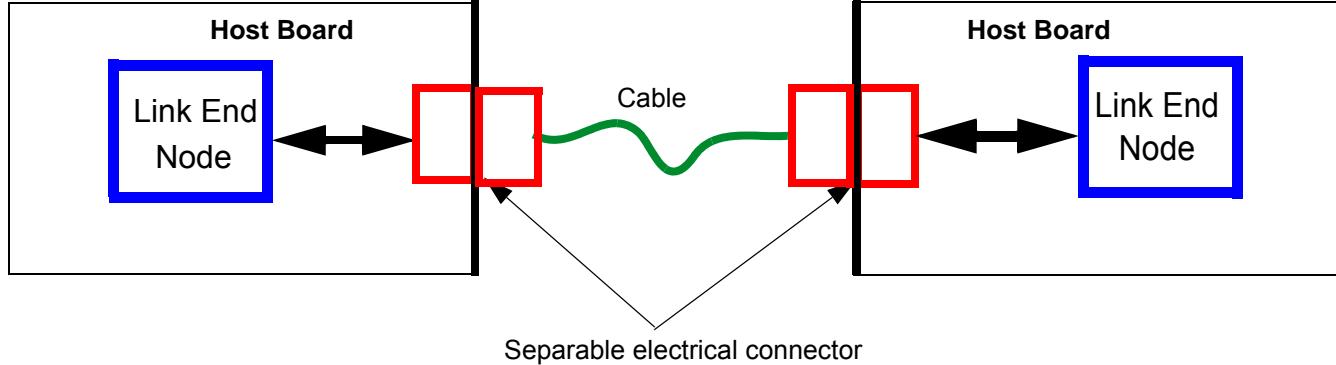


Figure 74 High-level topology block diagram

The electrical implementation details on the host board are vendor-specific and not defined within this document. All electrical host compliance measurements will be performed at the output of the host board (or input, depending on signal direction) by plugging a host compliance board into the separable electrical connector of the host board. Since electrical host board compliance is evaluated as a 'black box' measurement, host board implementers are free to design the host board using whatever means are necessary to guarantee electrical compliance at the host board output/input. For example, one unique implementation could incorporate an HCA connecting to a backplane connecting to a pass-through card with a repeater connected to the cable connector. The cable port on the pass-through module would be the compliance interface, and the input to and output from the cable port would have to be compliant with the receiver and transmitter compliance specifications, respectively.

Cabled interconnects may be implemented as a cable assembly in which the separable electrical connector is permanently connected to the cable transmission medium. Alternatively, a separable transceiver may be implemented, where a passive (typically optical) cable is attached to a pluggable transceiver through a second separable connector interface, as shown in [Figure 75 on page 268](#). Separable fiber optic cables utilize a transceiver module that plugs into a mating connector on the host board and has a connector such as an MPO for a fiber optic cable. The fiber optic cable can therefore be unplugged from the transceiver while the module stays in place, allowing the user to use different lengths or types of fiber medium. The transceiver module is always an active device.

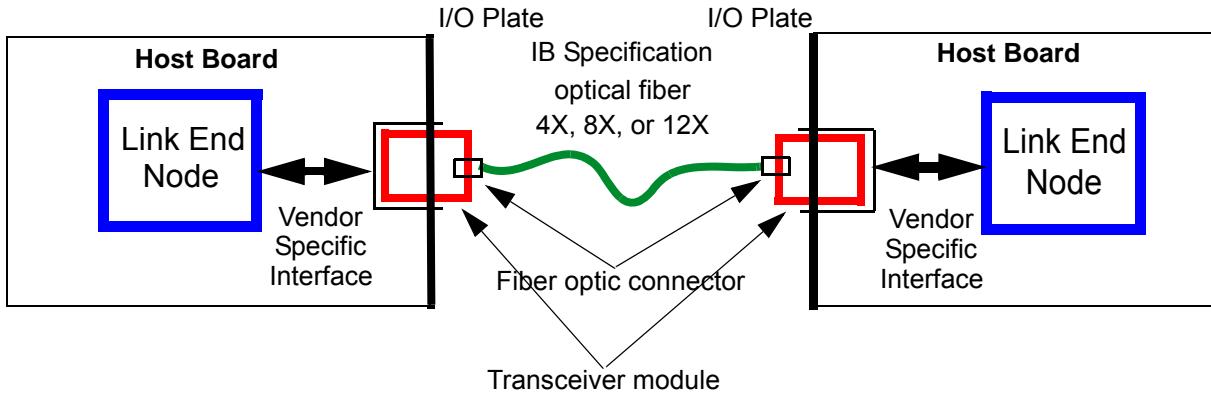


Figure 75 Optical Fiber Interconnect Topology

Like the host board, the cable electrical compliance is defined and evaluated at separable connector interfaces. Electrical implementation details between compliance points are left up to the implementer.

6.2.1 REPEATERS

The InfiniBand Architecture allows for the use of “signal repeaters” to recover from potentially weakened signal strength and, in some cases, built-up jitter between two end nodes of a link. The following definitions will be used in this specification when referring to various types of repeaters. This list is provided to define terminology used throughout the specification and is not necessarily exhaustive or exclusive.

Repeaters may be implemented at various locations throughout the electrical channel between two Link End Nodes. Repeaters are expected to produce compliant outputs at indicated compliance points. Specifically,

- Repeaters implemented on the host in the transmit channel between the Link End Node and the pluggable cable interface are expected to produce a signal at the host output that is compliant to [Section 6.6 on page 282](#).
- Repeaters implemented inside cable assemblies are expected to receive and transmit signals that are compliant to [Section 6.8 on page 323](#).
- Repeaters implemented on the host in the receive channel between the pluggable cable interface and the Link End Node are expected to successfully receive input signals compliant with [Section 6.7 on page 309](#).
- A repeater which retimes by recovering the clock and filtering it is allowed. The transmit port of any repeater shall meet the relevant transmitter specifications for its intended signaling rate. The receive port of any repeater shall meet the specifications of the relevant receiver specifications for its intended signaling rate. Cables that contain repeaters that retime, like those that do not, are expected to receive and transmit signals that are compliant to [Section 6.8 on page 323](#).

- Repeaters are not required to be protocol aware nor capable of any error detection.
- Repeater outputs shall be capable of being made quiescent. This can be achieved through LOS detection and corresponding output squelch or by disabling the repeater transmitter using commands sent through its management interface. For example, repeaters on the transmit (input) end of an active cable shall implement disable through the I2C management interface, while repeaters on the receive (output) end of an active cable shall implement LOS detection and squelch. See [Section 7.5.3.4 on page 393](#) or [Section 7.8.3.5 on page 435](#) for details on requirements for capabilities by cable type.
- Repeaters shall be capable of handling SDR signaling in addition to any other signaling rate with which compliance is claimed. Repeaters capable of higher speeds ((i.e., DDR, QDR, FDR, EDR) may need to be protocol aware to a limited degree in order to participate in the speed negotiation and equalization or have capabilities to handle SDR data with higher data rate clocking. A repeater may use the specifications of an intended higher rate of operation during the Link Initialization and Training process, even if carrying signals at 2.5 GBd (SDR) (see, for example, [Table 84](#)).

6.2.1.1 LINEAR REPEATERS

The output voltage of a Linear Repeater must be linearly related to the input voltage across the frequency band associated with the maximum supported data rate. These devices do not retime the input signal and therefore do not reset the link jitter budget. Linear Repeaters appear to Link End Nodes as another linear element in the electrical channel and are only participants in the link training via proxy through an optional low-speed management interface. These devices do not reset the link equalization budget.

6.2.1.2 LIMITING NON-RETIMING REPEATERS

Limiting Non-Retiming Repeaters use a limiting amplifier to sense the input voltage and retransmit that voltage at their output. Due to the limiting amplifier, the relationship between the input and output voltage of the repeater may be non-linear. These devices do not incorporate clock recovery and so do not reset the link jitter budget. These repeaters are unable to directly participate in link training, but can participate by proxy through an optional low-speed management interface to a Link End Node. Since these devices are non-linear in nature, they must be considered when determining the link equalization budget.

6.2.1.3 TRANSPARENT RETIMING REPEATERS

Transparent Retiming Repeaters utilize a limiting amplifier to sense their input voltage and also recover the clock from the input waveform. Due to the limiting amplifier, the relationship between the input and output voltage of the repeater may be non-linear. Since these devices recover the clock from the input waveform, they do reset the link jitter budget. However, because they do not possess an external oscillator, these devices do not perform SKIP insertion and are therefore transparent to Link End Nodes at the protocol level. These repeaters are unable to directly participate in link training, but can participate by proxy through an optional low-speed management interface to a Link End Node. Since these devices are non-linear in nature, they must be considered when determining the link equalization budget.

6.2.1.4 SKIP ORDERED-SET DEPENDENT RETIMING REPEATERS

SKIP-Ordered Set Retiming Repeaters use a local oscillator to clock incoming data and reset the link jitter budget. Since these devices may experience clock skew relative to Link End Nodes, they are required be able to insert or delete SKIP ordered sets for clock skew compensation. [Figure 5.12](#) defines the SKIP insertion and deletion behavior of these devices in detail. These repeaters are unable to directly participate in link training, but can participate by proxy through an optional low-speed management interface to a Link End Node. Since these devices are non-linear in nature, they must be considered when determining the link equalization budget.

6.2.2 CABLE TYPES

[Figure 74 on page 267](#) shows a topology diagram illustrating the use of cabling in an InfiniBand link. These interfaces are used for driving between TCAs or HCAs and switches through cables from one I/O Plate to another. A typical application of this topology might be a stand-alone storage box connecting to another such box or a processor. The following sections of this document describe various cable types including the placement of optional repeaters in the cable. Those cables sometimes referred to as “pluggable” cables may be of the active or passive type and utilize either copper or fiber interconnect media, with any active components being located inside the pluggable module. In these types of cables the copper wire or fiber transmission medium can not be separated from the plug or transceiver module. Fiber optic transceivers of this type are always active devices.

Passive copper cables, with or without passive equalization in cables or connectors, are typically used for links operating at lower speeds for shorter lengths. As data rates increase, the maximum achievable link length for passive cable decreases, driving the need for cables containing active components to condition and amplify the signal to achieve longer interconnect lengths. Release 1.2.1 and later versions of this document support both passive cables and active cables, with both copper and optical transmission technology, and with both limiting and linear amplification in the active cables.

In active cables the active components are located inside the cable plug shell, and only electrical interfaces at the input and output of the cable plug shell are specified. Active cables may utilize either optical or copper, or other material, as the transmission medium. The presence of active components impacts the link equalization requirements and must be considered during the link training process.

Various types of active cables are defined. Active optical cables are assumed to have limiting amplifiers at input and output of the cable plug shell. Active copper cables may be one of several types. The most commonly useful type is a full-active limiting cables, with limiting amplifiers at both input and output. Half-active cables have active components at either the input or output signal ends of the cable. Active linear cables incorporate linear amplifiers to emulate the performance of passive cables, but with longer link length capabilities for a given cable construction. The maximum distance, power dissipation, and jitter specifications will vary depending on the type of cable.

Active cables are required to provide a mechanism for quiescence of their output when no signal is being applied to their input. This quiescence can be implemented as LOS detection and corresponding output squelch, but in some cases (e.g. optical transmitters) this functionality is optional. In cases where LOS/squelch is not required, the host shall use the management interface to disable the repeater output when no signal is being applied to the repeater input. The only exception to this requirement is the Polling.Quiet state. Refer to [Section 7.5.3.4 on page 393](#) or [Section 7.8.3.5 on page 435](#) for details on squelch requirements for specific cable connector form factors.

Beginning with QSFP and CXP cables, EEPROMs were added to each cable plug. Prior to link training, the host reads the cable EEPROM to identify the appropriate electrical settings to use during link training. Since different cable types can require different approaches to link training, the cable type field is required and critical to successful link training. Chapter 8 contains definitions of the required and optional contents of cable EEPROM memory maps for InfiniBand applications.

For the purposes of this specification, it is helpful to categorize types of cables based on the electrical characteristics requirements at the host output, or cable input, and host input, or cable output. Each of the following definitions can, in many cases, represent several cable type definitions in the memory map of the cable EEPROM. These combinations are possible because the multiple cable types in the EEPROM have the same signal electrical characteristics required of them. New cable types not anticipated at the time of publication of this spec are possible, although typically they will fall under one of these existing cable type categories.

The link equalization budget is different for each of the cable types. Refer to [Section 6.5 on page 279](#) for an overview of the link equalization budget.

6.2.2.1 LINEAR CABLES

Linear cables consist of any cable where the output signal is a frequency-dependent linear multiple in amplitude and phase of the input signal across a range of frequencies. Examples of linear cables include passive copper cables, with or without equalization, and active cables with linear repeaters. Passive copper cables do not contain any active circuitry within the InfiniBand data path.

6.2.2.2 FULL LIMITING ACTIVE CABLES

Full limiting active cables (also referred to as "limiting active cables") consist of any cable that contains a limiting amplifying repeater at its input and a limiting amplifying repeater at its output. Some examples of full limiting active cables are active optical cables (AOCs), optical transceivers with separable optical cabling, and active copper cables that contain two limiting amplifying repeaters, one on each end of the cable. This type of cable is illustrated in [Figure 76](#).

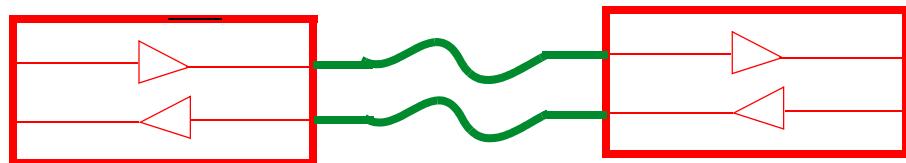


Figure 76 Full limiting active cable topology

6.2.2.3 NEAR-END LIMITING ACTIVE CABLES

Near end limiting active cables consist of any cable that contains a limiting amplifying repeater at its input, which in turn transmits the conditioned and amplified signal through a linear interconnect into the host receiver. Near-End Limiting Active cables are typically implemented with copper bulk cabling. This type of cable is illustrated in [Figure 77](#). These cables are also sometimes referred to as “near-end half active” cables.

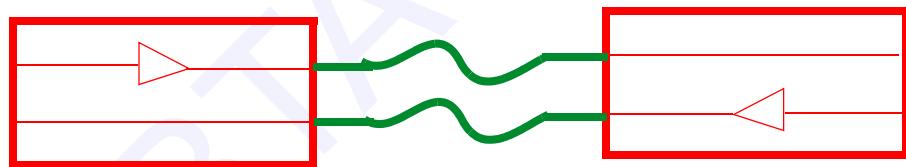


Figure 77 Near-end limiting active cable topology

6.2.2.4 FAR-END LIMITING ACTIVE CABLES

Far-End Limiting Active Cables are cables that contain a limiting amplifying repeater only at their outputs. In these cables, the cable interconnect up to the far end limiting amp device is expected to be linear. Far-End Limiting Active cables are typically implemented with copper bulk cabling. This type of cable is illustrated in [Figure 78 on page 272](#). These cables are also sometimes referred to as “far-end half active” cables.

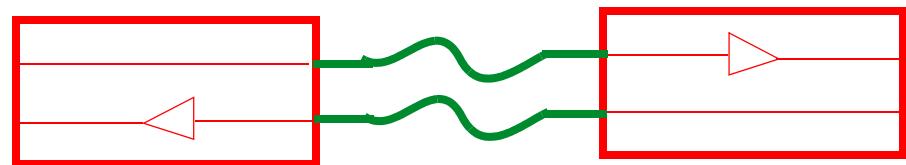


Figure 78 Far-end limiting active cable topology

6.3 GENERAL REQUIREMENTS

Connections are point to point and signaling is unidirectional. By definition, a physical lane comprises a differential pair or fiber in each direction, i.e. a transmit signal and a receive signal at each end.

The characteristic impedance of the cables and printed wiring is nominally 100 ohms differential. The single ended impedance is more variable, depending on the amount of coupling and the package design. The details of properly designing the packaging and interconnect of an InfiniBand link are beyond the scope of this specification and are the responsibility of the designer.

An important requirement of InfiniBand is that devices which may be produced by different manufacturers shall be interoperable and hot-pluggable.

6.3.1 ELECTROSTATIC DISCHARGE (ESD)

InfiniBand ports shall withstand 2 kV of electrostatic contact discharge to the host device receptacle housing using the Human Body Model per JEDEC Standard JESD22-A114:B2000 without damage or non-recoverable error including but not limited to latchup. A recoverable error is one that does not require reset or replacement of the device.

Cables shall meet ESD requirements given in EN 61000-4-2, criterion B test specification such that when installed in a properly grounded housing and chassis the unit is subjected to 15 kV air discharges during operation and 8 kV direct contact discharges to the case.

Further ESD specifications are given in [Section 7.3.4.1](#), [Section 7.5.4.1](#), [Section 7.7.4.1](#), or [Section 7.8.7.6](#), depending on connector type.

6.4 COMPLIANCE FACILITIES

6.4.1 COMPLIANCE POINTS

Since the ASIC or Serdes pins are not accessible, other points within the IB defined topologies are defined and the signal levels listed in [Table 46, “InfiniBand Signal Test Points,” on page 274](#) and depicted in [Figure 79](#) to [Figure 84](#) below.

Table 46 InfiniBand Signal Test Points

Test Point	Description	Test Point	Description
TP1	Transmitted signal at host board side of backplane connector	TP6a	Transmitted signal at HCB test connector
TP2	Transmitted signal at backplane side of backplane connector	TP7	Received signal at board side of cable connector
TP3	Received signal at host board side of backplane connector	TP7a	Received signal at MCB test connector
TP4	Received signal at backplane side of backplane connector	TP8	Received signal at cable side of cable connector
TP5	Transmitted signal at board side of cable connector	TP8a	Received signal at HCB test connector
TP5a	Transmitted signal at MCB test connector	TP9	Transmitted signal at input to pluggable module
TP6	Transmitted signal at cable side of cable connector	TP10	Received signal at output of pluggable module

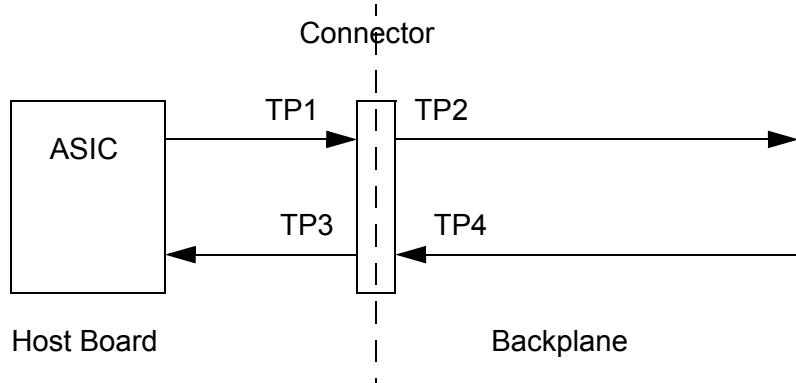


Figure 79 Board/Backplane test points

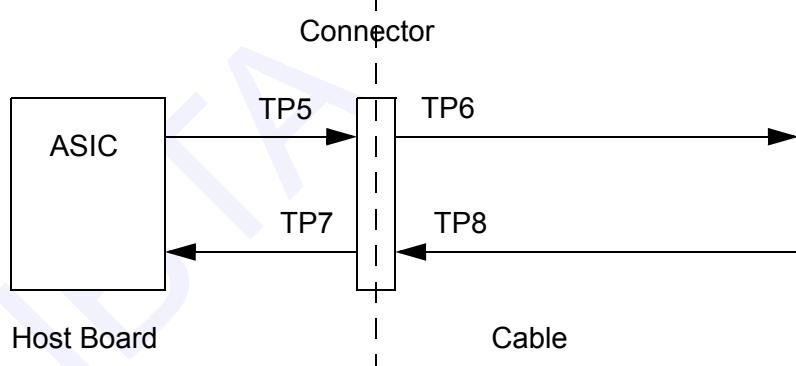


Figure 80 Board/Cable test points

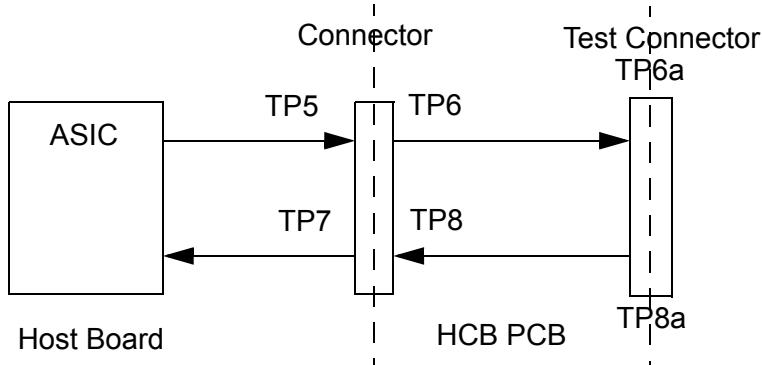


Figure 81 Host compliance board test points

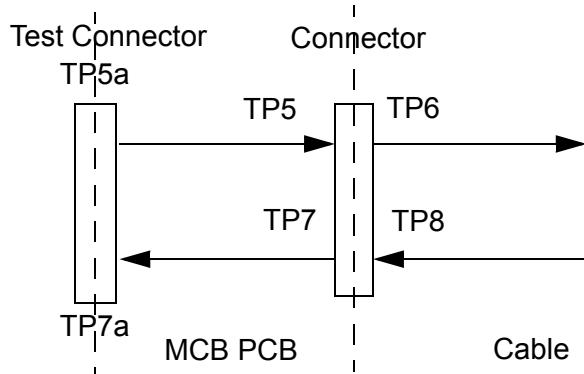


Figure 82 Module compliance board test points, cable

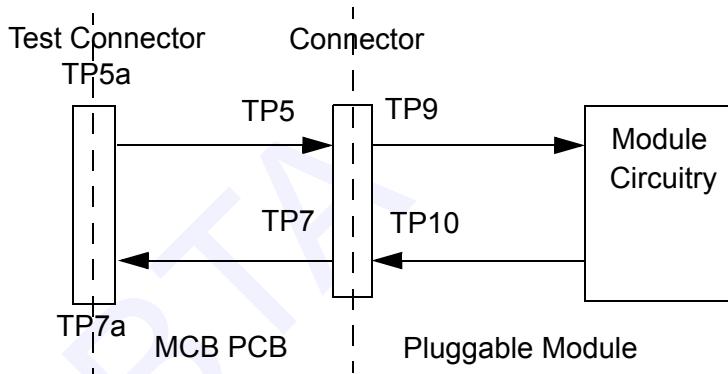


Figure 83 Module compliance board test points, pluggable module

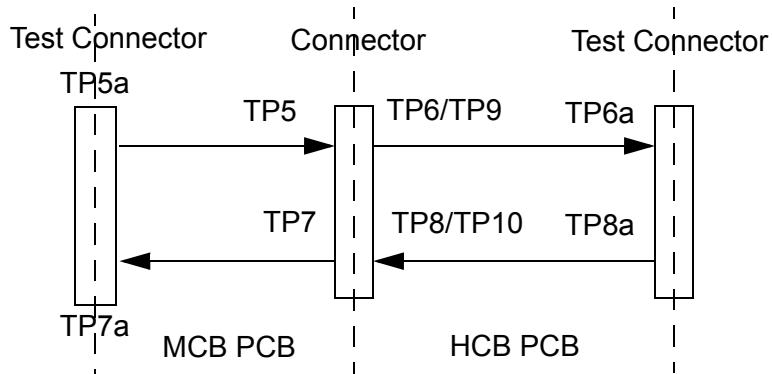


Figure 84 Mated Module and Host Compliance board test points

C6-0.2.1: SDR Backplane Connections shall comply with the amplitude and eye opening at TP1 and TP3. TP2 and TP4 are for reference.

C6-0.2.2: SDR Cable Connection shall comply with the amplitude and eye opening at TP5 and TP7. TP6 and TP8 are for reference.

C6-1: 1x Pluggable and 4x Pluggable Device Ports at SDR, DDR, and QDR speeds shall comply with the amplitude and eye opening at TP9 and TP10. Measurement may require de-embedding.

6.4.2 COMPLIANCE TESTING HARDWARE AND METHODOLOGY

This section provides only an introductory explanation of compliance testing hardware and methodology. Complete descriptions are provided in the Method of Implementation (MOI) documents for Active Time Domain Testing and for VNA Testing, generated by the IBTA Compliance and Interoperability Working Group (CIWG). Please refer to these documents for detailed specification of testing methodology, parameters, and recommended test equipment.

Testing of InfiniBand devices and cables requires the use of interface hardware between the device or cable connectors and the test equipment to be used. For QDR and lower data rates, QSFP to SMA or microGigaCN to SMA breakout cables are typically used for transmitter and receiver characterization. These breakout cables are not suitable at FDR and higher data rates. For FDR and higher data rates, transmitter output and receiver input characteristics are defined at the test connectors of a Host Compliance Board (HCB) as described in [Annex 1: FDR and EDR Compliance Boards and Test Setups on page 614](#).

Cable test boards are used in the measurement of cable electrical characteristics. For QDR and lower data rates, the electrical parameters of the test boards are left to the implementer. However, the effects of the compliance boards shall be de-embedded from the measured data.

For FDR and higher data rates a Module Compliance Board (MCB) and a Host Compliance Board (HCB), as described in [Annex 1: FDR and EDR Compliance Boards and Test Setups on page 614](#), shall be used.

Many of the electrical specification values at FDR and higher data rates are set based on the compliance board characteristics since they are measured using those boards without de-embedding. Therefore, use of compliance boards with different characteristics than those defined in the CIWG MOIs and in [Annex A1: FDR and EDR Compliance Boards and Test Setups](#) is discouraged.

At FDR and higher data rates, test equipment must be able to generate and receive several different test patterns, including at least PRBS31 and PRBS9. PRBS31 or Idle test patterns are used for characterization of parameters which don't require capturing a full test pattern (i.e., RJ, TJ, J2, J9, Qsq, Vcm), and parameters that don't require separating total jitter (TJ) into data dependent jitter (DJ) and random jitter (RJ). For parameters that

do require distinguishing DDJ from RJ (TJ-DDJ, DDPWS), the RJ determined from a shorter repeating pattern (PRBS9) is subtracted from the TJ in a PRBS31 pattern, assuming that RJ is unchanged between short and longer patterns. S-parameters and ILD, ICN, ICMCN are measured with a network analyser, although in some circumstances the item under test is operated using PRBS9 (or PRBS7, -11 -15 or -31, depending on available test equipment capabilities).

6.4.3 LINK/PHY COMPLIANCE PROVISIONS

Test, characterization, and compliance verification of an InfiniBand port requires additional states, modes, or accessible signals that are not required for functional operation. The test and characterization modes are described in [Section 5.17 on page 256](#) of this specification. These facilities are required for devices claiming compliance with Rel. 1.2 or later but are not required for legacy ports.

C6-1.2.1: Any device claiming InfiniBand compliance at the slot interface, or copper cable interface shall comply with the requirements of [Chapter 5: Link/Phy Interface](#) for Port Type 1 when operating at 2.5 Gb/s (SDR).

C6-1.2.2: Any 1x pluggable device claiming InfiniBand compliance at the socket interface shall comply with the requirements of [Chapter 5: Link/Phy Interface](#) for Port Type 2 when operating at 2.5 Gb/s (SDR).

o6-1.2.1: Any device claiming InfiniBand Rel. 1.2 Enhanced Signaling compliance **may** support data rates higher than SDR. A port may support a non-contiguous set of link speeds, e.g., SDR and QDR without DDR, SDR and FDR without DDR, etc.

C6-1.2.3: Any port claiming support for DDR signaling rate and InfiniBand Rel. 1.2 Enhanced Signaling compliance at the slot interface or copper cable interface shall comply with the requirements of [Chapter 5: Link/Phy Interface](#) while operating at 5.0 Gb/s (DDR).

C6-1.2.4: Any port claiming support for QDR signaling rate and InfiniBand Rel. 1.2 Enhanced Signaling compliance at the slot interface or copper cable interface shall comply with the requirements of [Chapter 5: Link/Phy Interface](#) while operating at 10 Gb/s (QDR).

C6-1.2.5: Any port claiming support for FDR signaling rate and InfiniBand Rel. 1.3 Enhanced Signaling compliance at the slot interface or copper cable interface shall comply with the requirements of [Chapter 5: Link/Phy Interface](#) while operating at 14.0625 Gb/s (FDR).

C6-1.2.6: Any port claiming support for EDR signaling rate and InfiniBand Rel. 1.3.1 Enhanced Signaling compliance at the slot interface or copper cable interface shall comply with the requirements of [Chapter 5: Link/Phy Interface](#) while operating at 25.78125 Gb/s (EDR).

6.5 EQUALIZATION METHODOLOGY

The frequency dependent attenuation of the interconnection media degrades the signal and thus produces Inter-Symbol Interference or Data Dependent Jitter. The effects of high frequency attenuation can be reduced by equalization, through techniques such as:

- Pre-distortion or pre-emphasis of the signal produced at the driver. (The term “de-emphasis” is also used.) This refers to a technique where the bits in which a logical transition occurs (0 to 1 or 1 to 0) have a larger amplitude than bits in which no transition has occurred. [Figure 85](#) shows an example waveform using this technique.

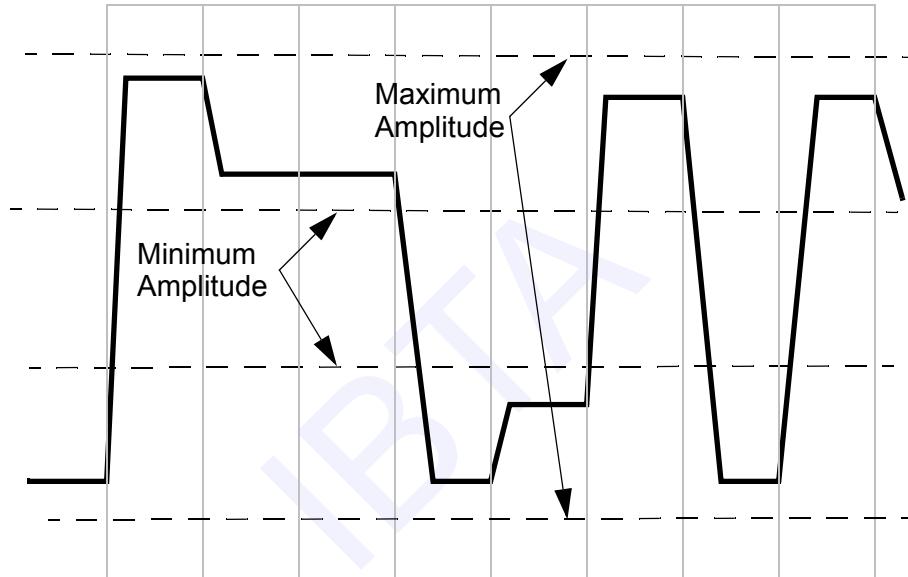


Figure 85 Waveform with De-emphasis

- Addition of a passive high pass filter network which has a frequency response complementary to that of the interconnect between driver and receiver. This is sometimes called Passive Equalization.
- Addition of a linear equalizing redriver in the channel
- Adaptive equalization using techniques such as partial response or DFE may be implemented at the receiver.

The method of equalization selected is dependent on the frequency of operation and the generation (version) of the ports.

Receivers shall incorporate equalization since the allowed de-emphasis will not be sufficient to guarantee an open eye under all circumstances.

6.5.1 EQUALIZATION FOR INFINIBAND RELEASE 1.1 DEVICES

InfiniBand Release 1.1 specifies fixed pre-emphasis or filtering rather than adaptive equalization when operating at 2.5 Gbits/second (SDR) for the first generation backplane interface. A fixed Total Jitter of 0.30 UI is allocated to the interconnect between the transmitter ASIC and the Receiver ASIC. Release 1.1 SDR cables incorporate equalization filters as necessary to limit total jitter.

6.5.2 EQUALIZATION FOR INFINIBAND RELEASE 1.2.1 DEVICES

Ports supporting Rel. 1.2.1 Enhanced Signaling use a combination of fixed or adaptive transmit equalization and variable receive equalization.

C6-1.2.7: InfiniBand Rel. 1.2.1 Enhanced Signaling connection equalization shall conform to the following restrictions for ports when operating at 2.5 Gb/s (SDR)

- Backplane connections shall use driver de-emphasis to compensate for the maximum loss as specified in [Table 72 on page 327](#).
- Cable ports shall use the specified driver de-emphasis to compensate for the maximum cable loss allowed. This is necessary for compatibility with legacy Receivers. It is recommended that between 1.5 and 6 dB of driver de-emphasis be used on a port with an attached active cable.
- Receive equalization may be provided however no training pattern is available other than the SDR initialization sequence as defined in [Chapter 5: Link/Phy Interface](#).
- Equalization for copper cables may be provided external to the transmitter for those connections with legacy transmitters.

C6-1.2.8: InfiniBand Rel. 1.2.1 Enhanced Signaling connection equalization shall conform to the following restrictions for ports when operating at 5.0 Gb/s (DDR) or 10 Gb/s (QDR).

- The transmitter shall provide the specified de-emphasis or equalization. It is recommended that between 1.5 and 6 dB of driver de-emphasis be used on a port with an attached active cable.
- The transmitter shall transmit a training pattern for the receive equalizer as defined in [Chapter 5: Link/Phy Interface](#).
- Filters shall not be used in the cable or backplane. The allowed interconnect characteristics are defined in [Section 6.8, “Compliant Channels,” on page 323](#).

For legacy devices with no access to cable attenuation information, transmitter equalization is defined statically in a manner that will appropriately equalize the full range of compliant channels. For some very long or very short passive copper channels, the single-setting transmit equalization method may produce symbol errors, due to under- or over-equalization for the two channel types, respectively. Similar errors can be seen with fiber optic cables, which convert the over-equalized input signal into non-compensable random jitter. Implementations using a single static transmit equalization should avoid

very short or very long passive copper as well as fiber topologies unless a mechanism is provided to adjust the transmit equalization for the channel.

The host device transmitter is required to choose the optimal transmitter equalization for any given channel. The portion of the channel that is to be equalized by the transmitter is defined by the cable type and described in [Table 47](#). For all cable types, the receiver is responsible for equalizing its receive card loss. In cases where cable attenuation information is available to the transmitter such as in a QSFP-based implementation, the transmitter may implement the attenuation-based algorithm (ABA). In this algorithm, the transmitter chooses an appropriate transmit equalization for the channel based on the cable's attenuation.

In cases where the transmitter is using the ABA, the transmitter shall obtain the attenuation of the cable from the cable EEPROM and use that value in its transmit equalization calculations.

Table 47 Equalization responsibilities

Cable type	Tx card	Connector	Cable	Connector	Rx card
Linear passive copper	Tx responsible for equalizing card + cable				Rx responsible for equalizing
Linear active copper	Tx responsible for equalizing card + cable Cable contains linear re-driver in Tx-side and/or Rx-side connector to reduce effective overall loss of the cable.				Rx responsible for equalizing
Full limiting active copper or fiber (Active Optical)	Tx responsible for equalizing	Redriver		Redriver	Rx responsible for equalizing
Near-end limiting active copper	Tx responsible for equalizing	Redriver	Redriver responsible for equalizing cable		Rx responsible for equalizing
Far-end limiting active copper	Tx responsible for equalizing cable + cable			Redriver	Rx responsible for equalizing

When choosing the transmit equalization to apply, the transmitter should consider the attenuation of its host board, the cable type, the attenuation reported by the cable, and whether or not the cable is indicated to be equalized. In cases where the cable type indicates a near-end re-driver, the transmitter shall not consider the cable attenuation in its transmit equalization calculations. The receiver is responsible for equalizing its receive card and any variations not accounted for by the transmit equalization, regardless of cable type.

The cable type and attenuation value are described in [Chapter 8: Management Interface](#).

6.5.3 EQUALIZATION FOR INFINIBAND RELEASE 1.3 DEVICES

For ports supporting Rel. 1.3 signaling at FDR and higher data rates, the receiver instructs the transmitter as to what equalization to apply from among the following three choices:

- A specified transmit equalization preset
- An incremental increase or decrease in transmit equalization from the current setting
- Transmitter-defined transmit equalization

Equalization step sizes, ranges, and presets are defined in [Section 6.6, “Differential Driver Outputs,” on page 282](#) for the applicable data rate. Link training state machine implementation details can be found in [Chapter 5: Link/Phy Interface](#).

For QDR and DDR data rates, the transmitter shall determine the optimum equalization for the channel using the attenuation-based algorithm as described in [Section 6.5.2](#) or an alternate method.

For SDR data rates, a fixed transmit equalization as described in [Section 6.5.1](#) may be employed.

6.6 DIFFERENTIAL DRIVER OUTPUTS

6.6.1 GENERAL REQUIREMENTS

For QDR and slower speeds, transmitter parameters are specified at the board side of the backplane or cable connector, TP2 in [Figure 79 on page 275](#) or TP6 in [Figure 80 on page 275](#) unless otherwise specified. Values at the SERDES (ASIC) pin are informative.

For FDR and higher speeds, transmitter parameters are specified at TP6a in [Figure 81 on page 275](#), unless otherwise specified.

C6-1.2.9: All output ports shall comply with the parameters and notes of [Table 48 Host Driver Characteristics for 2.5 Gb/s \(SDR\)](#) using appropriate parameters as noted while operating at SDR. The parameters are defined in terms of values at IB port pins.

In addition to the parameters defined for “normal” IB signaling from device to device defined as Port Type 1, additional more restrictive parameters are defined as Port Type 2 for 1x Pluggable Devices such as Optical Transceivers in the Small Form Factor Pluggable (SFP) form factor in order to be compatible with the MSA for that package.

C6-1.2.10: 1x Pluggable ports shall meet the jitter J_{T2} and J_{D2} defined for Port Type 2 in addition to the other parameters of [Section 6.6.2, “Host Driver Output Characteristics for SDR,” on page 285](#).

o6-1.2.1: Any output port claiming InfiniBand Rel. 1.2 Enhanced Signaling compliance shall comply with the parameters and notes of [Section 6.6.3, “Host Driver Output Characteristics for DDR,” on page 287](#) while operating at DDR.

o6-1.2.2: Any output port claiming InfiniBand Rel. 1.2 Enhanced Signaling compliance shall comply with the parameters and notes of [Section 6.6.4, “Host Driver Output Characteristics for QDR,” on page 290](#) while operating at QDR.

o6-1.2.3: Any output port claiming InfiniBand Rel. 1.3 compliance shall comply with the parameters and notes of [Section 6.6.5, “Host Driver Output Characteristics for FDR,” on page 293](#) while operating at FDR.

o6-1.2.4: Any output port claiming InfiniBand Rel. 1.3.1 compliance shall comply with the parameters and notes of [Section 6.6.6, “Host Driver Output Characteristics for EDR,” on page 302](#) while operating at EDR.

Some interface specifications use a diamond-shaped eye mask, with one X (time) parameter and 2 Y (differential amplitude) parameters, shown in [Figure 86 on page 284](#). Some use a hexagonal eye mask, with two X (time) parameters and 2 Y (differential amplitude) parameters, shown in [Figure 87 on page 284](#). Receivers at SDR, DDR, and QDR signaling rates use a diamond-shaped eye mask with other specifications, as shown in [Figure 94 on page 312](#).

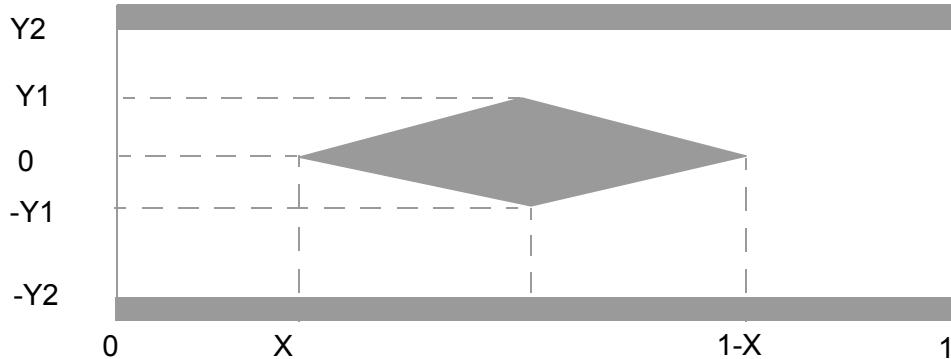


Figure 86 Diamond-shaped eye mask

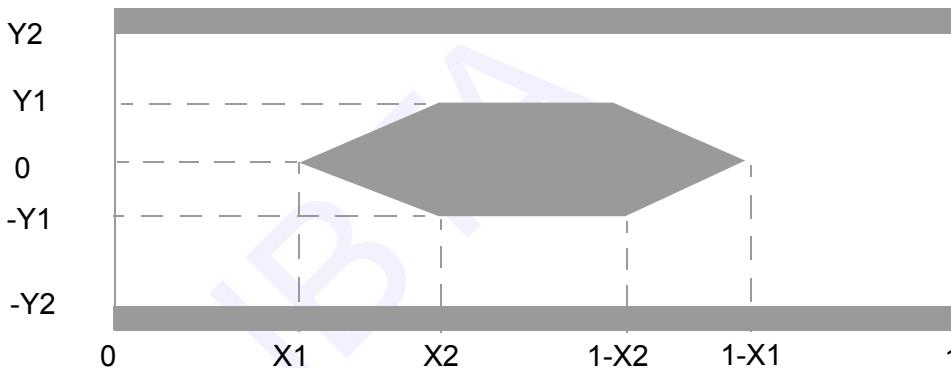


Figure 87 Hexagonal eye mask

A diamond-shaped eye mask is used to specify driver output characteristics at QDR, DDR, and SDR speeds. Diamond-shaped eye mask parameters are also used to specify the input and output signal specifications for Limiting Active cables at DDR and QDR speeds, as shown in [Section 6.8.5.2 on page 332](#) and [Section 6.8.6.2 on page 335](#), and for the output signal specifications for Limiting Active cables at FDR speeds, as described in [Section 6.8.7.2.2 on page 342](#).

At FDR and EDR speeds, driver output characteristics are specified in terms of a hexagonal eye mask. Hexagonal eye mask parameters are also used to describe the input signal specifications for Limiting Active cables at FDR and EDR speeds, as described in [Section 6.8.7.2.1 on page 341](#), and [Section 6.8.8.2.1 on page 348](#).

6.6.2 HOST DRIVER OUTPUT CHARACTERISTICS FOR SDR

The SDR driver characteristics are specified so as to guarantee the specified differential signal characteristics at the receiver as described in [Section 6.7.5, "Host Receiver input characteristics for SDR," on page 311](#) within defined InfiniBand topologies. [Table 48](#) defines the driver output characteristics for SDR rate links. Some driver parameters are not specified, but can be derived from other specified parameters. The eye mask in [Figure 86 on page 284](#) illustrates the definitions of parameters X, Y1, and Y2.

Table 48 Host Driver Characteristics for 2.5 Gb/s (SDR)

Symbol	Parameter	Maximum	Minimum	Unit	Notes
V_{CM}	Output Common Mode Voltage (backplane connections)	1.2	0.30	V	$(V_{high}+V_{low})/2$ The common mode is undefined if DC blocking capacitors are used.
V_{diff} (SERDES)	Differential output voltage At Transmit SerDes pins (Informative)	1.6	1.0	V	Differential unsigned waveform amplitude into 100 Ohm differential load.
X	eye mask parameter, time	0.15		UI	At TP6
Y1	eye mask parameter, voltage		0.45	V	At TP6; Differential unsigned waveform amplitude into 100 Ohm differential load.
Y2	eye mask parameter, voltage	0.8		V	FDR-capable devices: See Note 13.
V_{diffb}	Differential output voltage At TP1; (normative) (backplane connections)	1.6	0.80	V	Differential unsigned waveform amplitude into 100 Ohm differential load.
$V_{disabled}$	Disabled Mode output	1.6	0	V	The output in disabled or quiescent state may be zero volts differential. (Note 5)
$V_{standby}$	Standby Mode output	1.6	0	V	(Note 5)
t_{DTRF}	Driver Transition Time		30	ps	at 20-80% at the connector pins into 100 Ohm load
I_{ACCM}	AC Common Mode current (RMS)	5		uA	Determined by EMI restrictions and shielding effectiveness. 30 MHz to 6.25 GHz. (Notes 3, 6)
V_{ACCM}	AC Common Mode Voltage (RMS)	25		mV	(Notes 3, 6, 7)
S_{DD22}	Differential Output Return Loss	-10		dB	differential mode (Notes 1, 11)
S_{11}	Single Ended Output Return Loss (Note 1)	-8		dB	Single ended, either output, current - 15 to 15 mA, both pins driven, 50 Ohm reference impedance. (Note 11)
Z_{SEDC}	Single Ended Output Impedance-- Low Frequency.	10,000	30	Ω	Single ended, either output, current - 15 to 15 mA, both pins driven. (Notes 10, 11, 12)

Table 48 Host Driver Characteristics for 2.5 Gb/s (SDR) (Continued)

Symbol	Parameter	Maximum	Minimum	Unit	Notes
S_{DC22}	Common Mode to Differential return loss (mode conversion)	-20		dB	Reflected common mode that appears as differential (Note 11)
I_{Dshort}	Short Circuit Current	100	-100	mA	To any voltage between 1.6 and ground, power on or off.
S_{DBtB}	Skew	500		ps	Between any two physical lanes within a single transmitter.
J_D	Deterministic Jitter	0.17		UI	Without pre-emphasis (Notes 2, 8)
J_T	Total Jitter	0.35		UI	At +/- 7σ (10^{-12} BER) (Note 8)
J_{D2}	Deterministic Jitter	0.09		UI	Without pre-emphasis, Port Type 2 (Note 8, Note 9)
J_{T2}	Total Jitter	0.24		UI	At +/- 7σ , Port Type 2 (Note 8, Note 9)
UI_D	Unit Interval	400	400	ps	+/- 100 ppm measured over a minimum of 10,000 UI.

1. At the InfiniBand connector pins from 100 MHz to 1.875 GHz.
2. A transmitter which implements pre emphasis shall meet the receiver mask with the minimum and maximum allowable interconnect configuration.
3. The AC common mode voltage is limited in order to control radiated EMI. Some causes of AC common mode are skew between the true and complement outputs of the differential driver, mismatched levels between the true and complement outputs of the differential driver, and power supply or ground noise such as that caused by switching activity being conducted through the driver.
4. The output level is increased above the minimum required by the specified backplane attenuation to allow for cable attenuation as well as provide headroom for noise. Amplitude is measured for transition bits.
5. [Chapter 5: Link/Phy Interface](#) defines states in which the transmitter is quiescent. In these states driving node is inactive and there are no transitions on the link. Driver outputs shall be static in these states.
6. Common mode current is measured on an IB cable assembly using a detector bandwidth of 120 kHz below 1 GHz and 1 MHz above 1 GHz. All physical lanes shall be transmitting the idle pseudo-random character sequence.
7. Since cable shield effectiveness is in excess of 40 dB and common mode impedance is approximately 50 Ohms, driver common mode voltage of 25 mV is specified
8. Jitter measurement methods are defined in IBTA CIWG Methods of Implementation documents.
9. Port Type 2, see Architecture Note in [Section 7.4.1](#).
10. DC to 100 MHz
11. Driver impedances selected to adequately absorb reflections and other noise Return Loss measured with respect to 100 ohms, 100 MHz to 3.75 GHz
12. In the absence of DC blocking capacitors. Undefined if DC blocking capacitors present.
13. For FDR device transmitters operating at SDR speeds, the specifications are $Y_1 = 0.25$ V & $Y_2 = 0.6$ V except when driving limiting active cables capable of FDR or faster, when they are 0.095 V and 0.35 V. See [Table 55 on page 301](#). As described in [Section 6.1 on page 265](#), FDR hosts and cables are not required to be interoperable with SDR-only host and cables.

6.6.3 HOST DRIVER OUTPUT CHARACTERISTICS FOR DDR

[Table 49 on page 287](#) defines the driver output characteristics for DDR rate links using all cable types. [Table 50 on page 289](#) defines the requirements for driving active cables. The eye mask in [Figure 86](#) illustrates the definitions of parameters X, Y1, and Y2.

Table 49 Host Driver Characteristics for 5.0 Gb/s (DDR)

Symbol	Parameter	Maximum	Minimum	Unit	Notes
V_{CM}	Common Mode Voltage (Backplane connections)	1.2	0.75	V	$(V_{high}+V_{low})/2$ The common mode is undefined if DC blocking capacitors are used. (Note 12)
V_{diff}	Differential output voltage (SERDES pins, informative)	1.6	0.8	V	Differential unsigned waveform amplitude into 100 Ohm differential load. (Note 4)
X	eye mask parameter, time	0.15		UI	At TP6
Y1	eye mask parameter, voltage		0.33	V	At TP6; Differential unsigned waveform amplitude into 100 Ohm differential load.
Y2	eye mask parameter, voltage	0.8		V	FDR: See Note 13.
V_{diffb}	Differential output voltage (TP1, normative) (backplane connections)	1.6	0.60	V	Differential unsigned waveform amplitude into 100 Ohm differential load (Note 4)
$V_{disabled}$	Disabled Mode output	1.6	0	V	The output in disabled or quiescent state may be zero volts differential. DC value. (Note 5)
$V_{standby}$	Standby Mode output	1.6	0	V	DC value (Note 5)
t_{DTRF}	Driver Transition Time		30	ps	At 20-80% at the package pins into 100 Ohm load
I_{ACCM}	AC Common Mode current (RMS)	5		uA	Determined by EMI restrictions and shielding effectiveness. 30 MHz to 6.25 GHz (Notes 3, 6, 22)
V_{ACCM}	AC Common Mode voltage (RMS)	25		mV	(Notes 3, 6, 7)
S_{DD22}	Differential Output Return Loss	-10		dB	Differential mode (Note 11)
S_{22}	Single Ended Return Loss	-8		dB	Single ended, either output, current -15 to 15 mA, both pins driven. 50 Ohms reference impedance (Note 11)
Z_{SEDC}	Single Ended Output Impedance--Low Frequency	10,000	30	Ω	Single ended, either output, current -15 to 15 ma, both pins driven (Notes 10, 11, 12)

Table 49 Host Driver Characteristics for 5.0 Gb/s (DDR) (Continued)

Symbol	Parameter	Maximum	Minimum	Unit	Notes
S _{DC22}	Common Mode to Differential return loss (mode conversion)	-20		dB	Common mode reflected as differential (Note 11)
I _{DShort}	Short Circuit Current	100	-100	mA	To any voltage between 1.6 and ground, power on or off.
S _{DBtB}	Skew	500		ps	Between any two physical lanes within a single transmitter.
J _{D1}	Deterministic Jitter	0.15		UI	(Notes 2, 8)
J _{T1}	Total Jitter	0.30		UI	At +/- 7 σ (Note 8)
UI _D	Unit Interval	200	200	ps	+/- 100 ppm measured over a minimum of 10,000 UI

1. Obsolete.
 2. Obsolete.
 3. The AC common mode voltage is limited in order to control radiated EMI. Some causes of AC common mode are skew between the true and complement outputs of the differential driver, mismatched levels between the true and complement outputs of the differential driver, and power supply or ground noise such as that caused by switching activity being conducted through the driver.
 4. The output level is increased above the minimum required by the specified backplane attenuation to allow for cable attenuation as well as provide headroom for noise. Amplitude is measured for the first bit in a run.
 5. [Chapter 5: Link/Phy Interface](#) defines states in which the transmitter is quiescent. In these states driving node is inactive and there are no transitions on the link. Driver outputs shall be static in these states.
 6. Common mode current is measured on an IB cable assembly using a detector bandwidth of 120 kHz below 1 GHz and 1 MHz above 1 GHz. All physical lanes shall be transmitting the idle pseudo-random character sequence.
 7. Since cable shield effectiveness is in excess of 40 dB and common mode impedance is approximately 50 Ohms, driver common mode voltage of 25 mV is specified
 8. Jitter is measured as defined in IBTA CIWG Methods of Implementation documents.
 9. N/A - obsolete.
 10. DC to 100 MHz
 11. Driver impedances selected to adequately absorb reflections and other noise Return loss with respect to 100 ohms 100 MHz to 6.25 GHz
 12. In the absence of DC blocking capacitors. Undefined if DC blocking capacitors present.
 13. For FDR device transmitters operating at DDR speeds, the specifications are Y1 = 0.33 V and Y2 = 0.6 V.

6.6.3.1 HOST DRIVER OUTPUT CHARACTERISTICS FOR DRIVING DDR LIMITING ACTIVE CABLES

Table 50 defines the driver output characteristics for DDR rate links using active cables. The eye mask in Figure 87 on page 284 illustrates the definitions of parameters X1, X2, Y1, and Y2.

Table 50 Host Driver Characteristics for driving 5.0 Gb/s (DDR) limiting cables

Symbol	Parameter	Maximum	Minimum	Unit	Notes
X	eye mask parameter, time	0.15		UI	
Y1	eye mask parameter, voltage		0.25	V	At TP6; Differential unsigned waveform amplitude into 100 Ohm differential load. (Notes 1, 3, 4, 5)
Y2	eye mask parameter, voltage	0.6		V	
J _{D1}	Deterministic Jitter	0.15		UI	(Note ^a)
J _{T1}	Total Jitter	0.30		UI	At +/- 7 σ (Note ^b)
UI _D	Unit Interval	200	200	ps	+/- 100 ppm measured over a minimum of 10,000 UI

a. Jitter is measured as defined in IBTA CIWG Methods of Implementation documents.

b. Jitter is measured as defined in IBTA CIWG Methods of Implementation documents.

6.6.4 HOST DRIVER OUTPUT CHARACTERISTICS FOR QDR

[Table 51 on page 290](#) defines the driver output characteristics for QDR rate links using all cable types. [Table 52 on page 292](#) defines the requirements for driving active cables. The eye mask in [Figure 86 on page 284](#) illustrates the definitions of parameters X, Y1, and Y2.

Table 51 Host Driver Characteristics for 10 Gb/s (QDR)

Symbol	Parameter	Maximum	Minimum	Unit	Notes
V _{CM}	Common Mode Voltage Backplane connection	1.2	0.75	V	(V _{high} +V _{low})/2 The common mode is undefined if DC blocking capacitors are used. Note 12
V _{diff}	Differential output (Note 4) (SERDES pins, informative)	1.6	0.6	V	Differential unsigned waveform amplitude into 100 Ohm differential load See Notes 13,14,15
X	eye mask parameter, time	0.15		UI	At TP6
Y1	eye mask parameter, voltage (normative)		0.25	V	At TP6; Differential unsigned waveform amplitude into 100 Ohm differential load (Notes 13,14,15)
Y2	eye mask parameter, voltage (normative)	0.8		V	
V _{diffb}	Differential output (TP1, normative) (backplane connections)	1.6	0.45	V	Differential unsigned waveform amplitude into 100 ohm differential load (Notes 4, 13,14,15)
V _{disable}	Disabled Mode output	1.6	0	V	The output in disabled or quiescent state may be zero volts differential.DC (Note 5)
V _{standby}	Standby Mode output	1.6	0	V	DC (Note 5)
t _{DRF}	Driver Transition Time		30	ps	at 20-80% at the connector pins into 100 Ohm load
I _{ACCM}	AC Common Mode current (RMS)	5		uA	Determined by EMI restrictions and shielding effectiveness. 30 MHz to 6.25 GHz (Notes 3, 6)
V _{ACCM}	AC Common Mode Voltage (RMS)	25		mV	(Notes 3, 6, 7)
S _{DD22}	Differential Output Return Loss	-8		dB	Differential mode (Notes 1, 11.)

Table 51 Host Driver Characteristics for 10 Gb/s (QDR) (Continued)

Symbol	Parameter	Maximum	Minimum	Unit	Notes
S ₂₂	Single Ended Return Loss	-8		dB	Single ended, either output, current -15 to 15 mA, both outputs driven. 50 Ohms reference impedance. (Notes 1, 11)
Z _{SEDC}	Single Ended Output Impedance--Low Frequency	10,000	30	W	Single ended, either output, current -15 to 15 mA, both pins driven. (Notes 10, 11, 12)
S _{DC22}	Common Mode to Differential return loss (mode conversion)	-20		dB	Common mode reflected as differential. See Note 11
I _{Dshort}	Short Circuit Current	100	-100	mA	To any voltage between 1.6 and ground, power on or off.
S _{DBtB}	Skew	500		ps	Between any two physical lanes within a single transmitter.
J _{D1}	Deterministic Jitter	.15		UI	(Notes 2, 8)
J _{T1}	Total Jitter	.30		UI	At +/- 7 σ (Note 8)
UI _D	Unit Interval	100	100	ps	+/- 100 ppm measured over 10,000 UI

1. Obsolete.
2. Obsolete.
3. The AC common mode voltage is limited in order to control radiated EMI. Some causes of AC common mode are skew between the true and complement outputs of the differential driver, mismatched levels between the true and complement outputs of the differential driver, and power supply or ground noise such as that caused by switching activity being conducted through the driver.
4. Amplitude is measured for the first bit in a run if a Backplane port, for all bits of a Cable Port.
5. [Chapter 5: Link/Phy Interface](#) defines states in which the transmitter is quiescent. In these states driving node is inactive and there are no transitions on the link. Driver outputs shall be static in these states.
6. Common mode current is measured on an IB cable assembly using a detector bandwidth of 120 kHz below 1 GHz and 1 MHz above 1 GHz. All physical lanes shall be transmitting the idle pseudo-random character sequence.
7. Since cable shield effectiveness is in excess of 40 dB and common mode impedance is approximately 50 Ohms, driver common mode voltage of 25 mV is specified
8. Jitter is measured as defined in IBTA CIWG Methods of Implementation (MOI) documents.
9. N/A - obsolete
10. DC to 100 MHz
11. Driver impedances selected to adequately absorb reflections and other noise. Return Loss with respect to 100 ohms 100 MHz to 12.5 GHz.
12. In the absence of DC blocking capacitors. Undefined if DC blocking capacitors present.
13. Refer to IEEE Std.181-2003 for definitions and procedures around un-signed waveform amplitude measurements. (amplitude unsigned.)
14. Waveform amplitude measurements shall be conducted only on consecutive transition bits, over a 2UI Epoch (IEEE Std. 181-2003) unit interval, to a minimum population of 10E4 UI's. This test should apply to any unsigned consecutive transition bit in the test pattern.
15. Top and Base level calculations should be determined based on the histogram mean technique as described in IEEE Std. 181-2003 (Determining State Levels from the Histogram).

6.6.4.1 HOST DRIVER OUTPUT CHARACTERISTICS FOR DRIVING QDR LIMITING ACTIVE CABLES

Table 50 defines the driver output characteristics for QDR rate links using limiting active optical and limiting active copper cables. The eye mask in Figure 86 on page 284 illustrates the definitions of parameters X, Y1, and Y2.

Table 52 Host Driver Characteristics for driving 10.0 Gb/s (QDR) limiting cables

Symbol	Parameter	Maximum	Minimum	Unit	Notes
X	eye mask limit, time	0.15		UI	
Y1	eye mask parameter, voltage		0.25	V	At TP6; Differential unsigned waveform amplitude into 100 Ohm differential load
Y2	eye mask parameter, voltage	0.6		V	
J _{D1}	Deterministic Jitter	0.15		UI	(Note ^a)
J _{T1}	Total Jitter	0.30		UI	At +/- 7 σ (Note ^b)
UI _D	Unit Interval	100	100	ps	+/- 100 ppm measured over a minimum of 10,000 UI

a. Jitter is measured as defined in IBTA CIWG Methods of Implementation documents.

b. Jitter is measured as defined in IBTA CIWG Methods of Implementation documents.

6.6.5 HOST DRIVER OUTPUT CHARACTERISTICS FOR FDR

The FDR host transmitter output specifications are defined with reference to similar specifications for data communications at speeds between 10.3 and 28 Gb/s defined in IEEE 802.3-2015, OIF-CEI-03.0, and ANSI T11 FC-PI-5.

The transmitter is assumed to have the capabilities of a three tap FIR filter with pre- and post-cursor taps and the ability to adjust all three tap weights independently. Tap weights can be adjusted either incrementally, or with the use of 16 Preset equalization settings. Preset 0 is nominally a “fully-equalized” setting, where the driver equalization fully compensates for the S-parameters of the host PCB, such that the output at TP6 is suitable for driving a limiting active optical or copper cable. Preset 1 is an unequalized setting, which is required for measurement calibration. Presets 2-15 have various pre-cursor and post-cursor values, encompassing the full required range of driver equalization capabilities. Additionally, two values of the amplitude setting, AMP, accommodate channels with either high or low loss, reducing required receiver dynamic range.

The host transmitter shall comply with the specifications in [Table 53](#) for all transmit equalization settings and cable types except where indicated. All parameters are defined at TP6a unless otherwise indicated. Test patterns used to characterize various parameters are summarized in [Section 6.4.2. “Compliance Testing Hardware and Methodology.” on page 277](#), and specified in detail in the relevant CIWG MOI documents.

Defined preset equalization settings and output amplitude settings are specified in [Table 54 on page 299](#). Further specifications applicable to Preset 0, including eye mask parameters defined in [Figure 87](#), are given in [Table 55 on page 301](#).

Unless otherwise explicitly stated, for all FDR time-domain specifications (including eye, jitter, DDPWS, transition time, etc.), the waveform concerned is observed through a 17 GHz low-pass filter response (such as a Bessel-Thomson response), and includes the effects of the HCB and/or MCB as appropriate.

Table 53 Host Driver Characteristics for 14.0625 Gb/s (FDR)

Symbol	Parameter	Reference	Max	Min	Unit	Comment
V _{se}	Single-ended output voltage		4.0	-0.3	V	
V _{diffb}	Differential output (TP1, normative)		See Table 54 on page 299			
V _{diffpp}	Differential peak to peak output voltage, driver disabled	802.3-2015 85.8.3	30		mV	
	Termination mismatch		5		%	1 MHz
S _{DD22}	Differential output return loss	802.3-2015 85.8.3.1	see Eq. 1		dB	50 MHz to 14.1 GHz
S _{CC22}	Common mode output return loss		-2		dB	200 MHz to 14.1 GHz
S _{DC22}	Common mode to differential reflection		Eq. 2		dB	50 MHz to 14.1 GHz
V _{cm}	AC common mode output voltage (RMS)	802.3-2015 86A.5.3.1	30		mV	Except Preset 0
	Far end transmit output noise (low insertion loss channel)	802.3-2015 85.8.3.2	2		mV	Except Preset 0,
	Far end transmit output noise (high insertion loss channel)		1		mV	See reference for test point
	Equalizer coefficient step size	802.3-2015 85.8.3.3.2	0.05	0.0083		Except Preset 0 See Section 6.6.5.4 on page 299
	Txpre_cursor full scale range	802.3-2015 85.8.3.3.3		2.33		See Table 54 on page 299 and Section 6.6.5.5 on page 300
	Txpost_cursor full scale range	802.3-2015 85.8.3.3.3		7		
RJ	Random jitter		0.15		UI	BER = 1x10 ⁻¹²
DCD	Duty Cycle Distortion		0.035		UI	
TJ-DDJ	Total jitter, excluding data dependent jitter		0.25		UI	
	Peak linear fit pulse response	Section 6.6.5.2		0.6 x transmitter steady state output voltage	V	Preset 1 only Note ^a 1.
e	Normalized error (linear fit)	Section 6.6.5.2	0.037			Except Preset 0

a. Note that Steady-state output voltage is called “DC amplitude” in the 802.3-2015 documentation.

The differential return loss requirement is defined in [Equation 1](#) and shown in [Figure 88](#). Note that this definition is equivalent, other than frequency range limits, to the definition in IEEE 802.3bm-2015, Equation (83E-5).

$$(S_{DDxx}(f)) \leq \begin{cases} -9.5 + 0.37(f), & 0.05 \leq f < 8 \\ -4.75 + 7.4 \cdot \log 10\left(\frac{f}{14}\right), & 8 \leq f < 14.1 \end{cases} \quad \text{Eq. 1}$$

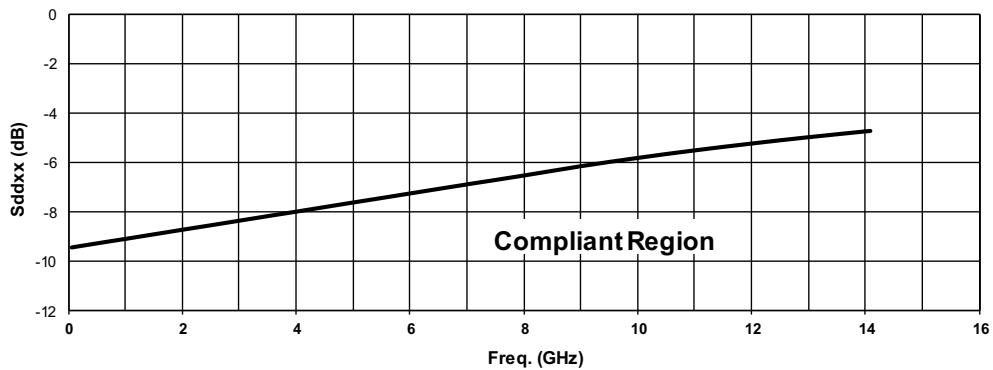


Figure 88 Limits on S_{DD11} and S_{DD22} vs. Frequency for FDR Hosts and Cables

The differential to common mode input return loss requirement is defined in [Equation 2](#) and shown in [Figure 89](#). Note that this definition is equivalent, other than frequency range limits to the definition in IEEE 802.3bm-2015, Equation (83E-6).

$$(S_{DCxx}(f)) \leq \begin{cases} -22 + 20\left(\frac{f}{25.78}\right), & 0.01 \leq f < 12.89 \\ -15 + 6\left(\frac{f}{25.78}\right), & 12.89 \leq f < 14.1 \end{cases} \quad \text{Eq. 2}$$

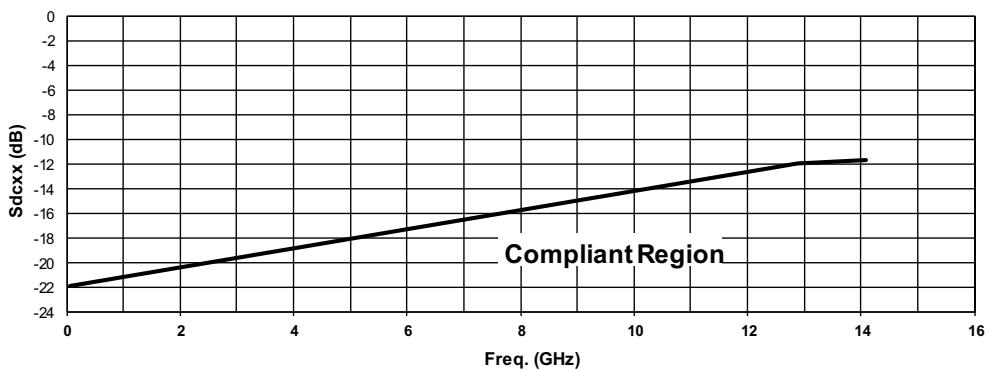


Figure 89 Limits on S_{DC11} and S_{DC22} vs. Frequency for FDR Hosts and Cables

6.6.5.1 FDR TRANSMITTER EQUALIZER IMPLEMENTATION

Transmit equalization coefficients are based on the FIR filter structure shown in [Figure 90 on page 296](#). C_{-1} and C_{+1} are the coefficients used in the FIR equation and represent the pre-cursor and post-cursor, respectively. C_0 represents the cursor coefficient setting and is a positive entity. Transmitter tap weights are adjusted during the equalization procedure in which a receiver may ask the transmitter to change one, two or all three FIR taps. FDR output driver presets shall be as defined in [Table 54 on page 299](#).

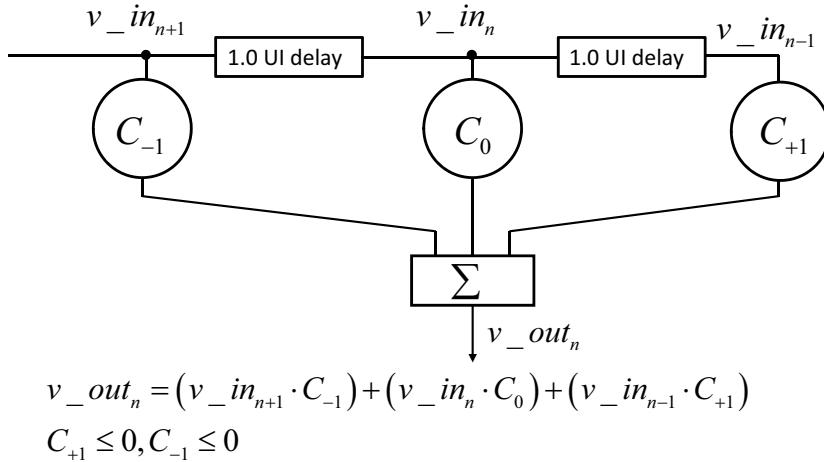


Figure 90 Tx Equalization FIR Representation

6.6.5.2 TRANSMIT EQUALIZATION MEASUREMENT

Given that the transmitter output can not be directly probed at the SERDES pins, and given that the equalizer tap weights cannot be directly measured, the transmitter's output waveform for various equalization settings is calculated from measurements made at the HCB output, comparing unequalized and equalized waveforms for a PRBS9 bit pattern.

Measurement of the transmit equalization for the purpose of transmitter characterization is done using the Host Compliance Boards described in [Section 6.4.2, “Compliance Testing Hardware and Methodology,” on page 277](#).

The method for measuring C_{-1} , C_0 and C_{+1} incorporates the following steps. The techniques for doing the calculations are only summarized here. They are described in full detail in IEEE 802.3-2015 clause 85.8.3.3.

- 1) The transmitter under test is configured for Preset 1 (i.e., $[C_{-1}, C_0, C_{+1}] = [0, 1, 0]$), and for the amplitude to be tested (amplitude bit=0 or 1).
- 2) Capture at least one complete cycle of the PRBS9 test pattern with a sample rate of M times the signaling rate of the transmitter under test, with M an integer ≥ 7 .

- 3) Compute the linear fit to the captured waveform and the linear fit pulse response $p(k)$ using the procedure described in IEEE 802.3-2015 clause 85.8.3.3.5. 1
- 4) Define t_x to be the time where the rising edge of the linear fit pulse, p , from step 3) crosses 50% of its peak amplitude. 2
- 5) Sample the linear fit pulse, $p(k)$, at symbol-spaced intervals relative to the time $t_0 = t_x + 0.5 \text{ UI}$, interpolating as necessary to yield the sampled pulse p_i . 3
- 6) Use p_i to compute the vector of coefficients, w , of a N_w -tap symbol-spaced transversal filter that equalizes for the transfer function from the driver to TP2 using the procedure described in IEEE 802.3-2015 clause 85.8.3.3.6. 4

To improve accuracy of the calculation, in recognition of the larger number of bit periods between the driver and TP2 for FDR and higher bit-rates relative to the 10.3125 GBd signaling rate assumed in IEEE 802.3-2015 clause 85.8.3.3, without significantly increasing the calculation complexity, the parameters of this calculation are defined as follows:

- N_p , the length of the pulse used for the linear fit, is set to 14 UI (802.3-2015 value: 8 UI). 14
- D_p , the delay used for the linear fit pulse response, is set to 2 UI (802.3-2015 value: 2 UI). 15
- N_w , the length of the symbol-spaced transversal filter, is set to 7 UI (802.3-2015 value: 7 UI). 16
- D_w , the delay between taps in the transversal filter, is set to 2 UI (802.3-2015 value: 1 UI). 17

To measure the tap weights at the various preset values, the following steps are performed.

- 7) Configure the transmitter under test as required by the test. 22
- 8) Capture at least one complete cycle of the PRBS9 test pattern with the same sample rate. 23
- 9) Compute the linear fit to the captured waveform and the linear fit pulse response $p(k)$ using the same procedure. 24
- 10) Define t_x to be the time where the rising edge of the linear fit pulse, p , from step 9) crosses 50% of its peak amplitude. 25
- 11) Sample the linear fit pulse, $p(k)$, at symbol-spaced intervals relative to the time $t_0 = t_x + 0.5 \text{ UI}$, interpolating as necessary to yield the sampled pulse p_i . 26
- 12) Equalize the sample pulse p_i using the coefficient vector computed in step 6), using the procedure described in IEEE 802.3-2015 clause 85.8.3.3.6, to yield the equalized pulse q_i . 27

For this calculation, because of the desire to accommodate realistic receiver DFE designs, the parameters for the linear fit are as follows:

- N_p , the length of the pulse used for the linear fit, is set to 14 UI (802.3-2015 value: 8 UI). 28

- D_p , the delay used for the linear fit pulse response, is set to 2 UI (802.3-2015 value: 2 UI).
- N_w , the length of the symbol-spaced transversal filter, is set to 7 UI (802.3-2015 value: 7 UI).
- D_w , the delay between taps in the transversal filter, is set to 2 UI (802.3-2015 value: 1 UI).

The C_{-1} , C_0 , and C_{+1} tap weights are read from the equalized pulse q_i as follows:

$$[C_{-1}, C_0, C_{+1}] = [q_i \text{ at time } t_0 + (D_p-1) \text{ UI}, q_i \text{ at time } t_0 + D_p \text{ UI}, q_i \text{ at time } t_0 + (D_p+1) \text{ UI}]$$

As specified in [Table 53](#), the peak of this linear fit pulse response must be ≥ 0.6 times the transmitter steady state output voltage, and the RMS value of the error between the captured waveform (normalized to the peak value of the pulse p) and the linear fit to the captured waveform from step [3](#)) shall be no greater than 0.037.

6.6.5.3 TRANSMIT EQUALIZATION PRESETS

There are 16 possible transmit equalization presets and 2 amplitude settings, for a total of 32 independent transmitter settings. [Table 54 on page 299](#) provides a summary of the 16 presets and the corresponding tap weights, as measured using the de-embedding technique described in [Section 6.6.5.2 on page 296](#).

The transmitter transmits signals using the equalization preset and amplitude, when requested to by the peer receiver through the use of TS3 DDS fields and the TS3 Amp bit.

The use of Preset 0 is recommended for cases where a copper or optical active cable with limiting amplifiers at both ends is used. The presence of limiting amplifiers in between Link End Nodes can make it difficult for the receiving Link End Node to determine the appropriate equalization settings for the transmitting Link End Node. If the receiver requests Preset 0, the transmitter shall choose the appropriate equalization setting for its own portion of the link and the receiver shall equalize its own portion of the link, as described in [Section 6.5.2, "Equalization for InfiniBand Release 1.2.1 Devices," on page 280](#). [Table 47 on page 281](#) describes the equalization responsibilities for the transmitter and receiver for various cable types when Preset 0 is used.

When Preset 0 is requested and a passive cable is connected, the transmitter may optionally adjust its output amplitude based on the requested amplitude range.

Tolerance requirements on each of the tap weight values (C_{-1} , C_0 and C_{+1}) is either ± 0.015 , or $\pm 7.5\%$ of the specified value, whichever is larger.

The steady state output voltage for amplitude bit = 0 for preset 1 (no equalization) shall be no larger than 70% of the steady state output voltage for amplitude bit = 1 for preset 1.

Table 54 Tx FIR Filter Coefficients and Amplitudes for 14.0625 Gb/s (FDR)

Preset	C_{-1}	C_0	C_{+1}	Precursor ratio	Postcursor ratio	Steady state output voltage range, amplitude bit=0 ^a	Steady state output voltage range, amplitude bit=1
0	Tx defined					As defined in Table 55	
1	0	1	0	1.00	1.0	400-600 mV	700-900 mV
2	0	0.90	-0.10	1.00	1.26		
3	-0.03	0.87	-0.10	1.07	1.27		
4	0	0.81	-0.19	1.00	1.61		
5	-0.02	0.79	-0.18	1.08	1.63		
6	-0.07	0.76	-0.18	1.26	1.69		
7	0	0.73	-0.27	1.00	2.22		
8	-0.02	0.71	-0.27	1.10	2.28		
9	-0.06	0.68	-0.26	1.33	2.42		
10	-0.10	0.66	-0.25	1.61	2.59		
11	-0.02	0.66	-0.33	1.12	3.09		
12	-0.06	0.63	-0.31	1.42	3.38		
13	-0.09	0.61	-0.30	1.81	3.76		
14	-0.05	0.58	-0.37	1.64	5.70		
15	-0.08	0.56	-0.36	2.36	6.97		

a. Steady state output voltage is the output voltage in the case of long strings of 1s and 0s.

6.6.5.4 TRANSMIT EQUALIZATION INCREMENTAL CHANGES

During equalization adjustments, the receiver may ask the transmitter to make an incremental change to one or more of the FIR taps using the Rev 1 TS3, as described in [Section 5.5.2.4.2 on page 147](#). The change request is indicated as either an increase, decrease or hold request for each tap. An increase request shall result in a normalized tap adjustment of between 0.0083 and 0.05 for the specified tap, and a decrease request shall result in a normalized tap adjustment between -0.05 and -0.0083, as specified in [Table 53](#). The transmitter is responsible for assuring that the output is not inverted as a result of receiver requests for incremental equalization changes.

6.6.5.5 TRANSMIT EQUALIZATION RANGES

The transmitter post-cursor ratio is calculated according to [Equation 3](#). The transmitter pre-cursor ratio, $Txpre$, is calculated according to [Equation 4](#).

$$Txpost = \frac{c_0 - c_1 + c_{-1}}{c_0 + c_1 + c_{-1}} \quad \text{Eq. 3}$$

$$Txpre = \frac{c_0 + c_1 - c_{-1}}{c_0 + c_1 + c_{-1}} \quad \text{Eq. 4}$$

The transmitter shall be capable of supporting Tx post_cursor values, $Txpost$, between 1 and 7 inclusive. Values greater than 7 are permissible but not required. This limit can be verified by setting C_{-1} to 0 and then adjusting C_0 and C_1 until the limit of device capability is reached.

The transmitter shall be capable of supporting $Txpre$ _cursor values between 1 and 2.36 inclusive. Values greater than 2.36 are permissible but not required. This limit can be verified by setting C_1 to 0 and then adjusting C_0 and C_{-1} until the limit of device capability is reached.

6.6.5.6 HOST TRANSMITTER SPECIFICATIONS FOR LIMITING ACTIVE CABLES WITH NEAR END REPEATER

When a limiting active cable with a near end repeater is connected, the host transmitter shall comply with the specifications in [Table 55](#). In cases where the same parameter is specified in [Table 55](#) and [Table 53](#), the [Table 55](#) specification shall take precedence.

The transmitter shall also comply with the specifications in [Table 55](#) when a receiver requests Preset 0. When a receiver requests Preset 0, the transmitter shall ignore the amplitude bit, as shown in [Table 54](#).

Each electrical output lane and signal of the FDR host when measured at TP6a shall meet the specifications of [Table 55 on page 301](#) while the specified crosstalk sources are applied to all lanes of the host's electrical input. [Figure 87 on page 284](#) illustrates the definitions of parameters X1, X2, Y1, and Y2. Parameters in [Table 55](#) are defined at TP6a unless otherwise stated, as observed through a 17 GHz low-pass filter response (such as a Bessel-Thomson response), and HCB and/or MCB as appropriate.

If the receiver specifies Preset 0, the transmitter is responsible for choosing the appropriate equalization to meet the output specifications described in [Table 55](#). Under these circumstances, the transmitter shall use the attenuation-based algorithm or some alternate method to select its transmit equalization as described in [Section 6.5.3 on page 282](#).

Table 55 FDR host output specifications at Preset 0, for Limiting Active Cables

Symbol	Parameter	Paragraph reference	Specification value(s)	Unit	Conditions
X1, X2	eye mask parameters, time		0.11, 0.31	UI	Hit ratio=5x10 ⁻⁵ ; Hexagonal mask, Figure 87 on page 284 .
Y1, Y2	eye mask parameters, voltage		95, 350	mV	When driving cables whose maximum data rate is FDR or EDR; Hit ratio=5x10 ⁻⁵
			250, 600		When driving cables whose max. data rate is SDR, DDR, or QDR; Hit ratio=1x10 ⁻¹²
	Crosstalk signal Vpk-pk, each aggressor		450 +/- 5%	mV	At TP7a Counter-propagating aggressor signals emulate active cable range 0 output signals.
	Crosstalk signal transition time, 90 mV - 360 mV (20%-80% nom.)		27 +/- 3	ps	

Symbol	Parameter	Paragraph reference	Max	Min	Unit	Conditions
t _r , t _f	Output transition time, 20%-80%		24	ps		Measured at the following locations in the PRBS9 test pattern: 11111111 1000001 11101...
DDPWS	Data Dependent Pulse Width Shrinkage	802.3-2015 86A.5.3.4	0.11		UI	
J2	J2 jitter	802.3-2015 86.8.3.3.1	0.19		UI	
J9	J9 jitter	802.3-2015 86.8.3.3.2	0.34		UI	
Q _{sq}	Signal to noise ratio	802.3-2015 86A.5.3.5		45	V/V	
V _{cm}	AC common mode output voltage (RMS)	802.3-2015 86A.5.3.1	20		mV	

6.6.6 HOST DRIVER OUTPUT CHARACTERISTICS FOR EDR

Similar to FDR, the EDR host transmitter output specifications are defined with reference to similar specifications for data communications at speeds between 10.3 and 28 Gb/s defined in IEEE 802.3-2015, OIF-CEI-03.0, and ANSI T11 FC-PI-5.

The transmitter is assumed to have the capabilities of a three tap FIR filter with pre- and post-cursor taps and the ability to adjust all three tap weights independently. Tap weights can be adjusted either incrementally, or with the use of 16 Preset equalization settings. Preset 0 is nominally a “fully-equalized” setting, where the driver equalization fully compensates for the S-parameters of the host PCB, such that the output at TP6 is suitable for driving a limiting active optical or copper cable. Preset 1 is an unequalized setting, which is required for measurement calibration. Presets 2-15 have various pre-cursor and post-cursor values, encompassing the full required range of driver equalization capabilities. Additionally, two values of the amplitude setting, AMP, accommodate channels with either high or low loss, reducing required receiver dynamic range.

The host transmitter shall comply with the specifications in [Table 56](#) for all transmit equalization settings and cable types except where indicated. All parameters are defined at TP6a unless otherwise indicated. Test patterns used to characterize various parameters are summarized in [Section 6.4.2. “Compliance Testing Hardware and Methodology.” on page 277](#), and specified in detail in the relevant CIWG MOI documents.

Defined preset equalization settings and output amplitude settings are specified in [Table 57 on page 306](#). Further specifications applicable to Preset 0, including hexagonal eye mask parameters defined in [Figure 87](#), are given in [Table 58 on page 308](#).

Unless otherwise explicitly stated, for all EDR time-domain specifications (including eye, jitter, DDPWS, transition time, etc.), the waveform concerned is observed through a 31 GHz low-pass filter response (such as a Bessel-Thomson filter response), and includes the effects of the HCB and/or MCB as appropriate.

Table 56 Host transmitter output characteristics for 25.78125 Gb/s (EDR)

Symbol	Parameter	Reference	Max	Min	Unit	Comment
V _{se}	Single-ended output voltage		4.0	-0.3	V	
V _{diffb}	Differential output (TP1, normative)		See Table 57 on page 306			
V _{diffpp}	Differential peak to peak output voltage, driver disabled	802.3-2015 85.8.3	30		mV	
	Termination mismatch		5		%	1 MHz
S _{DD22}	Differential output return loss	802.3-2015 85.8.3.1	see Eq. 5		dB	50 MHz to 26 GHz
S _{CC22}	Common mode output return loss		-2		dB	50 MHz to 26 GHz
S _{DC22}	Common mode to differential reflection		Eq. 6		dB	50 MHz to 26 GHz
V _{cm-ac}	AC common mode output voltage (RMS)	802.3-2015 86A.5.3.1	30		mV	Except Preset 0
V _{cm-dc}	DC common mode output voltage		2.8	-0.3	V	
	Far end transmit output noise (low insertion loss channel)	802.3-2015 85.8.3.2	2		mV	Except Preset 0, See reference for test point
	Far end transmit output noise (high insertion loss channel)		1		mV	
	Equalizer coefficient step size	802.3-2015 85.8.3.3.2	0.05	0.0083		Except Preset 0 See Section 6.6.5.4 on page 299
	Txpre_cursor full scale range	802.3-2015 85.8.3.3		2.33		See Table 57 on page 306 and Section 6.6.6 on page 307
	Txpost_cursor full scale range	802.3-2015 85.8.3.3		7		
RJ	Random jitter		0.15		UI	BER = 1x10 ⁻¹²
DCD	Duty Cycle Distortion		0.035		UI	
TJ-DDJ	Total jitter, excluding data dependent jitter		0.25		UI	
	Peak linear fit pulse response	Section 6.6.5.2		0.6 x transmitter steady state output voltage	V	Preset 1 only Note ^a 1.
e	Normalized error (linear fit)	Section 6.6.5.2	0.037			Except Preset 0

a. Note that Steady-state output voltage is called “DC amplitude” in 802.3-2015 documentation.

The differential return loss requirement is defined in [Equation 5](#) and shown in [Figure 91](#). Note that this definition is equivalent, other than frequency limits, to the definition in IEEE 802.3bm-2015, Equation (83E-5), and to [Equation 1](#).

$$(S_{DDxx}(f)) \leq \begin{cases} -9.5 + 0.37(f), & 0.05 \leq f < 8 \\ -4.75 + 7.4 \cdot \log 10\left(\frac{f}{14}\right), & 8 \leq f \leq 26 \end{cases} \quad \text{Eq. 5}$$

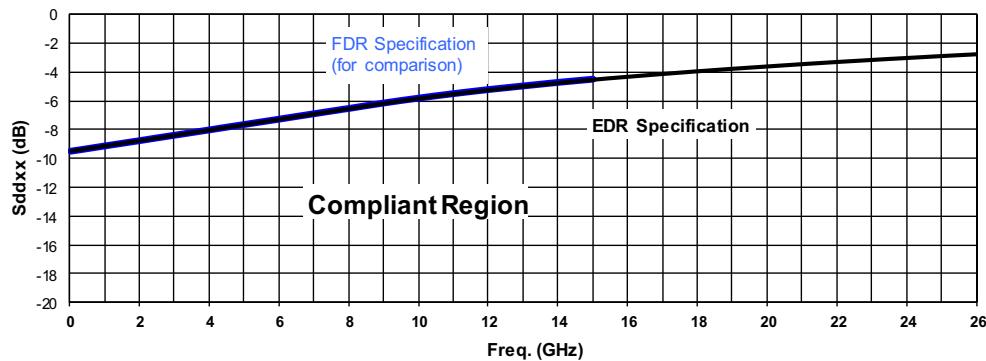


Figure 91 Limits on S_{DD11} and S_{DD22} vs. Frequency for EDR Hosts and Cables

The differential to common mode input return loss requirement is defined in [Equation 6](#) and shown in [Figure 92](#). Note that this definition is equivalent, other than frequency range limits, to the definition in IEEE 802.3bm-2015, Equation (83E-6), and to [Equation 2](#).

$$(S_{DCxx}(f)) \leq \begin{cases} -22 + 20\left(\frac{f}{25.78}\right), & 0.01 \leq f < 12.89 \\ -15 + 6\left(\frac{f}{25.78}\right), & 12.89 \leq f < 26 \end{cases} \quad \text{Eq. 6}$$

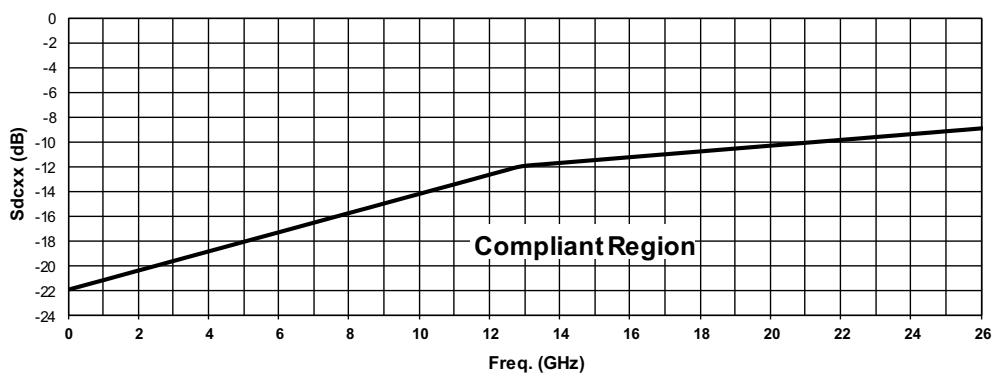


Figure 92 Limits on S_{DC11} and S_{DC22} vs. Frequency for EDR Hosts and Cables

6.6.6.1 EDR TRANSMITTER EQUALIZER IMPLEMENTATION

Transmit equalization coefficients are based on the same FIR filter structure described for the FDR speed, and shown in [Figure 90 on page 296](#) and described in [6.6.5.1 on page 296](#). EDR output driver presets shall be as defined in [Table 57 on page 306](#).

6.6.6.2 TRANSMIT EQUALIZATION MEASUREMENT

The method for measurement of transmitter equalization for EDR speed is identical to the method for FDR speed, as described in [6.6.5.2 on page 296](#), with the following exceptions or modifications.

- Measurement of the transmit equalization for the purpose of transmitter characterization is done using the Host Compliance Boards meeting loss specifications for EDR speed.

Note also that, as with FDR and for measurement simplicity and accuracy, the equalized transmitter signal is measured without a Continuous Time Linear Equalizer (CTLE) in the measurement path.

6.6.6.3 TRANSMIT EQUALIZATION PRESETS

Transmitter equalization presets at EDR speed are defined similarly to the transmitter equalization presets at FDR speed, as described in [6.6.5.3 on page 298](#). There are 16 possible transmit equalization presets and 2 amplitude settings, for a total of 32 independent transmitter settings. [Table 57 on page 306](#) provides a summary of the 16 presets and the corresponding tap weights, as measured using the de-embedding technique described in [Section 6.6.6.2 on page 305](#).

Table 57 Tx FIR Filter Coefficients and Amplitudes for 25.78125 Gb/s (EDR)

Preset	C_{-1}	C_0	C_{+1}	Precursor ratio	Postcursor ratio	Steady state output voltage range, amplitude bit=0 ^a	Steady state output voltage range, amplitude bit=1
0	Tx defined					As defined in Table 58	
1	0	1	0	1.00	1.0	400-600 mV	700-900 mV
2	0	0.90	-0.10	1.00	1.26		
3	-0.03	0.87	-0.10	1.07	1.27		
4	0	0.81	-0.19	1.00	1.61		
5	-0.02	0.79	-0.18	1.08	1.63		
6	-0.07	0.76	-0.18	1.26	1.69		
7	0	0.73	-0.27	1.00	2.22		
8	-0.02	0.71	-0.27	1.10	2.28		
9	-0.06	0.68	-0.26	1.33	2.42		
10	-0.10	0.66	-0.25	1.61	2.59		
11	-0.02	0.66	-0.33	1.12	3.09		
12	-0.06	0.63	-0.31	1.42	3.38		
13	-0.09	0.61	-0.30	1.81	3.76		
14	-0.05	0.58	-0.37	1.64	5.70		
15	-0.08	0.56	-0.36	2.36	6.97		

a. Steady state output voltage is the output voltage in the case of long strings of 1s and 0s.

6.6.6.4 TRANSMIT EQUALIZATION INCREMENTAL CHANGES

During equalization adjustments, the receiver may ask the transmitter to make an incremental change to one or more of the FIR taps using the Rev 1 TS3, as described in [Section 5.5.2.4.2 on page 147](#). The change request is indicated as either an increase, decrease or hold request for each tap. An increase request shall result in a normalized tap adjustment of between 0.0083 and 0.05 for the specified tap, and a decrease request shall result in a normalized tap adjustment between -0.05 and -0.0083, as specified in [Table 56](#). The transmitter is responsible for assuring that the output is not inverted as a result of receiver requests for incremental equalization changes.

6.6.6.5 TRANSMIT EQUALIZATION RANGES

The transmitter post-cursor ratio is calculated according to [Equation 7](#). The transmitter pre-cursor ratio, $Txpre$, is calculated according to [Equation 8](#).

$$Txpost = \frac{c_0 - c_1 + c_{-1}}{c_0 + c_1 + c_{-1}} \quad \text{Eq. 7}$$

$$Txpre = \frac{C_0 + C_1 - C_{-1}}{C_0 + C_1 + C_{-1}} \quad \text{Eq. 8}$$

The transmitter shall be capable of supporting Tx post_cursor values, $Txpost$, between 1 and 7 inclusive. Values greater than 7 are permissible but not required. This limit can be verified by setting C_{-1} to 0 and then adjusting C_0 and C_1 until the limit of device capability is reached.

The transmitter shall be capable of supporting $Txpre$ _cursor values between 1 and 2.36 inclusive. Values greater than 2.36 are permissible but not required. This limit can be verified by setting C_1 to 0 and then adjusting C_0 and C_{-1} until the limit of device capability is reached.

6.6.6.6 HOST TRANSMITTER SPECIFICATIONS FOR LIMITING ACTIVE CABLES WITH NEAR END REPEATER

When a limiting active cable with a near end repeater is connected, the host transmitter shall comply with the specifications in [Table 58](#). In cases where the same parameter is specified in [Table 58](#) and [Table 56](#), the [Table 58](#) specification shall take precedence.

The transmitter shall also comply with the specifications in [Table 58](#) when a receiver requests Preset 0. When a receiver requests Preset 0, the transmitter shall ignore the amplitude bit, as shown in [Table 57](#).

Each electrical output lane and signal of the EDR host when measured at TP6a shall meet the specifications of [Table 58 on page 308](#) while the specified crosstalk sources are applied to all lanes of the host's electrical input. [Figure 87 on page 284](#) illustrates the definitions of parameters X1, X2, Y1, and Y2. Parameters in [Table 58](#) are defined at TP6a unless otherwise stated, as observed through a 31 GHz low-pass filter response (such as a Bessel-Thomson response), and HCB and/or MCB as appropriate.

If the receiver specifies Preset 0, the transmitter is responsible for choosing the appropriate equalization to meet the output specifications described in [Table 58](#). Under these circumstances, the transmitter shall use the attenuation-based algorithm or some alternate method to select its transmit equalization as described in [Section 6.5.3 on page 282](#).

As shown in the J2 and J9 jitter specification rows in [Table 58](#), the transmitter's output jitter specifications can be relaxed if the transmitter is driving to a limiting active cable or module with TX CDR (Transmitter Clock and Data Recovery) circuit enabled.

Table 58 EDR host output specifications at Preset 0, for Limiting Active Cables

Symbol	Parameter	Paragraph reference	Specification value(s)	Unit	Conditions ^a
	Crosstalk signal Vpk-pk, each aggressor		450 +/- 5%	mV	At TP7a
	Crosstalk signal transition time, 90 mV - 360 mV (20%-80% nom.)		17 +/- 3	ps	Counter-propagating aggressor signals emulate active cable range 0 output signals.

Symbol	Parameter	Paragraph reference	Max	Min	Unit	Conditions
V _{cm}	AC common mode output voltage (RMS)	802.3-2015 86A.5.3.1	20		mV	
t _r , t _f	Output transition time, 20%-80%			13	ps	Measured at the following locations in the PRBS9 test pattern: <u>11111111100000111101...</u>
EH15	Eye Height, 1E-15			120	mV	
EW15	Eye Width, 1E-15		0.53		UI	When driving a limiting active cable or module with TX CDR enabled
			0.71		UI	When driving a limiting active cable or module with TX CDR bypassed (i.e., disabled)

a. EH15/EW15 measured with PRBS9, observed through Host Compliance Board, with specified scope BW and CTLE in scope reference receiver software.

6.7 DIFFERENTIAL RECEIVER INPUTS

6.7.1 GENERAL REQUIREMENTS

Receivers are specified at the board side of the backplane or cable connector, TP4 in [Figure 79 on page 275](#) or TP8 in [Figure 81](#). For FDR and higher speeds, transmitter parameters are specified at TP8a in [Figure 81 on page 275](#), unless otherwise specified. Values at the SERDES (ASIC) pin are informative.

C6-1.2.11: All input ports shall comply with the parameters and notes of [Table 59 on page 311](#) while operating at SDR. Parameters apply to the pin type as noted. The parameters are defined in terms of values at ASIC pins. They may be measured at accessible test points, with the values adjusted appropriately as defined in [Section 6.4.1, “Compliance Points,” on page 274](#). A BER of 10^{-12} shall be achieved when connected to the worst case transmitter through any compliant channel.

o6-1.2.1: All input ports claiming InfiniBand Rel. 1.2.1 Enhanced Signaling compliance shall comply with the parameters and notes of [Table 60 Host Receiver Characteristics for 5.0 Gb/s \(DDR\)](#) while operating at DDR. A BER of 10^{-12} shall be achieved when connected to the worst case transmitter through any compliant channel.

o6-1.2.2: All input ports claiming InfiniBand Rel. 1.2.1 Enhanced Signaling compliance shall comply with the parameters and notes of [Table 61 Host Receiver Characteristics for 10 Gb/s \(QDR\)](#) while operating at QDR. A BER of 10^{-12} shall be achieved when connected to the worst case transmitter through any compliant channel.

o6-1.2.3: All input ports claiming InfiniBand Rel. 1.3 compliance shall comply with the parameters and notes of [Table 63 Host Receiver characteristics for 14.0625 Gb/s \(FDR\)](#), [Table 64 FDR stressed input signal specifications - linear cables](#), and [Table 65 FDR receiver stressed input signal specifications - limiting cables](#) while operating at FDR. A BER of 10^{-12} shall be achieved when connected to the worst case transmitter through any compliant channel.

o6-1.2.4: All input ports claiming InfiniBand Rel. 1.3.1 compliance shall comply with the parameters and notes of [Table 63 Host Receiver characteristics for 14.0625 Gb/s \(FDR\)](#), [Table 64 FDR stressed input signal specifications - linear cables](#), and [Table 65 FDR receiver stressed input signal specifications - limiting cables](#) while operating at EDR. A BER of 10^{-12} shall be achieved when connected to the worst case transmitter through any compliant channel.

6.7.2 BEACON SIGNALING

C6-1.2.12: The beaconing sequence is described in detail in [Chapter 5: Link/Phy Interface](#). Devices which power down in X_{Sleep} state shall detect the beaconing sequence while operating on Auxiliary power. All devices shall detect the beaconing sequence. The beaconing sequence is transmitted at SDR only.

C6-1.2.13: The beacon detection shall meet the following requirements: The beaconing detector shall detect a minimum unsigned waveform amplitude of 175 mV p-p as a valid

signal present. Unsigned waveform amplitudes less than 85 mV p-p shall be considered absent. Signals with transitions only at less than 10 MHz shall be considered absent. The beaconing sequence is transmitted at the in band signaling rate of 2.5 Gbits/second with an active period of 2 ms and a quiescent period of 100 ms (See [Section 5.8.4.2, "Polling States," on page 197](#)).

o6-1.2.1: Note: Beaconing occurs at 2.5 Gb/s. All links begin the initialization process at 2.5 Gb/s. Continued link connection is verified by periodic transmission and reception of the "Link Heartbeat" ordered-set.

o6-1.2.2: All input ports capable of operating at FDR or higher speeds shall detect a minimum unsigned waveform amplitude of 100 mV p-p (i.e., voltage parameter Y1 = 50 mV in the eye mask of [Figure 86](#)). Unsigned waveform amplitudes less than 50 mV p-p shall be considered absent. Note that eye mask voltage parameter minimum cable output swing Y1 corresponds to half of these values.

6.7.3 SIGNALING CAUTIONS

Note that there are many circumstances in InfiniBand in which no differential voltage will be applied at the IB receiver. Examples include sleeping links, disconnected cables, and partially used ports resulting from width negotiation. Therefore means shall be provided to disable or de-gate the data coming from such a receiver to assure that false error signals are not generated. This gating shall be maintained across speed changes as appropriate. Note that the heartbeat will be absent in a disconnected link.

6.7.4 TERMINATION

[Figure 93](#) shows the recommended circuit schematic for proper signal termination. Differential and common mode (odd and even mode) termination is important to dampen noise and reflections in both modes.

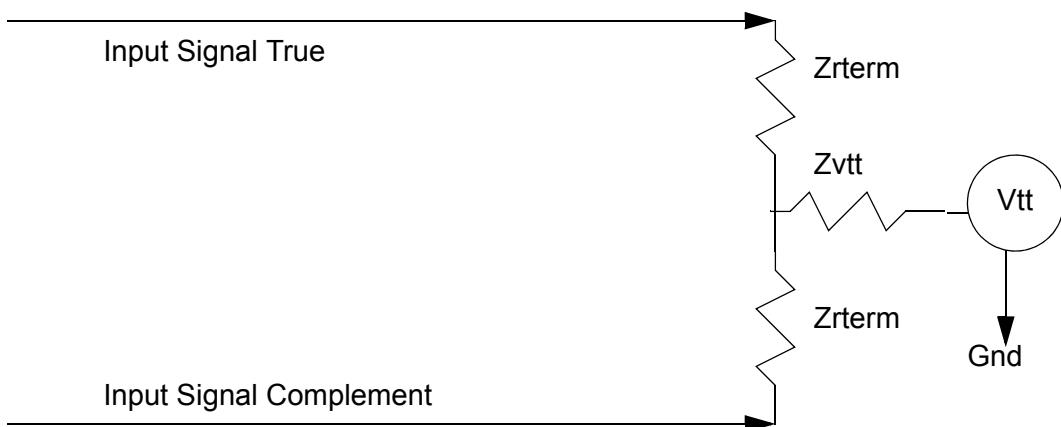


Figure 93 Termination and Signaling

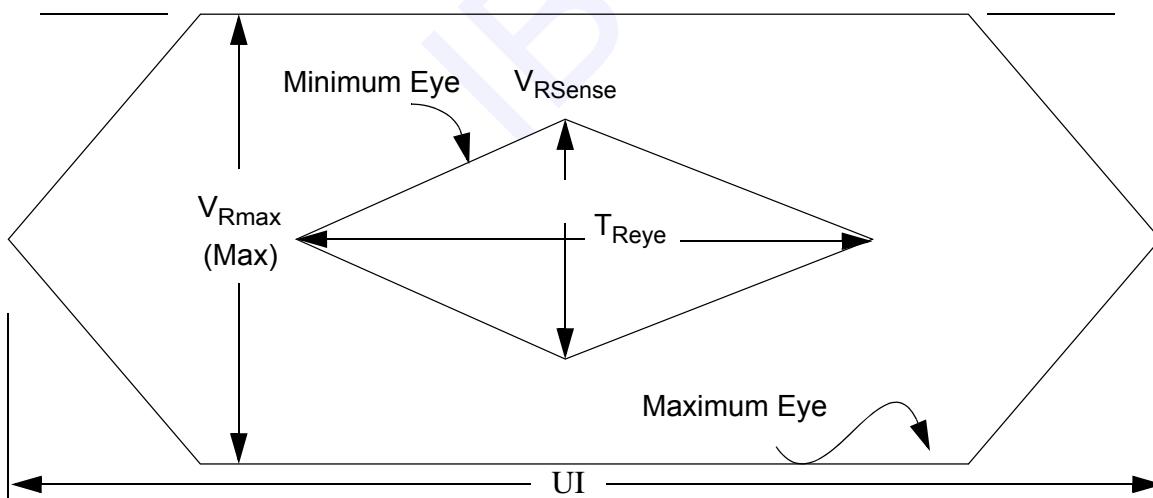
6.7.5 HOST RECEIVER INPUT CHARACTERISTICS FOR SDR**Table 59 Host Receiver Characteristics for 2.5 Gb/s (SDR)**

Symbol	Parameter	Maximum	Minimum	Unit	Conditions or Remarks
BER	Bit Error Ratio	10^{-12}			With minimum input
J_{DR}	Deterministic Jitter at Receiver	0.47		UI	at TP8
J_{TR}	Total Jitter at Receiver	0.65		UI	at 10^{-12} BER
Z_{RTerm}	Termination	62.5	40	Ω	To V_{tt} (differential impedance is double) See Figure 93 .
V_{tt}	Termination Voltage	1.0	0.5	V	See Figure 93 See Notes 3, 6
S_{DC11}	Mode Conversion return loss		30	dB	The plus and minus rails of the signal are each terminated to V_{tt} and must be matched to avoid common mode to differential conversion. See Note 4. Measured at TP1, TP6.
Z_{Vtt}	Vtt Impedance	30	0	Ω	See Note 4.
Z_{Vtt}	Vtt Impedance	10,000	0	Ω	See Notes 5, 6 Use of large values can reduce excess power caused by Vcm conflicts between driver and receiver.
L_{DR}	Differential Return Loss		10	dB	See Note 4
L_{CMR}	Common Mode Return Loss		6	dB	See Note 4
V_{RSense}	Input Sensitivity		175	mV	Minimum differential unsigned waveform amplitude. See Notes 1, 7, 8 See Figure 94 on page 312
V_{RSD}	Signal Threshold		85	mV	Minimum differential unsigned waveform amplitude. See Note 2 and Figure 94 on page 312
V_{max}	Maximum Input Voltage	1.6		V	Maximum differential unsigned waveform amplitude.
V_{RCM}	Common Mode Voltage See Note 6	1.25	0.25	V	$(V_{high}+V_{low})/2$ See Note 3.
I_{ROff}	Off Current	50	-50	mA	Max. current into a pin with power off. Max voltage 1.6

Table 59 Host Receiver Characteristics for 2.5 Gb/s (SDR) (Continued)

Symbol	Parameter	Maximum	Minimum	Unit	Conditions or Remarks
V_{RHP}	Hot Plug Voltage (Voltage applied with power off or on)	1.6	-0.5	V	Applied without damage to any IB connector signal pin.
t_{REye}	Eye width	140	ps		Note 9 and Figure 94 on page 312 .
S_{RBtB}	Total Skew	24	ns		See Section 6.8.3 on page 325

1. Signals meeting the InfiniBand input specification shall be received with a maximum bit error ratio of 1×10^{-12} .
 2. Signals having an un-signed differential amplitude less than VRSD shall be ignored. Signals having an un-signed differential amplitude greater than VRSD and less than VRSense may be ignored.
 3. Unless DC blocking capacitors are present between the termination and the InfiniBand module pins.
 4. Over a frequency range of 100 MHz to 1.875 GHz.
 5. Frequency from DC to 100 MHz.
 6. In the absence of DC blocking capacitors. Undefined if DC blocking capacitors present.
 7. Refer to IEEE Std.181-2003 (Amplitude, Waveform, Unsigned) for definitions and measurement procedures.
 8. Waveform amplitude measurement's shall be conducted only on consecutive transition bits, over a 2UI Epoch (IEEE Std. 181-2003), over a minimum population of $10E3$ UI's. This test should apply to any unsigned consecutive transition bit in the test pattern.
 9: Top and Base level calculation should be determined based on the histogram mean technique as described in IEEE Std. 181-2003 (Determining State Levels from the Histogram)

**Figure 94 Eye Opening at receiver for SDR, DDR, & QDR signaling (Differential)**

6.7.6 HOST RECEIVER INPUT CHARACTERISTICS FOR DDR

Receiver input specifications for DDR for linear cables are shown in [Table 60](#). Specifications for limiting cables are shown in [Table 62](#).

Table 60 Host Receiver Characteristics for 5.0 Gb/s (DDR)

Symbol	Parameter	Maximum	Minimum	Unit	Conditions or Remarks
BER	Bit Error Ratio	10^{-12}			Driven by compliant transmitter through compliance channel
S_{DD11}	Differential return loss	8	dB		Reference to 100 ohms
S_{CC11}	Common Mode Return Loss	6	dB		reference to 30 ohms. Note 4
V_{tt}	Termination Voltage	1.2	1.0	V	See Figure 93 and Notes 3 and 6
S_{DC11}	Mode Conversion at input	-20		dB	The plus and minus rails of the signal are each terminated to V_{tt} and must be matched to minimize common mode to differential conversion. See Note 4 Measured at IB connector pins.
Z_{Vtt}	Vtt Impedance	0		Ω	Note 5, Note 6 Use of larger values can reduce excess power caused by Vcm conflicts between driver and receiver.
V_{RSense}	Input Sensitivity (informative, at SERDES input)	80	mV		Minimum differential unsigned waveform amplitude. See Notes 1, 2, 7, 8 and Figure 94 on page 312
V_{CSense}	Input Sensitivity at cable connector (informative)	90	mV		Minimum differential unsigned waveform amplitude. Note 1, 2, 7, 8, 10 See Figure 94 on page 312
V_{BSense}	Input Sensitivity at backplane connector (informative)	95	mV		Minimum differential unsigned waveform amplitude. See Notes 1, 2, 7, 8, 10 and Figure 94 on page 312
V_{rmax}	Maximum Input Voltage	1.6		V	Maximum differential unsigned waveform amplitude.
V_{RCM}	Common Mode Voltage	0.70	1.2	V	$(V_{high}+V_{low})/2$. See Notes 3, 6.
I_{Roff}	Off Current	50	-50	mA	Max. current into a pin with power off. Max 1.6 V

Table 60 Host Receiver Characteristics for 5.0 Gb/s (DDR) (Continued)

Symbol	Parameter	Maximum	Minimum	Unit	Conditions or Remarks
V_{RHP}	Hot Plug Voltage (Voltage applied with power off or on)	1.6	-0.5	V	Applied without damage to any IB connector signal pin.
t_{REye}	Eye width	N/A	N/A	ps	Note 9
S_{RBtB}	Total Skew	12		ns	See Section 6.8.3 on page 325

1. Signals meeting the InfiniBand input specification shall be received with a maximum bit error ratio of 10^{-12} .
 2. Signals amplitude informative only. DDR signal amplitude determined by transmitter and channel specifications.
 3. Unless DC blocking capacitors are present between the termination and the InfiniBand module pins.
 4. Over a frequency range of 100 MHz to 3.75 GHz.
 5. Frequency from DC to 100 MHz.
 6. In the absence of DC blocking capacitors. Undefined if DC blocking capacitors present.
 7. Refer to IEEE Std.181-2003 (Amplitude, Waveform, Unsigned) for definitions and measurement procedures.
 8. Waveform amplitude measurement's shall be conducted only on consecutive transition bits, over a 2 UI Epoch (IEEE Std. 181-2003), over a minimum population of 1000 UI's. This test should apply to any unsigned consecutive transition bit in the test pattern.
 9: At DDR speed, eye width may not be meaningful, as eye diagrams may be fully closed at cable connector and SERDES input. Required waveform at receiver is determined by requirements on driver and compliant channel, specified in [Table 49 Host Driver Characteristics for 5.0 Gb/s \(DDR\) on page 287](#), and [Table 75 Linear Channel S Parameter Requirements for 5.0 Gb/s \(DDR\) on page 331](#) respectively.
 10. Does not supersede Compliant channel requirement.

6.7.7 HOST RECEIVER INPUT CHARACTERISTICS FOR QDR

Receiver input specifications for DDR for linear cables are shown in [Table 61](#). Specifications for limiting cables are shown in [Table 62](#).

Table 61 Host Receiver Characteristics for 10 Gb/s (QDR)

Symbol	Parameter	Maximum	Minimum	Unit	Conditions or Remarks
BER	Bit Error Ratio	10 ⁻¹²			Driven by compliant transmitter through compliance channel
V _{tt}	Termination Voltage	1.2	0.9	V	See Figure 93 See Notes 3, 6
S _{D211}	Differential Return Loss		8	dB	Note 4
S _{C211}	Common Mode Return Loss		6	dB	Note 4
S _{CD}	Mode Conversion, Common mode to differential	-20		dB	The plus and minus rails of the signal are each terminated to V _{tt} and must be matched to avoid common mode to differential conversion. See Note 4 Measured at IB connector pins.
Z _{Vtt}	V _{tt} Impedance	10,000	0	Ω	See Note 5, 6 Use of large values can reduce excess power caused by V _{cm} conflicts between driver and receiver.
V _{RSense}	Input Sensitivity (informative, at serdes pins)		60	mV	Minimum differential unsigned waveform amplitude. See Notes 1, 2, 7, 8 and Figure 94 on page 312
V _{CSense}	Input Sensitivity at cable connector (informative)		70	mV	Minimum differential unsigned waveform amplitude. See Notes 1, 2, 7, 8, 10 and Figure 94 on page 312
V _{BSense}	Input Sensitivity at Backplane connector (informative)		75	mV	Minimum differential unsigned waveform amplitude. See Notes 1, 2, 7, 8, 10 and Figure 94 on page 312
V _{rmax}	Maximum Input Voltage	1.6		V	Maximum differential unsigned waveform amplitude.
V _{RCM}	Common Mode Voltage	1.25	.8	V	(V _{high} +V _{low})/2 See Note 3, 6
I _{ROff}	Off Current	50	-50	mA	Max. current into a pin with power off.

Table 61 Host Receiver Characteristics for 10 Gb/s (QDR) (Continued)

Symbol	Parameter	Maximum	Minimum	Unit	Conditions or Remarks
V_{RHP}	Hot Plug Voltage (Voltage applied with power off or on)	1.6	-0.5	V	Applied without damage to any IB connector signal pin.
t_{REye}	Eye width	N/A	N/A	ps	See Note 9
S_{RBtB}	Total Skew	6		ns	See Section 6.8.3 on page 325

1. Signals meeting the InfiniBand input specification shall be received with a maximum bit error ratio of 10^{-12} .
2. Signals amplitude informative only. QDR signal amplitude determined by transmitter and channel specifications.
3. Unless DC blocking capacitors are present between the termination and the InfiniBand module pins.
4. Over a frequency range of 100 MHz to 7.5 GHz.
5. Frequency from DC to 100 MHz.
6. In the absence of DC blocking capacitors. Undefined if DC blocking capacitors present.
7. Refer to IEEE Std.181-2003 (Amplitude, Waveform, Unsigned) for definitions and measurement procedures.
8. Waveform amplitude measurement's shall be conducted only on consecutive transition bits, over a 2 UI Epoch (IEEE Std. 181-2003), over a minimum population of 1000 UI's. This test should apply to any unsigned consecutive transition bit in the test pattern.
9. At QDR speed, eye width may not be meaningful, as eye diagrams may be fully closed at cable connector and SERDES input. Required waveform at receiver is determined by requirements on driver and compliant channel, specified in [Table 51 Host Driver Characteristics for 10 Gb/s \(QDR\) on page 290](#), and [Table 79 Linear Channel S Parameter Requirements for 10 Gb/s \(QDR\) on page 334](#) respectively.
10. Does not supersede Compliant channel requirement.

Table 62 Host Receiver Characteristics for 5.0 Gb/s and 10 Gb/s limiting cables

Symbol	Parameter	Maximum	Nominal	Minimum	Unit	Conditions
X	Eye mask parameter, Time	0.36			UI	$\pm 7\sigma$ (BER = 1E-12) measured at TP7
Y1,Y2	Eye mask parameters, voltage		100, 600		mV	
J_{D1}	Deterministic Jitter	0.40			UI	At TP7
J_{T1}	Total Jitter	0.72			UI	$\pm 7\sigma$ (BER = 1E-12) measured at TP7
S_{DD11}	Differential output return loss	Eq. 12 on page 336				measured at TP7

a. An FDR-capable active cable, if requested to by the host, may comply with the reduced Y1,Y2 amplitude specifications described in [Table 85, "FDR limiting active cable output electrical specifications," on page 342](#) while operating at QDR data rate.

6.7.8 HOST RECEIVER INPUT CHARACTERISTICS FOR FDR

An FDR receiver must meet the common electrical requirements described in [Table 63 on page 317](#), as well as the requirements described for linear cables in [Section 6.7.8.1](#) and limiting active cables in [Section 6.7.8.2](#). The FDR receiver requirements follow closely the definition for 16 GFC as described in ANSI T11 FC-PI-5 and IEEE 802.3-2015, with scaling as appropriate for data rate differences. The receiver is assumed to have the capabilities of an adaptive decision feedback equalizer (DFE) with a minimum of five taps.

Table 63 Host Receiver characteristics for 14.0625 Gb/s (FDR)

Symbol	Parameter	Paragraph reference	Maximum	Minimum	Unit	Comment
	Bit error ratio		1E-12			
S _{RBtB}	Lane-to-lane skew		9.387		ns	See Section 6.8.3 on page 325
	Differential peak to peak input amplitude tolerance	FC-PI-5 9.3.1		1200	mV	Equivalent to Y2 = 600 mV in eye mask voltage.
V _{CM}	AC Common mode voltage tolerance (RMS)			33	mV	
S _{DD11}	Differential input return loss	FC-PI-5 9.3.3, 802.3-2015 85.8.4.1	See Equation 1 on page 295		dB	50 MHz-14.1 GHz
S _{DC11}	Common mode to differential reflection		See Equation 2 on page 295		dB	50 MHz-14.1GHz

6.7.8.1 RECEIVER TESTS FOR USE WITH LINEAR CABLES

Receivers for FDR shall meet the required bit error ratio when tested using stressed input signals as specified in [Table 64](#), using the test patterns specified in the relevant CIWG MOI documents. These signals are generated by passing a signal from a pattern generator through a low loss channel (test 1) and a maximum loss channel with less noise degradation (test 2). The values of the loss have been rounded up by 0.16 dB to allow for the fact that the channel ILD is not at worst case. The apportionment of the loss to the various coefficients is based on the 16GFC specification with the a2 term increased by 0.1 for test 2 so that the total loss using the assigned max coefficients is approximately 0.5 dB larger than the allowed maximum value.

The pattern generator has 800 mV-pp output amplitude (same as 16GFC and 802.3-2015) and 47 ps rise/fall times (same as 802.3-2015). If the rise and fall times t_r of the pattern generator are less than 47 ps the value of a4 in [Table 64](#) is increased by $\Delta a4$ from [Equation 9](#) where t_r is the rise time, in ps.

$$\Delta a4 = 60.51 \times 10^{-6} (35^2 - t_r^2) \quad \text{Eq. 9}$$

Table 64 FDR stressed input signal specifications - linear cables

Symbol	Parameter	Paragraph reference	Test 1 value	Test 2 value	Unit	Comments
	Insertion loss at 7.03 GHz	802.3-2015 85.8.4.2, FC-PI-5 9.7.3.3	10.5	19.71	dB	Test 2: 15.00 (Cable+2xMCB IL) - 0.55 (1x MCB IL) + 5.26 (assumed host IL) = 19.71 dB
	Fitted insertion loss coefficients a0 a1 a2 a4	802.3-2015 85.8.4.2, FC-PI-5 9.7.3.3	0 1.7 0.8 0.02	0 3.2 1.4 0.04	dB $(\text{dB}/\text{GHz})^{0.5}$ dB/GHz $(\text{dB}/\text{GHz})^2$	
ILD	Insertion loss deviation		$\pm(0.5 + 0.2*f)$	$\pm(0.5 + 0.3*f)$	dB, f in GHz	Allows deviation at 7.03 GHz, of ± 1.9 dB for Test 1, and ± 2.6 dB for Test 2
	Applied RJP-p	802.3-2015 85.8.4.2, FC-PI-5 9.7.1	0.15	0.15	UI	
	Applied DCDP-p	802.3-2015 85.8.4.2, FC-PI-5 9.7.1	0.035	0.035	UI	
	Applied TJ	802.3-2015 85.8.4.2, FC-PI-5 9.7.1	0.25	0.25	UI	
	Calibrated far end crosstalk (RMS)	802.3-2015 85.8.4.2	6.3	4.1	mV	
	Calibrated ICN (includes FEXT) (RMS)	802.3-2015 85.8.4.2, FC-PI-5	10	5.5	mV	

6.7.8.2 RECEIVER TESTS FOR USE WITH LIMITING CABLES

Receivers for FDR shall meet the required bit error ratio when tested using a PRBS31 pattern and a stressed input signal as specified in [Table 65](#) which just meets the inner eye amplitude. These signals are generated using the test system shown in [Figure 198](#) on page 616 and set up as described in IEEE 802.3-2015 Annex 86A.

Table 65 FDR receiver stressed input signal specifications - limiting cables

Symbol	Parameter	Test Value	Unit	Conditions
X	eye mask parameter, time	0.30	UI	Hit ratio=5E-5 with 100 Ohm load at TP7a
Y1, Y2	Differential unsigned input voltage tolerance range 0 (required) range 1 (optional) range 2 (optional)	50, 225 100, 350 150, 450	mV	Note: These specifications emulate limiting active cable range 0 output
tr	Transition time	17	ps	20-80%, Measured at this PRBS9 test pattern transition 11111111 <u>1</u> 00000111101...
DDPWS	Data Dependent Pulse Width Shrinkage	0.34	UI	At TP7a
J2	J2 jitter	0.44	UI	At TP7a, if RX CDR bypass is bypassed (i.e., disabled)
		0.19	UI	At TP7a, if RX CDR bypass is enabled
J9	J9 jitter	0.69	UI	At TP7a, if RX CDR bypass is bypassed (i.e., disabled)
		0.34	UI	At TP7a, if RX CDR bypass is enabled
	Crosstalk signal Vpk-pk	450	mV	At TP7a. Co-propagating aggressor. ^a Transition time measured at this PRBS9 test pattern transition 11111111 <u>1</u> 00000111101...
	Crosstalk signal transition time, 20%-80%	17	ps	
	Crosstalk calibration signal Vpk-pk	700	mV	At TP6a Counter-propagating aggressor, apply during crosstalk calibration only
	Crosstalk calibration signal transition time, 20%-80%	24	ps	

a. Host input crosstalk tolerance specifications are based on the requirement that FDR limiting active cables support range 0 differential unsigned output voltage. This host input spec shall be considered a minimum requirement for hosts; hosts may optionally choose to support a wider input range for FDR limiting active cables.

6.7.9 HOST RECEIVER INPUT CHARACTERISTICS FOR EDR

An EDR receiver must meet the common electrical requirements described in [Table 66 on page 320](#), as well as the requirements described for linear cables in [Section 6.7.9.1](#) and limiting active cables in [Section 6.7.9.2](#). The EDR receiver requirements follow closely the definition for 16 GFC as described in ANSI T11 FC-PI-5 and IEEE 802.3-2015, with scaling as appropriate for data rate differences. The receiver is assumed to have the capabilities of an adaptive decision feedback equalizer (DFE) with a minimum of five taps.

Table 66 Host Receiver characteristics for 25.78125 Gb/s (EDR)

Symbol	Parameter	Paragraph reference	Maximum	Minimum	Unit	Comment
	Bit error ratio		1E-12			
S _{RBtB}	Lane-to-lane skew		5.120		ns	See Section 6.8.3 on page 325
	Differential peak to peak input amplitude tolerance	FC-PI-5 9.3.1		1200	mV	Equivalent to Y2 = 600 mV in eye mask voltage.
V _{CM}	AC Common mode voltage tolerance (RMS)			33	mV	
S _{DD11}	Differential input return loss	FC-PI-5 9.3.3, 802.3-2015 85.8.4.1	See Equation 5 on page 304		dB	50 MHz-26 GHz
S _{DC11}	Common mode to differential reflection		See Equation 6 on page 304		dB	50 MHz-26 GHz

6.7.9.1 RECEIVER TESTS FOR USE WITH LINEAR CABLES

Receivers for EDR shall meet the required bit error ratio when tested using stressed input signals as specified in [Table 67](#), using the test patterns specified in the relevant CIWG MOI documents. These signals are generated by passing a signal from a pattern generator through a low loss channel (test 1) and a maximum loss channel with less noise degradation (test 2). The values of the loss have been rounded up by 0.16 dB to allow for the fact that the channel ILD is not at worst case. The apportionment of the loss to the various coefficients is based on the 16GFC specification with the a2 term increased by 0.1 for test 2 so that the total loss using the assigned max coefficients is approximately 0.5 dB larger than the allowed maximum value.

The pattern generator has 800 mV-pp output amplitude (same as 16GFC and 802.3-2015) and 47 ps rise/fall times (same as 802.3-2015). If the rise and fall times t_r of the pattern generator are less than 47 ps the value of a4 in [Table 67](#) is increased by $\Delta a4$ from [Equation 10](#) where t_r is the rise time, in ps.

$$\Delta a4 = 60.51 \times 10^{-6} (35^2 - t_r^2) \quad \text{Eq. 10}$$

Table 67 EDR stressed input signal specifications - linear cables

Symbol	Parameter	Paragraph reference	Test 1 value	Test 2 value	Unit	Comments
	Insertion loss at 12.89 GHz	802.3-2015 85.8.4.2, FC-PI-5 9.7.3.3	10.5	19.71	dB	Test 2: 15.00 (Cable+2xMCB IL) - 0.55 (1x MCB IL) + 5.26 (assumed host IL) = 19.71 dB
	Fitted insertion loss coefficients a0 a1 a2 a4	802.3-2015 85.8.4.2, FC-PI-5 9.7.3.3	0 1.7 0.8 0.02	0 3.2 1.4 0.04	dB (dB/GHz) ^{0.5} dB/GHz (dB/GHz) ²	
ILD	Insertion loss deviation		$\pm(0.5 + 0.1*f)$	$\pm(0.5 + 0.17*f)$	dB, f in GHz	Allows deviation at 12.89 GHz of ± 1.9 dB for Test 1, and ± 2.6 dB for Test 2
	Applied RJP-p	802.3-2015 85.8.4.2, FC-PI-5 9.7.1	0.15	0.15	UI	
	Applied DCDP-p	802.3-2015 85.8.4.2, FC-PI-5 9.7.1	0.035	0.035	UI	
	Applied TJ	802.3-2015 85.8.4.2, FC-PI-5 9.7.1	0.25	0.25	UI	
	Calibrated far end crosstalk (RMS)	802.3-2015 85.8.4.2	6.3	4.1	mV	
	Calibrated ICN (includes FEXT) (RMS)	802.3-2015 85.8.4.2, FC-PI-5	10	5.5	mV	

6.7.9.2 RECEIVER TESTS FOR USE WITH LIMITING CABLES

Receivers for EDR shall meet the required bit error ratio when tested using a PRBS31 pattern and a stressed input signal as specified in [Table 68](#) which just meets the inner eye amplitude. These signals are generated using the test system shown in [Figure 198](#) on page 616 and set up as described in IEEE 802.3bm-2015 Clause 83E, and in the CIWG ATD MOI.

Table 68 EDR receiver stressed input signal specifications - limiting cables

Symbol	Parameter	Test Value	Unit	Conditions
X	eye mask parameter, time	0.30	UI	Hit ratio=5E-5 with 100 Ohm load at TP7a
Y1, Y2	Differential unsigned input voltage tolerance range 0 (required) range 1 (optional) range 2 (optional)	50, 225 100, 350 150, 450	mV	Note: These specifications emulate limiting active cable range 0 output
tr	Transition time	13	ps	20-80%, Measured at this PRBS9 test pattern transition 11111111 <u>1</u> 00000111101...
J2	J2 jitter	0.44	UI	At TP7a
J9	J9 jitter	0.69	UI	At TP7a
	Crosstalk signal Vpk-pk	450	mV	At TP7a. Co-propagating aggressor. ^a
	Crosstalk signal transition time, 20%-80%	12	ps	Transition time measured at this PRBS9 test pattern transition 11111111 <u>1</u> 00000111101...
	Crosstalk calibration signal Vpk-pk	700	mV	At TP6a Counter-propagating aggressor, apply during crosstalk calibration only
	Crosstalk calibration signal transition time, 20%-80%	14	ps	

a. Host input crosstalk tolerance specifications are based on the requirement that EDR limiting active cables support range 0 differential unsigned output voltage. This host input spec shall be considered a minimum requirement for hosts; hosts may optionally choose to support a wider input range for EDR limiting active cables.

6.8 COMPLIANT CHANNELS

The signal characteristics seen by an InfiniBand receiver are defined by the combination of the Transmitter characteristics as defined in [6.6 Differential Driver Outputs on page 282](#) and any Compliant Channel as defined in this section. It is intended that the channel will be capable of being equalized by the receiver using equalizers such as a 5-tap Decision Feedback Equalizer (DFE).

This section outlines the channel topologies accommodated within the InfiniBand specification. On-board and backplane topologies are obsolete. Information on those topologies is contained in InfiniBand specification volume 2, rev. 1.2.1.

Integrated Circuits (ICs) that are depicted represent elements that generate and accept a specified InfiniBand interface. These ICs may provide any one of the following types of functions:

- Host Channel Adapter (HCA)
- Target Channel Adapter (TCA)
- Switch
- Repeater (non-addressable retimer and redrive component)
- Electrical/Optical Translation function
- Pluggable Device

6.8.1 DC BLOCKING

o6-1.2.1: The use of DC blocking capacitors is optional for backplane connections at SDR rate.

Explanation: The specified levels and termination allow the direct connection of driver to receiver if desired in environments with a common signal ground reference. Direct connection will minimize cost and maximize signal integrity. At higher speeds, DC blocking capacitors may be required.

C6-2: Cable ports shall incorporate DC blocking capacitors located at the receiver in the case of the microGigaCN interface, and in either the transmitter or the receiver end of the cable in the case of passive cables with the QSFP+ or CXP interface. Active cables with the QSFP+ or CXP interface shall present a high DC common-mode impedance on the high-speed signal inputs and outputs.

o6-2.2.1: DC blocking capacitors may be located at any IB pin (or pin pair).

Any loss or jitter caused by the addition of capacitors must be accounted for as part of the allocation for the printed wiring board on which the capacitors are mounted.

C6-2.2.1: DC blocking capacitors shall not be mounted inside the cable assembly for cables using the microGigaCN interface.

Capacitor value selection is left to the designer. Capacitors used in links for QDR and slower data rates shall be large enough to provide a -3dB point below 100 MHz. Capacitors used in linear cables for FDR and EDR shall be large enough to result in a cutoff frequency of 50 kHz or lower; 100 nF is recommended. Note that impedance and voltage values and measurements below the cutoff frequency will be affected by the presence of blocking capacitors.

Capacitor DC working voltage is left to the designer. Since the differences in common mode voltage are limited by the power supply voltage and the grounding of the shields, a value of 10 Volts should be more than adequate. ESD resistance should be taken into account, since the Cable and Backplane connector pins also need to meet the ESD specifications.

Layout of the mounting pads and vias for the blocking capacitor, and selection of the capacitor value and model must be carefully done in order to minimize impedance discontinuities. At higher speeds, blind or back drilled vias may be required.

6.8.2 LINEAR CABLE ELECTRICAL REQUIREMENTS

This section defines linear cable requirements for all data rates.

C6-3: Linear cable assemblies to be used for InfiniBand shall meet the electrical requirements listed in [Table 69](#) for all link widths.

Table 69 Cable Assembly Electrical Requirements

Symbol	Parameter	Maximum	Minimum	Unit	Conditions/Comments
$Z_{dca}(\text{nom})$	Differential Impedance, nominal value	110	95	Ω	linear and far end half active only, measured per EIA-364-108 over the length of each signal pair, from the unequalized end (if equalizer is used). Rise/Fall time: 100 ps (20-80%). Includes connector cable to connector interface and board termination pads and vias.
$Z_{dca}(\text{peak})$	Differential Impedance, deviation from nominal value	5		Ω	measured over the length of each signal pair

Architecture Note

Within pair skew may cause excessive jitter or induce sufficient shield current to exceed EMI radiation limits. It is expected that shield currents over 1 μ A may be sufficient to cause radiation problems.

For QDR and lower data rates, it is recommended that within pair skew be no more than 0.3 UI total for any given pair in a cable assembly. The effect of within pair skew will be reflected in the eye parameter measurements and bounded by those specifications. At FDR and higher data rates, the within pair skew is bounded by the ICMCN specification.

6.8.2.1 CABLE EQUALIZATION

Equalization in the cable, either fabricated as part of the bulk wire or in the form of discrete components in the cable assembly (e.g., on a printed circuit board inside the cable connector backshell), is permitted. However, the cable assembly shall be required to meet the electrical requirements as listed in the appropriate section of this specification for the data rate used in all cases.

6.8.3 LANE-TO-LANE SKEW

C6-4: [Table 70](#) defines the allowable lane-to-lane skew across the physical lanes. All IB ports shall limit skew to the values in [Table 70](#).

Table 70 Lane-to-lane skew maximum values

Skew Parameter	Symbol	SDR max.	DDR max.	QDR max.	FDR max.	EDR max.	Comment
Total Skew Note ^a	S _{RBtB}	24 ns (60 UI)	12 ns (60 UI)	6 ns (60 UI)	9.387 ns (132 UI)	5.120 ns (132 UI)	
Host transmitter Note ^b	S _{DBtB}	500 ps	500 ps	500 ps	500 ps	500 ps	Transmitter plus host board, up to and including I/O connector
Host receiver Note ^c	S _{RBtB}	500 ps	500 ps	500 ps	500 ps	500 ps	Receiver plus host board, up to and including I/O connector
Cable Assembly Note ^{d, e}	S _{CBtB}	23.0 ns	11.0 ns	5.0 ns	8.387 ns	4.120 ns	Cable assembly including module plugs

a. Defined in [Table 59 Host Receiver Characteristics for 2.5 Gb/s \(SDR\) on page 311](#) and included for reference.

b. Defined in [Table 48 Host Driver Characteristics for 2.5 Gb/s \(SDR\) on page 285](#) and included for reference.

c. Defined in [Table 59 Host Receiver Characteristics for 2.5 Gb/s \(SDR\) on page 311](#) and included for reference.

d. Defined in [Chapter 7: Electrical Connectors for Modules and Cables on page 351](#).

e. Retiming repeaters may independently insert or delete skip symbols on a per lane basis.

Fiber optic cables using detachable passive fiber shall meet the skew limits in [Table 71](#).

Table 71 Lane-to-lane skew maximum values for separable fiber optic cables

Skew Parameter	Symbol	SDR max.	DDR max.	QDR max.	FDR max.	EDR max.	Comment
Optical transceiver ^a		500 ps	250 ps	125 ps	125 ps	125 ps	Per transceiver, if separable from optical cable
Passive optical cable		22.0 ns	10.5 ns	4.75 ns	8.137 ns	3.87 ns	Fiber cable that is separable from transceiver

a. Defined in [Table 48 Host Driver Characteristics for 2.5 Gb/s \(SDR\) on page 285](#) and included for reference.

c. Defined in [Table 59 Host Receiver Characteristics for 2.5 Gb/s \(SDR\) on page 311](#) and included for reference.

d. Defined in [Chapter 7: Electrical Connectors for Modules and Cables on page 351](#).

e. Retiming repeaters may independently insert or delete skip symbols on a per lane basis.

At QDR and slower speeds, skew at any point is measured using the zero crossings of the differential voltage of the commas present in the training sequences TS1 and TS2 or in the Skip ordered-set as defined in [Section 5.5.2, “Control Ordered-sets 8b/10b encoding,” on page 142](#). Each of these sequences transmits a rank of commas on all physical lanes simultaneously.

At FDR and higher speeds, the appropriate test mode in Phy Test shall be used to generate TS1s to measure lane-to-lane skew. See [Section 5.17, “Physical Layer Compliance Testing,” on page 256](#).

6.8.4 SDR (2.5 GB/s) COMPLIANT CHANNEL

6.8.4.1 BACKGROUND

Allowable interconnect is specified in terms of attenuation at various package boundaries and at two or more frequencies. This bounds the acceptable configurations but does not guarantee operation. Both the driver and the interconnect are responsible for producing the specified signal at the receiver pins.

6.8.4.2 LOSS VALUES

C6-4.2.1: [Table 72](#) defines the maximum values of the loss assumed in IB interconnect topologies as shown in [Figure 95 on page 327](#). The following sections define each of the parameters. InfiniBand components shall limit loss at each package level to provide the eye openings and amplitudes at the Compliance Test Points as shown in [Table 46 InfiniBand Signal Test Points on page 274](#), when operating at 2.5 Gb/s (SDR).

Table 72 SDR Loss Parameter Maximum Values

Loss Parameter	Symbol	dB @ 2.5Gb/s, 1.25 GHz	dB at 625 MHz
Total Loss	L_T	15 dB	8 dB
Adapter Board ¹	L_A	2.5 dB	1.8 dB
Cable Assembly ²	L_C	See Chapter 7: Electrical Connectors for Modules and Cables on page 351	
I/O Plate	L_S	1 dB	0.6 dB
Active Backplane	L_{BA}	9.5 dB	6.5 dB
Passive Backplane	L_{BP}	7 dB	5 dB
Crosstalk	L_{XT}	3 dB	N/A

¹ includes a loss based on 1 pf via at the connector. Implementations must account for larger vias, if used.

² Cable attenuation is measured from the connecting vias on the IB board. Allowable values are defined in [Chapter 7: Electrical Connectors for Modules and Cables on page 351](#).

3. Loss at 625 MHz specified for predictable equalization. Attenuation is assumed to be smoothly varying with frequency.

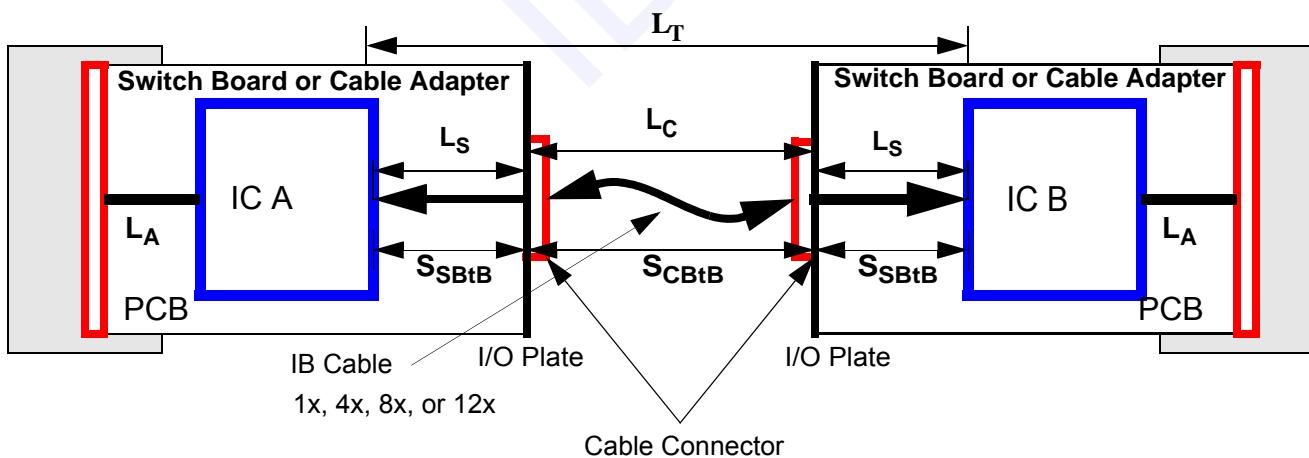


Figure 95 I/O Plate to I/O Plate via Cable Topology

Backplanes (or “System Boards”) may be connected together using cables as shown in [Figure 74 on page 267](#).

6.8.4.3 OPTICAL FIBER

This topology may be used for driving remote IB ports at distances of up to several hundred meters (or several km using the long haul fiber and laser transmitters) through optical interfaces. Each end of the link may be mounted on pluggable boards with each housing optical transceiver components as shown in [Figure 74 on page 267](#) or be directly mounted on a system board as shown in [Figure 75 on page 268](#). This release of the InfiniBand specification specifies 1X and 4X short and long haul links, and 1X, 4X, 8X, and 12X short haul optical links, at several speeds. Backplanes may be connected together using optical links. The transceivers depicted translate the Infiniband™ copper interface into an optical interface to drive the optical componenrty.

6.8.4.4 JITTER AND EYE OPENING

Eye opening values for SDR channels at the various test points are listed in [Table 73](#) and [Table 74 on page 329](#). These signal levels and eye openings are derived from the already defined driver, receiver, and interconnect characteristics. All values are for an equivalent 100 Ohm differential impedance load located at the test point. Correction or de-embedding must be performed to derive actual measured data, taking into account the configuration of the test setup.

Table 73 InfiniBand SDR Signal eye opening

Test Point	Description	SDR Unsigned Waveform Amplitude Notes 1,2	SDR Eye Width
TP1	Transmitted signal at board side of backplane connector	0.85	250 ps
TP2	Transmitted signal at backplane side of backplane connector	0.8	240 ps
TP3	Received signal at board side of connector	0.316	150 ps
TP4	Received signal at backplane side of connector	0.335	160 ps
TP5	Transmitted signal at board side of cable connector	0.96	250 ps
TP6	Transmitted signal at cable side of cable connector	0.89	240 ps
TP7	Received signal at board side of cable connector	0.282	150 ps
TP8	Received signal at cable side of cable connector	0.30	160 ps
TP9	Transmitted signal at board side of pluggable interface socket	0.89	296 ps
TP10	Received signal at board side of pluggable interface socket	0.282	126 ps

Note: All measurements for only LUT active, 2.5 Gb/s (SDR)

1. Refer to IEEE Std.181-2003 (Amplitude, Waveform, Unsigned) for definitions and procedures around unsigned waveform amplitude measurements.

2. Waveform amplitude measurements shall be conducted only on consecutive transition bits, over a 2UI Epoch (IEEE Std. 181-2003) unit interval, over a minimum population of 10E4 UI.

Table 74 SDR Cable Assembly Electrical Requirements

Symbol	Parameter	Minimum	Maximum	Unit	Conditions/Comments
J_{ca}	Jitter, SDR		0.25	UI	per EIA-364-107 and IEEE 191-2003, with 1 V differential Fibre Channel CJTPAT stimulus until 10000 hits in max. 20 mV high jitter box or equivalent, with equipment and fixture contribution de-embedded; worst case pair, with three adjacent pairs on one side of that pair (if they exist) to be driven at the same end of the cable as the measurement is performed, by an asynchronous source with 1.5 V unsigned differential amplitude and transition time of 100 ps or less at the board pins. The far end of each adjacent pair should be terminated in 50 Ohms to Ground. A PRBS generator or other source may be used for driving the adjacent pairs, at a minimum signaling rate of $f_{Nyquist}$. See Figure 96 .
V_{Cout}	Eye height (voltage), SDR	196		mV	minimum unsigned differential amplitude for 1 V differential Fibre Channel CJTPAT stimulus until 10000 hits in max. 20 mV high jitter box or equivalent, measured at board connector pins under the same conditions as J_{ca} (see above). See Figure 96 . For information, 316 mV if measured without crosstalk.
T_{Reye}	Eye width (time), SDR	0.75		UI	minimum time opening for 1 V differential Fibre Channel CJTPAT stimulus until 10000 hits in max. 20 mV high jitter box or equivalent, measured at board connector pins under the same conditions as J_{ca} (see above). See Figure 96 .

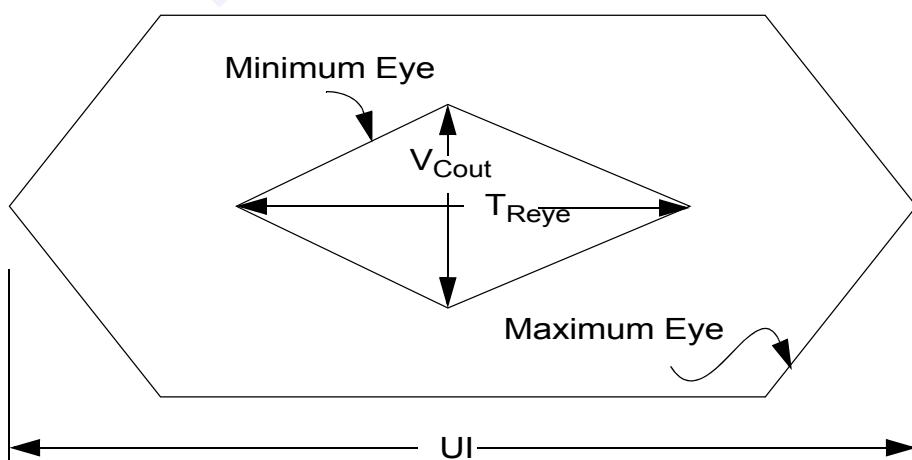


Figure 96 Eye Opening at receiving board connector pins, SDR rate (differential)

6.8.4.5 TOTAL LOSS (LT)

Total loss is measured between the pins of communicating IC packages or between the receiving IC and the passive equalizer at the transmitter.

6.8.4.6 ADAPTER BOARD LOSS (L_A)

The maximum loss from the IC package to the board connector is 2.5 dB at 1250 MHz, including the connector.

6.8.4.7 CABLE ASSEMBLY Loss (L_C)

The maximum loss in a cable assembly, including the connector loss, is 10 dB at 1250 MHz (value given here for reference only) as defined in [Chapter 7: Electrical Connectors for Modules and Cables on page 351](#).

6.8.4.8 I/O PLATE Loss (L_S)

The maximum loss L_S from the IC package to the I/O plate connector shall be 1 dB at 1250 MHz not including the connector.

6.8.4.9 PLUGGABLE DEVICE LOSS

The maximum loss from the IC package to the Pluggable Device Connector shall be 1 dB at 1250 MHz not including the connector.

6.8.4.10 CROSSTALK

80 mV referenced to the receiver of the loss budget is reserved for worst case crosstalk and other noise sources. Crosstalk is specified for the Backplane Connector in *InfiniBand Architecture Specification volume 2, version 1.2.1*, and for the cable connectors in [Table 93. "MicroGigaCN Cable connector electrical performance requirements." on page 365](#).

6.8.5 DDR (5.0 GB/s) LINEAR CHANNELS AND LIMITING ACTIVE CABLES

The maximum loss L_s from the IC package to the I/O plate connector shall be 4 dB at 2.5 GHz not including the connector.

6.8.5.1 DDR (5.0 GB/s) LINEAR CHANNEL

A Compliant linear channel for DDR (5.0 Gb/s) signaling is defined in terms of frequency-dependent S Parameters for the path between TP2 and TP4 in [Figure 79 on page 275](#) and between TP6 and TP8 in [Figure 80 on page 275](#). [Table 75](#) defines the values. S parameters for the full path from transmitter to receiver SERDES pins are listed in [Table 76 on page 331](#). These values are informative, rather than normative, since serdes pins are not normally accessible for testing.

Table 75 Linear Channel S Parameter Requirements for 5.0 Gb/s (DDR)

Frequency	S_{DD11} (max.) (dB) ^{a b}	S_{DD21} (min.) (dB) ^c
100 MHz	-10	-8
200 MHz	-10	-8
625 MHz	-10	-8
1250 MHz	-9.76	-8
1875 MHz	-9.26	-9.5
2500 MHz	-8.84	-11
4000 MHz	-8.0	
5000 MHz	-6.8	

a. Return loss measurements are not required on passive equalized cables on the signal pins at the cable end containing the equalizer components.

b. Note that in Rel. 1.3.1, specifications for S_{DD11} at these specific frequencies have been tightened somewhat, to match the S_{DD11} specifications for DDR limiting active cables defined in [Equation 11](#).

c. Parameter values are scaled for maximum allowable attenuation (-11 dB) at a frequency equal to half the bit rate (2500 MHz). S_{DD21} must decrease smoothly as frequency increases, with no notch-like behavior at frequencies up to the maximum frequency listed.

Table 76 SERDES to SERDES pin S Parameters for 5.0 Gb/s (Informative)

Frequency	S_{DD11} (max.) (dB)	S_{DD21} (min.) (dB) ^a
100 MHz	-10	-8.5
200 MHz	-10	-8.5
625 MHz	-10	-9.5
1250 MHz	-10	-13.5
1875 MHz	-10	-16.5
2500 MHz	-10	-19

a. Parameter values are scaled for maximum allowable attenuation (-19 dB) at a frequency equal to half the bit rate (2500 MHz). S_{DD21} must decrease smoothly as frequency increases, with no notch-like behavior at frequencies up to the maximum frequency listed.

6.8.5.2 DDR (5.0 GB/s) LIMITING ACTIVE CABLES

The electrical characteristics of active cables whose maximum data rate is DDR shall comply with the requirements listed in [Table 77](#) and [Table 78 on page 332](#). Parameters in this table are defined with respect to the diamond-shaped eye mask for QDR, DDR, and SDR speeds in [Figure 86 on page 284](#).

Table 77 DDR active cable input electrical specifications

Symbol	Parameter	Maximum	Nominal	Minimum	Unit	Conditions
X	Eye mask parameter, Time	0.15			UI	$+/-7\sigma$ (BER = 1E-12) measured at TP6
Y1,Y2	Eye mask parameters, voltage		250, 600		mV	
J _{D1}	Deterministic Jitter	0.15			UI	measured at TP6
J _{T1}	Total Jitter	0.30			UI	$+/-7\sigma$ (BER = 1E-12) measured at TP6
S _{DD11}	Differential input return loss	see Eq. 11 on page 333				measured at TP6

Table 78 DDR active cable output electrical specifications

Symbol	Parameter	Maximum	Nominal	Minimum	Unit	Conditions
X	Eye mask parameter, Time	0.36			UI	$+/-7\sigma$ (BER = 1E-12) measured at TP7
Y1,Y2	Eye mask parameters, voltage		100, 600		mV	
J _{D1}	Deterministic Jitter	0.40			UI	At TP7
J _{T1}	Total Jitter	0.72			UI	$+/-7\sigma$ (BER = 1E-12) measured at TP7
S _{DD22}	Differential output return loss	see Eq. 11 on page 333				measured at TP7

a. An FDR-capable active cable, if requested to by the host, may comply with the reduced Y1,Y2 amplitude specifications described in [Table 85, “FDR limiting active cable output electrical specifications,” on page 342](#) while operating at DDR data rate.

$$S_{DDxx} \leq \begin{cases} -10, 0.01 \leq f < 1 \text{ GHz} \\ -12 + 2\sqrt{f}, 1 \leq f < 4.1 \text{ GHz} \\ -6.3 + 13\log_{10}(f/5.5), 4.1 \leq f \leq 5 \text{ GHz} \end{cases} \quad \text{Eq.11}$$

f is frequency, in GHz

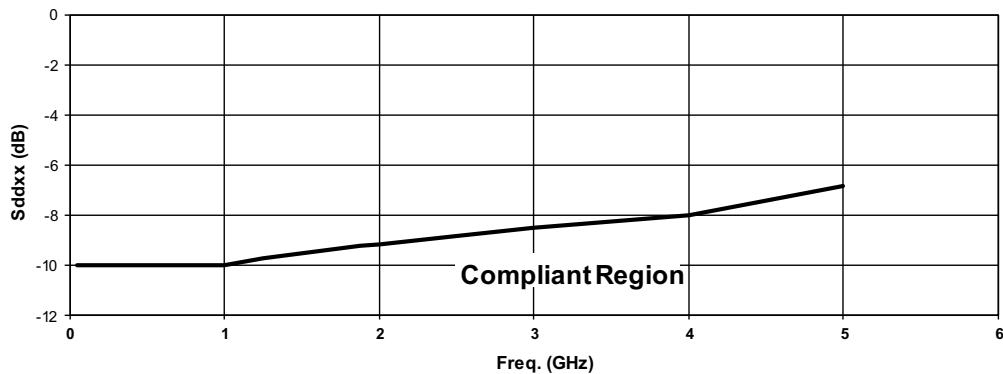


Figure 97 Limits on S_{DD11} and S_{DD22} vs. Frequency for DDR Active Cables

6.8.6 QDR (10.0 GB/S) LINEAR CHANNELS AND LIMITING ACTIVE CABLES

The maximum loss L_s from the IC package to the I/O plate connector shall be 6 dB at 5.0 GHz including the connector.

6.8.6.1 QDR (10.0 GB/S) LINEAR CHANNELS

A compliant linear channel for QDR (10.0 Gb/s) signaling is defined in terms of frequency-dependent S Parameters for the path between TP2 and TP4 in [Figure 79 on page 275](#) and between TP6 and TP8 in [Figure 80 on page 275](#). [Table 79](#) and [Table 80 on page 335](#) define the values. S parameters for the full path from transmitter to receiver SERDES pins are listed in [Table 80 on page 335](#). These values are informative rather than normative, since SERDES pins are not normally accessible for testing.

Table 79 Linear Channel S Parameter Requirements for 10 Gb/s (QDR)

Frequency	S_{DD11} (max.) (dB) ^a ^b	S_{DD21} (min.)(dB) ^c
100 MHz	-10	-8
200 MHz	-10	-8
625 MHz	-10	-8
1250 MHz	-9.76	-8
1875 MHz	-9.26	-8.5
2500 MHz	-8.84	-9.3
3750 MHz	-8.13	-11.2
5000 MHz	-6.84	-13
10000 MHz	-2.92	N/A

a. Return loss measurements are not required on passive equalized cables on the signal pins at the cable end containing the equalizer components.

b. Note that in Rel. 1.3.1, specifications for S_{DD11} at these specific frequencies have been slightly modified, to match the S_{DD11} specifications for QDR limiting active cables defined in [Equation 12](#).

c. Parameter values are scaled for maximum allowable attenuation (-13 dB) at a frequency equal to half the bit rate (5000 MHz). S_{DD21} must decrease smoothly as frequency increases, with no notch-like behavior, at frequencies up to the maximum frequency listed.

**Table 80 SERDES pin to SERDES pin S Parameters for 10 Gb/s (QDR)^a
(Informative)**

Frequency	S _{DD11} (max.)	S _{DD21} (min.) ^b
100 MHz	-10	-8.5
200 MHz	-10	-8.5
625 MHz	-10	-10.5
1250 MHz	-10	-12.5
1875 MHz	-10	-14.1
2500 MHz	-10	-16.4
3750 MHz	-10	-21
5000 MHz	-10	-25

a. All values are measured in dB.

b. Parameter values are scaled for maximum allowable attenuation (-25 dB) at a frequency equal to half the bit rate (5000 MHz). S_{DD21} must decrease smoothly as frequency increases, with no notch-like behavior, at frequencies up to the maximum frequency listed.**6.8.6.2 QDR (10.0 GB/s) LIMITING ACTIVE CABLES**

The electrical characteristics of active cables whose maximum data rate is QDR shall comply with the requirements listed in [Table 81](#) and [Table 82](#). Parameters in this table are defined with respect to the diamond-shaped eye mask for QDR, DDR, and SDR speeds in [Figure 86 on page 284](#).

Table 81 QDR active cable input electrical specifications

Symbol	Parameter	Maximum	Nominal	Minimum	Unit	Conditions
X	Eye mask parameter, Time	0.15			UI	+/-7σ (BER = 1E-12) measured at TP6
Y1,Y2	Eye mask parameters, voltage		250, 600		mV	
J _{D1}	Deterministic Jitter	0.15			UI	measured at TP6
J _{T1}	Total Jitter	0.30			UI	+/-7σ, BER = 1E-12, measured at TP6
S _{DD11}	Differential input return loss	Eq. 12 on page 336				measured at TP6

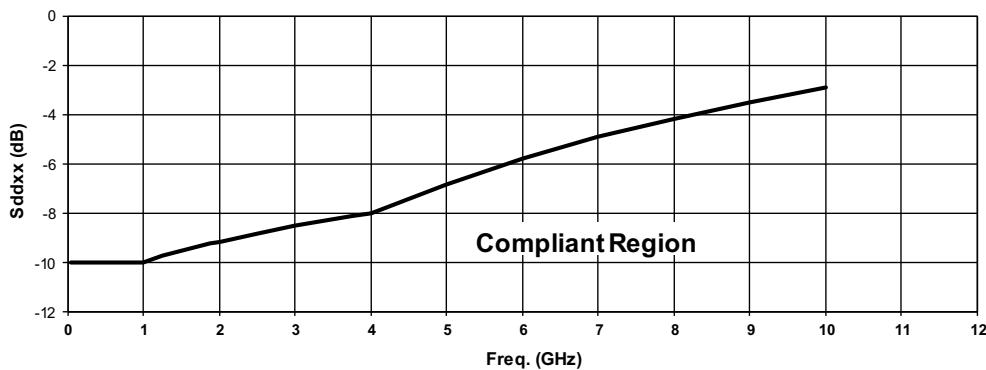
Table 82 QDR active cable output electrical specifications

Symbol	Parameter	Maximum	Nominal	Minimum	Unit	Conditions
X	Eye mask parameter, Time	0.36	100, 600		UI mV	+/-7σ (BER = 1E-12) measured at TP7
Y1,Y2	Eye mask parameters, voltage					Note: ^a
J _{D1}	Deterministic Jitter	0.40			UI	At TP7
J _{T1}	Total Jitter	0.72			UI	+/-7σ (BER = 1E-12) measured at TP7
S _{DD22}	Differential output return loss	Eq. 12 on page 336				measured at TP7

a. An FDR-capable active cable, if requested to by the host, may comply with the reduced Y1,Y2 amplitude specifications described in [Table 85, “FDR limiting active cable output electrical specifications,” on page 342](#) while operating at QDR data rate.

$$S_{DDxx} \leq \begin{cases} -10, 0.01 \leq f < 1\text{GHz} \\ -12 + 2\sqrt{f}, 1 \leq f < 4.1\text{GHz} \\ -6.3 + 13\log_{10}(f/5.5), 4.1 \leq f \leq 10\text{GHz} \end{cases} \quad \text{Eq.12}$$

f is frequency, in GHz

**Figure 98 Limits on S_{DD11} and S_{DD22} vs. Frequency for QDR Active Cables**

6.8.7 FDR (14.0625 GB/s) LINEAR CHANNELS AND LIMITING ACTIVE CABLES

Unless otherwise explicitly stated, for all FDR time-domain specifications (including eye, jitter, DDPWS, transition time, etc.), the waveform concerned is observed through a 17 GHz low-pass filter response (such as a Bessel-Thomson response), and includes the effects of the HCB and/or MCB as appropriate.

The maximum loss L_s from the IC package to the I/O plate connector is unspecified, but the link model assumes a loss budget of 5.26 dB at each end, with maximum total loss between IC packages of 24.42 dB at 7.03125 GHz, leaving 15.0 dB for the path between TP5a and TP7a, as illustrated in [Figure 211 FDR Overall Link Budget \(Informative\)](#).

6.8.7.1 FDR (14.0625 GB/s) LINEAR CHANNELS

A compliant linear channel for FDR (14.0625 Gb/s) signaling is defined in terms of frequency-dependent S Parameters for the path between TP2 and TP4 in [Figure 79 on page 275](#) and between TP6 and TP8 in [Figure 80 on page 275](#).

Linear channels for FDR signaling shall meet the fitted insertion loss as defined in [Eq. 13](#). The methodology for fitting the insertion loss is described in [Annex A2: Cable Electrical parameters for FDR and EDR on page 630](#).

$$IL_{fitted}(f) = a_0 + a_1 \sqrt{\frac{f}{f_b}} + a_2 \frac{f}{f_b} + a_4 \left(\frac{f}{f_b}\right)^2 \quad \text{Eq.13}$$

f is frequency, in GHz

where f_b is the signaling frequency (14.0625 GHz). Frequency-dependent channel insertion loss shall match the fitted insertion loss to within the Insertion Loss Deviation (ILD) limits specified in [Eq. 14](#), and the RMS ILD across frequencies shall match the limit described in [Table 83](#).

Linear cables for FDR signaling shall meet the integrated crosstalk noise limits as defined in [Table 83 on page 338](#). The methodology for calculating the integrated crosstalk noise ICN is described in [A2.2 Integrated Crosstalk Noise \(ICN\) on page 631](#). The methodology for calculating the integrated common mode conversion noise ICMCN is described in [A2.3 Integrated Common Mode Conversion Noise \(ICMCN\) on page 633](#).

Cable electrical outputs shall be AC coupled; i.e. they shall present a high DC common-mode impedance at TP5a. There may be various methods for AC coupling in actual implementations. Linear cables for use at 14 Gb/s FDR speed shall conform to the requirements listed in [Table 83](#).

Table 83 FDR compliant linear cable specifications

Symbol	Parameter	Paragraph reference	Max	Min	Unit	Comment
S _{DD21}	Insertion loss at 7.03125 GHz	IEEE 802.3-2015 clause 85.10.2	15.00		dB	TP5a to TP7a
	Fitted insertion loss coefficients a0 a1 a2 a4	OIF-CEI-03.0 clause 12.2	0.5 17.96 10.25 7.91	0 -0.75 0 0	dB dB dB dB	TP5a to TP7a; see Annex A2: Cable Electrical parameters for FDR and EDR on page 630
ILD _{ca}	Insertion loss deviation		Eq. 14		dB	TP5a to TP7a See description below
ILD _{rms}	Insertion loss deviation (RMS)		0.41 ^a Ref. Eq. 15		dB	TP5a to TP7a See description below
ICN	Integrated cross-talk noise (RMS)	Section A2.2 on page 631	Eq. 17 on page 340		mV	At TP7a, from 50 MHz to 14.1 GHz with crosstalk signals applied to both MCB TP5a ports See description below
S _{DD11} , S _{DD22}	Differential return loss	802.3-2015 clause 85.10.4	Eq. 1 on page 295		dB	At TP7a
S _{CC22}	Common mode return loss	FC-PI-5 clause 9.2.4	-2		dB	At TP7a, 200 MHz to 14.1 GHz
S _{DC11} , S _{DC22}	Common mode to differential reflection		Eq. 2 on page 295		dB	At TP5a, 50 MHz to 14.1 GHz
ICMCN	Integrated Common Mode Conversion Noise (RMS)		40		mV	At TP7a, with zero common mode input to the cable, with rise-time/falltime as specified in Table 55

a. Note that OIF-CEI-03.0 1st September 2011 has a formula for calculating ILD_{rms} which differs from [Eq. 15](#) by a weighting factor which is equivalent, for t_f=24 ps, f_b=14.0625 GHz, to a factor of 0.7355. This value here is equivalent to 0.3 dB in OIF-CEI-03.0.

The insertion loss deviation for the cable ILD_{ca} is the difference between the measured insertion loss IL and the fitted insertion loss and shall be between the upper and lower

limits defined by [Eq. 14](#).

$$|ILD_{ca}| \leq \begin{cases} 0.75, & 0.05 \leq f < 3.50 \\ 0.4286f - 0.75, & 3.50 \leq f < 7.0 \\ 2.25, & 7.0 \leq f \leq 10.5 \end{cases} \text{dB} \quad \text{Eq. 14}$$

f is frequency, in GHz

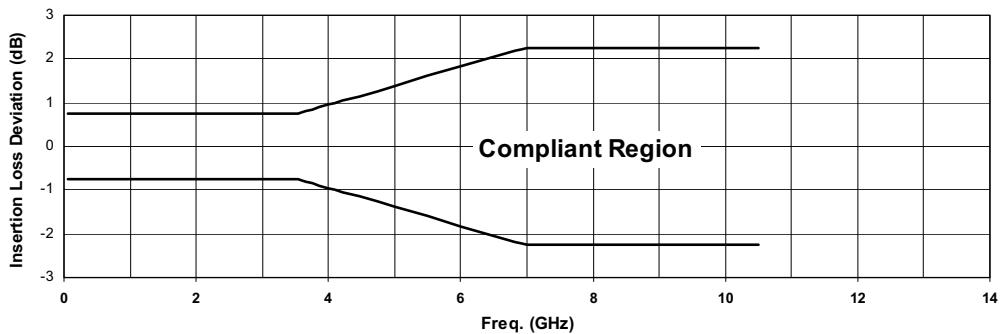


Figure 99 Insertion Loss Deviation (ILD) for FDR linear cables

ILD_{rms} is the RMS value of the ILD curve, and is calculated as indicated in [Eq. 15](#) and [Eq. 16](#) below. $ILD(f)$ is measured at N different frequencies in the frequency range from 50 MHz to 14.06 GHz, and the weighted root mean square of values is calculated using the formulae below. N must be high enough to capture the significant oscillations in the IL vs. frequency curve, so the frequency spacing should be 10 MHz or less. In the weighting function f_b is the bit rate of 14.0625 Gb/s, f_t is the equivalent frequency of the 20%-80% signal risetime $t_r = 24$ ps, calculated using $f_t = 0.2365 / t_r$, and f_r is the reference receiver bandwidth, which is defined as $(3/4)f_b = 10.546875$ GHz.

$$ILD_{rms} = \sqrt{\frac{\sum [W(f) \times (ILD^2(f))]}{\sum [W(f)]}} \quad \text{Eq. 15}$$

$$W(f) = \operatorname{sinc}^2(f/f_b) \left[\frac{1}{1 + (f/f_t)^4} \right] \left[\frac{1}{1 + (f/f_r)^8} \right] \quad \text{Eq. 16}$$

f is frequency, in GHz

The total Integrated Crosstalk Noise (ICN) shall meet the values specified in [Eq. 17](#), where IL is the cable assembly insertion loss in dB at 7.03125 GHz.

$$ICN \leq \begin{cases} 9, & 3 \leq IL \leq 7.65 \\ 12.75 - 0.49IL, & 7.65 < IL \leq 15.00 \end{cases} \text{ mV} \quad \text{Eq. 17}$$

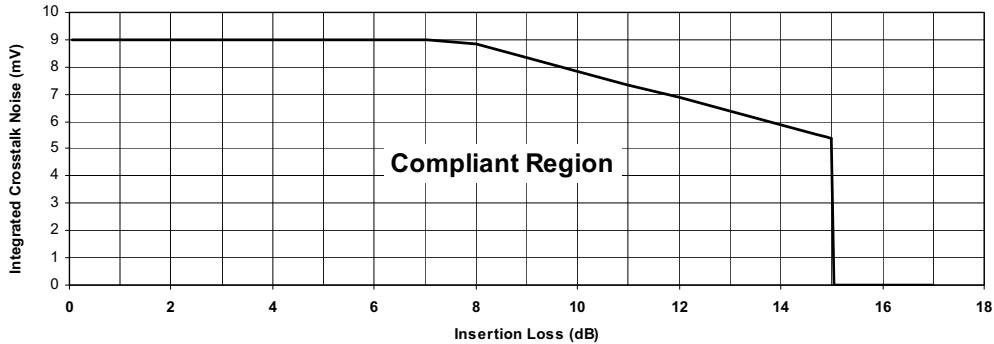


Figure 100 ICN vs. IL for FDR linear cables

6.8.7.1.1 MULTIPLE DISTURBER NEAR END CROSSTALK (MDNEXT)

The multiple disturber near end crosstalk, which is the crosstalk coupled into a receiver input from the adjacent transmit pairs when driven at the same end of the channel as the receiver, is defined in [Eq. 18](#), where N is the number of crosstalk aggressors (4 for a 4X interface, or 12 for a 12X interface) and $NEXT_i(f)$ is the induced crosstalk from each individual aggressor pair to the victim pair in question. These data are used in the ICN calculations described in [Annex 2.2 Integrated Crosstalk Noise \(ICN\) on page 631](#).

$$MDNEXT(f) = 10 \times \log_{10} \left(\sum_{i=1}^N 10^{NEXT_i(f)/10} \right) \quad \text{Eq. 18}$$

6.8.7.1.2 MULTIPLE DISTURBER FAR END CROSSTALK (MDFEXT)

The multiple disturber far end crosstalk, which is the crosstalk coupled into a receiver input from the transmitters adjacent to the channels' transmitter, is defined in [Equation 19](#), where N is the number of crosstalk aggressors (3 for a 4X interface, or 11 for a 12X interface) and $FEXT_i(f)$ is the induced crosstalk from each individual aggressor pair to the victim pair in question. These data are used in the ICN calculations described in [Annex 2.2 Integrated Crosstalk Noise \(ICN\) on page 631](#).

$$MDFEXT(f) = 10 \times \log_{10} \left(\sum_{i=1}^N 10^{FEXT_i(f)/10} \right) \quad \text{Eq. 19}$$

6.8.7.2 FDR (14.0625 GB/s) LIMITING ACTIVE CABLES**6.8.7.2.1 FDR LIMITING ACTIVE CABLE INPUT REQUIREMENTS**

Limiting Active cables for use at 14.0625 Gb/s (FDR) speed shall comply with the requirements listed in [Table 84](#) and [Figure 87 on page 284](#). The active cable electrical input shall be AC coupled; i.e. it shall present a high DC common-mode impedance at TP6a. There may be various methods for AC coupling in actual implementations. See [Figure 199 on page 617](#) for an example measurement setup.

Table 84 FDR limiting active cable input electrical specifications

Symbol	Parameter	Specification value(s)	Unit	Conditions
X1, X2	eye mask parameter, time; see Figure 87 on page 284	0.11, 0.31	UI	At TP6a, at FDR and higher data rates
Y1, Y2	eye mask parameter, voltage	95, 350	mV	Hit ratio=5x10 ⁻⁵
	Crosstalk signal Vpk-pk	+/- 5% (See Conditions)	mV	At TP6a. Co-propagating aggressors.
	Crosstalk signal transition time, 20%-80%	27	ps	Crosstalk signal Vpk-pk to match lane under test, to within +/- 5%. Transition time measured at this PRBS9 test pattern transition 11111111 1 00000111101...
	Crosstalk calibration signal Vpk-pk, each aggressor	450 +/- 10%	mV	At TP7a. Counter-propagating aggressors.
	Crosstalk calibration signal transition time, 20%-80%	27 +/- 3	ps	Apply during crosstalk calibration only ^a

Symbol	Parameter	Max	Min	Unit	Conditions
	Single-ended input voltage	4	-0.3	V	At TP6a
V _{CM}	AC common mode input voltage tolerance (RMS)	20		mV	At TP6a
DDPWS	Data Dependent Pulse Width Shrinkage	0.11		UI	At TP6a
J2	J2 Jitter tolerance	0.19		UI	At TP6a
J9	J9 Jitter tolerance	0.34		UI	At TP6a
S _{DD11}	Differential input return loss	Eq. 1 on page 295		dB	At TP5a, 50 MHz to 14.1 GHz
S _{DC11}	Common mode to differential reflection	Eq. 2 on page 295		dB	At TP5a, 50 MHz to 14.1 GHz

a. Please refer to CIWG Method of Implementation (MOI) document regarding Active Time Domain Testing for detailed specification of testing methodology and parameters.

6.8.7.2.2 FDR LIMITING ACTIVE CABLE OUTPUT REQUIREMENTS

Each electrical output lane and signal of the FDR active cable when measured at TP7a shall meet the specifications of [Table 85](#) while the signals on all input lanes at the other end of the cable comply with the specifications of [Table 84 on page 341](#) and the specified crosstalk signals are applied to all lanes of the active cable's electrical input at the output end of the cable. The active cable electrical output shall be AC coupled; i.e. it shall present a high DC common-mode impedance at TP7a. There may be various methods for AC coupling in actual implementations.

Table 85 FDR limiting active cable output electrical specifications

Symbol	Parameter	Specification value(s)	Unit	Conditions
X	eye mask parameter, time	0.30	UI	
Y1, Y2	Diff. unsigned output voltage range 0 (required) range 1 (optional) range 2 (optional)	50, 225 100, 350 150, 450	mV	Hit ratio=5E-5 with 100 Ohm load at TP7a (Note ^a)
	Crosstalk signal Vpk-pk, each aggressor	700 +/- 10%	mV	At TP6a. Counter-propagating aggressors. ^b
	Crosstalk signal transition time, 20%-80%	27 +/- 3	ps	Transition time measured at this PRBS9 test pattern transi- tion: 11111111 <u>0</u> 000011...

Symbol	Parameter	Max	Min	Unit	Conditions
Vout	Single-ended output voltage	4.0	-0.3	V	Referred to Signal Ground; measured at TP7a
V _{CM}	AC common mode output voltage (RMS)	20		mV	at TP7a
	Termination mismatch	5		%	1 MHz; at TP7a
S _{DD22}	Differential output return loss	Eq. 1 on page 295		dB	At TP7a, 50 MHz to 14.1 GHz
S _{CC22}	Common mode output return loss	-2		dB	At TP7a, 200 MHz to 14.1 GH
S _{DC22}	Common mode to differential reflec- tion	Eq. 2 on page 295		dB	At TP7a, 50 MHz to 14.1 GHz
t _r , t _f	Output transition time		13	ps	20-80%, Transition time mea- sured at these PRBS9 test pat- tern transitions: <u>111111110</u> 0000 <u>0</u> 11101...
J2	J2 Jitter	0.44		UI	At TP7a
J9	J9 jitter	0.69		UI	At TP7a

a. Output range is set for QSFP+ interfaces using page 03, addresses 238 & 239; see [Section 8.5](#).

For CXP interfaces, output range is set using Rx Addresses 62-67; see [Section 8.7.2](#).

b. Please refer to CIWG Method of Implementation (MOI) document Active Time Domain Testing for detailed specification of testing methodology and parameters.

Implementation Note - Output voltage ranges

The output voltage range 0 ($Y_1, Y_2 = 50, 225 \text{ mV}$) is defined to provide similar output swing as high-loss linear channels (long passive copper cables). Support of range 0 is required to reduce the dynamic range required at host device receiver circuits for systems with a mix of passive and active cables.

The default differential output voltage range for a limiting active cable that also operates at QDR and lower speeds must meet the eye mask defined in [Table 82 on page 336](#) and [Table 78 on page 332](#) for QDR and lower speeds (i.e., $Y_1, Y_2 = 100, 600 \text{ mV}, X=0.29 \text{ UI}$) to assure interoperability with host devices designed to meet Rel. 1.2.1 and earlier specifications.

An active cable may support both FDR voltage limits ($Y_1, Y_2 = 50, 225 \text{ mV}$) and QDR/DDR/SDR limits ($Y_1, Y_2 = 100, 600 \text{ mV}$) by either limiting its output voltage range to the intersection voltage range ($Y_1, Y_2 = 100, 225 \text{ mV}$), or by supporting output voltage range configuration as defined in Note 2 of [Table 85 on page 342](#).

6.8.8 EDR (25.78125 GB/s) LINEAR CHANNELS AND LIMITING ACTIVE CABLES

Unless otherwise explicitly stated, for all EDR time-domain specifications (including eye, jitter, DDPWS, transition time, etc.), the waveform concerned is observed through a 31 GHz low-pass filter response (such as a Bessel-Thomson response), and includes the effects of the HCB and/or MCB as appropriate.

The maximum loss L_s from the IC package to the I/O plate connector is unspecified, but the link model assumes a loss budget of 3.1 dB and 7.0 dB for the two ends, with maximum total loss between IC package BGA contacts of 24.5dB at 12.89 GHz, leaving budget of 14.4 dB for the cable and connectors. MCBs add 0.84 dB of trace loss per end, yielding 16.74 dB of maximum loss for the path between TP5a and TP7a, as illustrated in [Figure 212 EDR Overall Link Budget \(Informative\)](#).

6.8.8.1 EDR (25.78125 GB/s) LINEAR CHANNELS

A compliant linear channel for EDR (25.78125 Gb/s) signaling is defined in terms of frequency-dependent S Parameters for the path between TP2 and TP4 in [Figure 79 on page 275](#) and between TP6 and TP8 in [Figure 80 on page 275](#).

Linear channels for EDR signaling shall meet the fitted insertion loss as defined in [Eq. 20](#). The methodology for fitting the insertion loss is described in [Annex A2: Cable Electrical parameters for FDR and EDR on page 630](#).

$$IL_{fitted}(f) = a_0 + a_1 \sqrt{\frac{f}{f_b}} + a_2 \frac{f}{f_b} + a_4 \left(\frac{f}{f_b}\right)^2 \quad \text{Eq.20}$$

f is frequency, in GHz

where f_b is the signaling frequency (25.78125 GHz). Frequency-dependent channel insertion loss shall match the fitted insertion loss to within the Insertion Loss Deviation (ILD) limits specified in [Eq. 21](#), and the RMS ILD across frequencies shall match the limit described in [Table 86](#).

Linear cables for EDR signaling shall meet the integrated crosstalk noise limits as defined in [Table 86 on page 345](#). The methodology for calculating the integrated crosstalk noise ICN is described in [A2.2 Integrated Crosstalk Noise \(ICN\) on page 631](#). The methodology for calculating the integrated common mode conversion noise ICMCN is described in [A2.3 Integrated Common Mode Conversion Noise \(ICMCN\) on page 633](#).

Cable electrical outputs shall be AC coupled; i.e. they shall present a high DC common-mode impedance at TP5a. There may be various methods for AC coupling in actual implementations. Linear cables for use at 25 Gb/s EDR speed shall conform to the requirements listed in [Table 86](#).

Table 86 EDR compliant linear cable specifications

Symbol	Parameter	Paragraph reference	Max	Min	Unit	Comment
S _{DD21}	Insertion loss at 12.89 GHz	IEEE 802.3-2015 clause 85.10.2	16.74		dB	TP5a to TP7a
	Fitted insertion loss coefficients a0 a1 a2 a4	OIF-CEI-03.0 clause 12.2	0.5 17.96 10.25 7.91	0 -0.75 0 0	dB dB dB dB	TP5a to TP7a; see Annex A2: Cable Electrical parameters for FDR and EDR on page 630
ILD _{ca}	Insertion loss deviation		Eq. 21		dB	TP5a to TP7a See description below
ILD _{rms}	Insertion loss deviation (RMS)		0.41 ^a Ref. Eq. 22		dB	TP5a to TP7a See description below
ICN	Integrated cross-talk noise (RMS)	Section A2.2 on page 631	Eq. 24 on page 347		mV	At TP7a, from 50 MHz to 26 GHz with crosstalk signals applied to both MCB TP5a ports See description below
S _{DD11} , S _{DD22}	Differential return loss	802.3-2015 clause 85.10.4	Eq. 5 on page 304		dB	At TP7a
S _{CC22}	Common mode return loss	FC-PI-5 clause 9.2.4	-2		dB	At TP7a, 50 MHz to 26 GHz
S _{DC11} , S _{DC22}	Common mode to differential reflection		Eq. 6 on page 304		dB	At TP5a, 50 MHz to 26 GHz
ICMCN	Integrated Common Mode Conversion Noise (RMS)		60		mV	At TP7a, with zero common mode input to the cable, with risetime/fall-time as specified in Table 58

a. Note that OIF-CEI-03.0 1st September 2011 has a formula for calculating ILD_{rms} which differs from [Eq. 22](#) by a weighting factor which is equivalent, for t_f=24 ps, f_b=14.0625 GHz, to a factor of 0.7355. This value here is equivalent to 0.3 dB in OIF-CEI-03.0.

The insertion loss deviation for the cable ILD_{ca} is the difference between the measured insertion loss IL and the fitted insertion loss and shall have absolute value below the limits defined in [Eq. 21](#).

$$|ILD_{ca}| \leq \begin{cases} 0.75, & 0.05 \leq f < 5 \\ (0.26)(f-5) + 0.75, & 5 \leq f < 15.5 \\ 3.5, & 15.5 \leq f \leq 19.5 \end{cases} dB \quad Eq. 21$$

f is frequency, in GHz

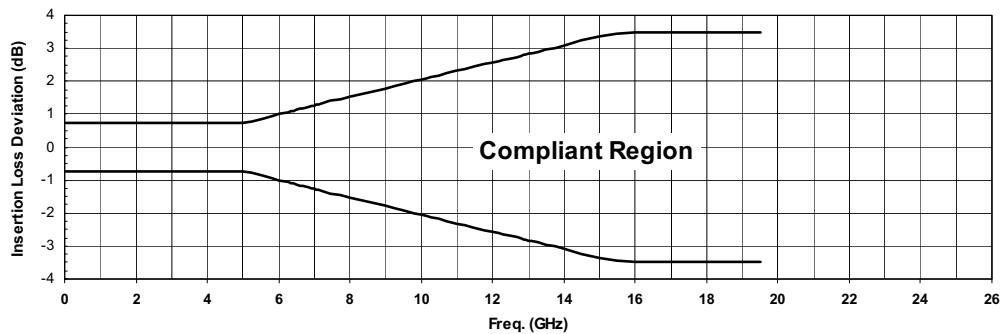


Figure 101 Insertion Loss Deviation (ILD) for EDR linear cables

ILD_{rms} is the RMS value of the ILD curve, and is calculated as indicated in of [Eq. 22](#) and [Eq. 23](#) below. ILD(f) is measured at *N* different frequencies in the frequency range from 50 MHz to 19.5 GHz and the weighted root mean square of values is calculated using the formulae below. *N* must be high enough to capture the significant oscillations in the IL vs. frequency curve, so the frequency spacing should be 10 MHz or less. In the weighting function *f_b* is the bit rate of 25.78125 Gb/s, *f_t* is the equivalent frequency of the 20%-80% signal risetime *t_r* = 13 ps, calculated using *f_t* = 0.2365 / *t_r*, and *f_r* is the reference receiver bandwidth, which is defined as (3/4)*f_b* = 19.3359375 GHz.

$$ILD_{rms} = \sqrt{\frac{\sum [W(f) \times (ILD^2(f))]}{\sum [W(f)]}} \quad Eq. 22$$

$$W(f) = \operatorname{sinc}^2(f/f_b) \left[\frac{1}{1 + (f/f_t)^4} \right] \left[\frac{1}{1 + (f/f_r)^8} \right] \quad Eq. 23$$

f is frequency, in GHz

The total Integrated Crosstalk Noise (ICN) shall meet the values specified in [Eq. 24](#), where IL is the cable assembly insertion loss in dB at 12.89 GHz.

$$ICN \leq \begin{cases} 9, & 3 \leq IL \leq 7.65 \\ 12.75 - 0.49IL, & 7.65 < IL \leq 26 \end{cases} \text{ mV} \quad \text{Eq.24}$$

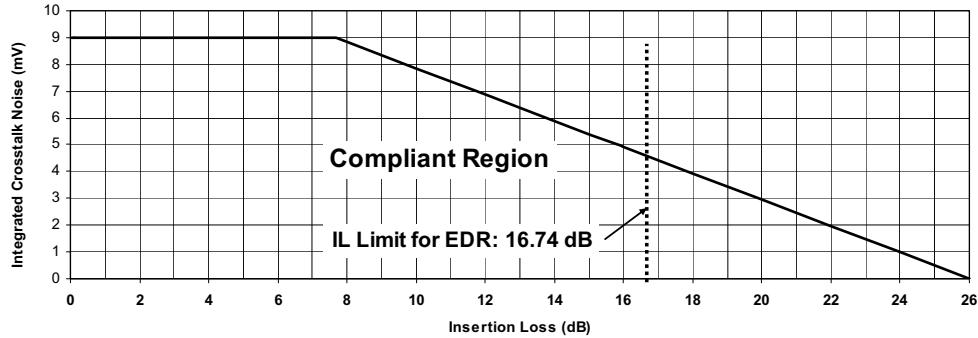


Figure 102 ICN vs. IL for EDR linear cables

6.8.8.1.1 MULTIPLE DISTURBER NEAR END CROSSTALK (MDNEXT)

The multiple disturber near end crosstalk, which is the crosstalk coupled into a receiver input from the adjacent transmit pairs when driven at the same end of the channel as the receiver, is defined in [Equation 25](#), where N is the number of crosstalk aggressors (4 for a 4X interface, or 12 for a 12X interface) and $NEXT_i(f)$ is the induced crosstalk from each individual aggressor pair to the victim pair in question. These data are used in the ICN calculations described in [Annex 2.2 Integrated Crosstalk Noise \(ICN\) on page 631](#).

$$MDNEXT(f) = 10 \times \log 10 \left(\sum_{i=1}^N 10^{NEXT_i(f)/10} \right) \quad \text{Eq.25}$$

6.8.8.1.2 MULTIPLE DISTURBER FAR END CROSSTALK (MDFEXT)

The multiple disturber far end crosstalk, which is the crosstalk coupled into a receiver input from the transmitters adjacent to the channels' transmitter, is defined in [Equation 26](#), where N is the number of crosstalk aggressors (3 for a 4X interface, or 11 for a 12X interface) and $FEXT_i(f)$ is the induced crosstalk from each individual aggressor pair to the victim pair in question. These data are used in the ICN calculations described in [Annex 2.2 Integrated Crosstalk Noise \(ICN\) on page 631](#).

$$MDFEXT(f) = 10 \times \log 10 \left(\sum_{i=1}^N 10^{FEXT_i(f)/10} \right) \quad \text{Eq.26}$$

6.8.8.2 EDR (25.78125 Gb/s) LIMITING ACTIVE CABLES**6.8.8.2.1 EDR LIMITING ACTIVE CABLE INPUT REQUIREMENTS**

Limiting Active cables for use at 25.78125 Gb/s (EDR) speed shall comply with the requirements listed in [Table 87](#) and [Figure 87 on page 284](#). The active cable electrical input shall be AC coupled; i.e. it shall present a high DC common-mode impedance at TP6a. There may be various methods for AC coupling in actual implementations. See [Figure 199 on page 617](#) for an example measurement setup.

Table 87 EDR limiting active cable input electrical specifications

Symbol	Parameter	Specification value(s)	Unit	Conditions
	Crosstalk signal Vpk-pk	+/- 5% (See Conditions)	mV	At TP6a. Co-propagating aggressors. Crosstalk signal Vpk-pk to match lane under test, to within +/- 20%.
	Crosstalk signal transition time, 20%-80%	17	ps	
	Crosstalk calibration signal Vpk-pk, each aggressor	450 +/- 10%	mV	At TP7a. Counter-propagating aggressors. Apply during crosstalk calibration only ^a
	Crosstalk calibration signal transition time, 20%-80%	17 +/- 3	ps	

Symbol	Parameter	Max	Min	Unit	Conditions
	Single-ended input voltage	3.3	-0.3	V	At TP6a
V_{CM-AC}	AC common mode input voltage tolerance (RMS)	20		mV	At TP6a
V_{CM-DC}	DC common mode input voltage tolerance	2850	-350	mV	At TP6a
S_{DD11}	Differential input return loss	Eq. 5 on page 304		dB	At TP5a, 50 MHz to 26 GHz
S_{DC11}	Common mode to differential reflection	Eq. 6 on page 304		dB	At TP5a, 50 MHz to 26 GHz
EH15	Eye Height tolerance, at 1E-15	120		mV	
EW15	Eye Width tolerance, at 1E-15	0.53		UI	At TP6a, with TX CDR enabled
		0.71		UI	At TP6a, with TX CDR bypassed (i.e., disabled)

a. Please refer to CIWG Method of Implementation (MOI) document Active Time Domain Testing for detailed specification of testing methodology and parameters.

6.8.8.2.2 EDR LIMITING ACTIVE CABLE OUTPUT REQUIREMENTS

Each electrical output lane and signal of the EDR active cable when measured at TP7a shall meet the specifications of [Table 88](#) while the signals on all input lanes at the other end of the cable comply with the specifications of [Table 87 on page 348](#) and the specified crosstalk signals are applied to all lanes of the active cable's electrical input at the output end of the cable. The active cable electrical output shall be AC coupled; i.e. it shall present a high DC common-mode impedance at TP7a. There may be various methods for AC coupling in actual implementations.

Table 88 EDR limiting active cable output electrical specifications

Symbol	Parameter	Specification value(s)	Unit	Conditions
X	eye mask parameter, time; see Figure 86 on page 284	0.30	UI	Hit ratio=5E-5 with 100 Ohm load at TP7a (Note ^a)
Y1, Y2	Diff. unsigned output voltage range 0 (required) range 1 (optional) range 2 (optional)	50, 225 100, 350 150, 450	mV	
	Crosstalk signal Vpk-pk, each aggressor	700 +/- 10%	mV	At TP6a. Counter-propagating aggressors. ^b
	Crosstalk signal transition time, 20%-80%	17 +/- 3	ps	Transition time measured at this PRBS9 test pattern transition: 111111110000011...

Symbol	Parameter	Max	Min	Unit	Conditions
Vout	Single-ended output voltage	4.0	-0.3	V	Referred to Signal Ground; measured at TP7a
V _{CM}	AC common mode output voltage	20		mV	(RMS); at TP7a
	Termination mismatch	5		%	1 MHz; at TP7a
S _{DD22}	Differential output return loss	Eq. 5 on page 304		dB	At TP7a, 50 MHz to 26 GHz
S _{CC22}	Common mode output return loss	-2		dB	At TP7a, 200 MHz to 26 GHz
S _{DC22}	Common mode to differential reflection	Eq. 6 on page 304		dB	At TP7a, 50 MHz to 26 GHz
t _r , t _f	Output transition time		10	ps	20-80%, Transition time measured at these PRBS9 test pattern transitions: 1111111100000111101...
J2	J2 Jitter	0.44		UI	At TP7a
J9	J9 jitter	0.69		UI	At TP7a

a. Output range is set for QSFP+ interfaces using page 03, addresses 238 & 239; see [Section 8.5](#).

For CXP interfaces, output range is set using Rx Addresses 62-67; see [Section 8.7.2](#).

b. Please refer to CIWG Method of Implementation (MOI) document Active Time Domain Testing for detailed specification of testing methodology and parameters.

Implementation Note - Output voltage ranges

The output voltage range 0 ($Y_1, Y_2 = 50, 225$ mV) is defined to provide similar output swing as high-loss linear channels (long passive copper cables). Support of range 0 is required to reduce the dynamic range required at host device receiver circuits for systems with a mix of passive and active cables.

The default differential output voltage range for a limiting active cable that also operates at QDR and lower speeds must meet the eye mask defined in [Table 82 on page 336](#) and [Table 78 on page 332](#) for QDR and lower speeds (i.e., $Y_1, Y_2 = 100, 600$ mV, $X=0.29$ UI) to assure interoperability with host devices designed to meet Rel. 1.2.1 and earlier specifications.

An active cable may support both EDR voltage limits ($Y_1, Y_2 = 50, 225$ mV) and QDR/DDR/SDR limits ($Y_1, Y_2 = 100, 600$ mV) by either limiting its output voltage range to the intersection voltage range ($Y_1, Y_2 = 100, 225$ mV), or by supporting output voltage range configuration as defined in Note a. of [Table 88](#).

Architecture Note - Active Limiting Cable Testing Methodology

Test methodology for EDR Active Limiting Cables - both AOCs (Active Optical Cables) and AECs (Active Electrical Cables) - will reference the procedures described in 802.3bm-2015 Clause 83E.3.4.1.1 "Module stressed input test procedure". This note describes similarities and differences vs. the Clause 83E procedure. The complete methodology is described in the CIWG Active Time Domain Testing MOI.

The EDR Active Cable test procedure will use a stressed input signal including a pattern generator signal, stressed by sinusoidal jitter, bounded uncorrelated jitter, and random jitter, followed by frequency-dependent attenuation, measured with a reference receiver. The sinusoidal jitter magnitude is set to 0.05 UI, at a frequency of 91 MHz (well above the scope's reference receiver CDR circuit), in accordance with a particular point on the curve described in 802.3 Table 88-13. The random jitter is set to 0.15 UI p-p at BER of 10^{-15} . The frequency-dependent attenuation is set to 10 dB (rather than the 13.8 dB value in Clause 83E.3.4.1.1) to match the lower attenuation typical for InfiniBand-compliant products. There is no low-loss test. Then, bounded uncorrelated jitter is adjusted so that the EW15 total jitter matches the limit described in [Table 58, "EDR host output specifications at Preset 0, for Limiting Active Cables," on page 308](#), when measured through the frequency-dependent attenuation, using the scope reference receiver with CTLE setting to maximize the product of eye height and eye width. Calibration is done using a PRBS9 data pattern, to allow full pattern capture. After calibration, DUT testing is done using a PRBS31 pattern.

As described in [Table 58, "EDR host output specifications at Preset 0, for Limiting Active Cables," on page 308](#) and [Table 87, "EDR limiting active cable input electrical specifications," on page 348](#), there are two mode of testing. The "Retimed" mode measurement is made with the Tx CDR (at the electrical input of the cable) enabled, and (optionally) with the Rx CDR enabled at cable output, as desired to optimize power utilization and output jitter. The "Unretimed" mode measurement is made with all cable/module CDRs bypassed, and places tighter constraints on the host output/cable input jitter specifications.

The output of the Limiting Active Cable is measured for BER (important for retimed mode) and measured against the limits in [Table 88, "EDR limiting active cable output electrical specifications," on page 349](#). At 25 Gb/s, measurement for 120 seconds (2 minutes) allows BER measurement across 3×10^{12} bits.

CHAPTER 7: ELECTRICAL CONNECTORS FOR MODULES AND CABLES

7.1 INTRODUCTION

The following sections define InfiniBand cables and the connectors used to attach InfiniBand cables or modules with electrical interfaces to InfiniBand host boards. InfiniBand cable connectors shall incorporate features shown in the appropriate sections below that specify the connector to board interfaces.

It is the responsibility of the cable and connector designers or suppliers to perform the indicated tests and supply the data to potential customer companies to indicate compliance. It is recommended that the appropriate test groups specified in EIA-364-1000.01 be used for qualification testing, using the following conditions:

- 50 mating cycles preconditioning
- Unmated exposure, option 2 mixed flowing gas exposure
- Five year product life
- Field operating temperature range up to 60 degrees C.

The requirements stated in this section apply to all connectors used to attach InfiniBand cables to InfiniBand host boards. The following individual sections that define the specific connectors to be used for a given link width may also define additional requirements for that specific connector.

Use of the keying definitions included herein is optional; however, utilization of these definitions is highly recommended to prevent plugging of InfiniBand cables into ports using incompatible interfaces such as Fibre Channel.

7.1.1 ENVIRONMENTAL PERFORMANCE REQUIREMENTS

C7-1: Connectors to be used on and for connection to InfiniBand host boards shall meet or exceed the environmental performance requirements of EIA-364-1000-2009, including exposure to Mixed Flowing Gas consistent with the required product life as defined in [Section 7.1](#).

Unless otherwise noted, successful completion of a given test is indicated by an acceptable Low Level Contact Resistance measurement upon completion of the test, as defined above.

7.1.2 PORT AND CONNECTOR LABELING

It is recommended that I/O ports be labeled to avoid confusion between InfiniBand ports and connections and those used for other incompatible interfaces.

The following icons have been chosen to identify InfiniBand connectors for both cables and chassis. There is one icon for each cable width and speed identified below. The use of the icon is optional but highly recommended.

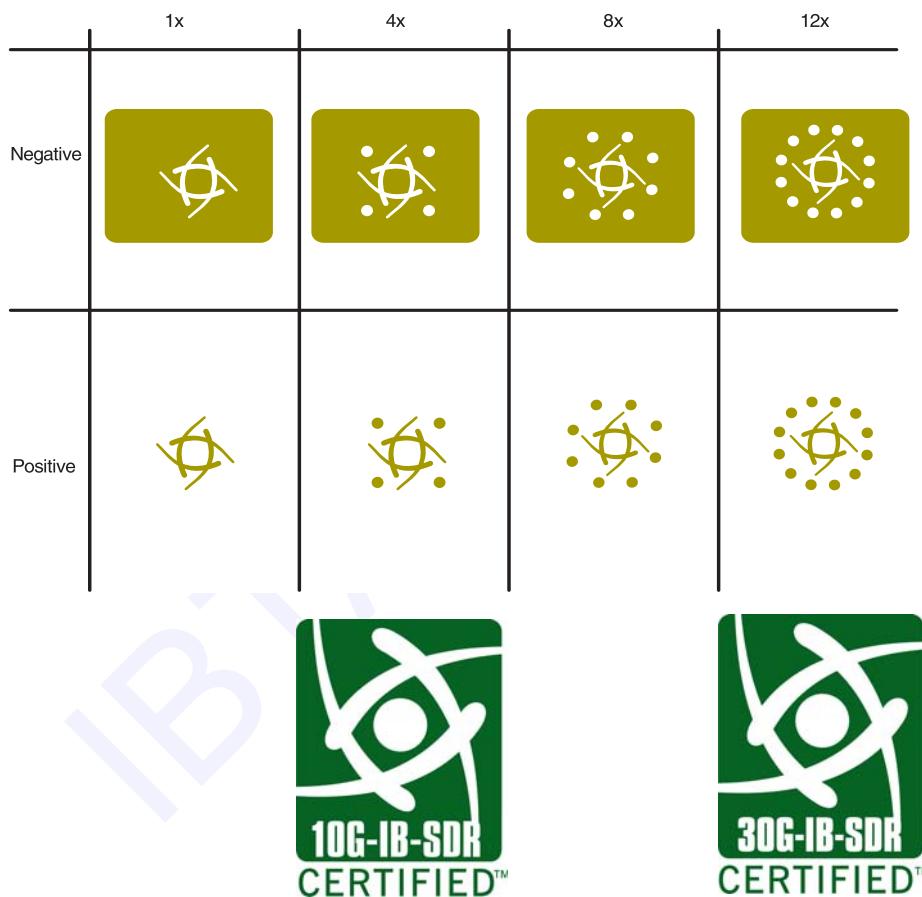


Figure 103 Icons for SDR cables and ports

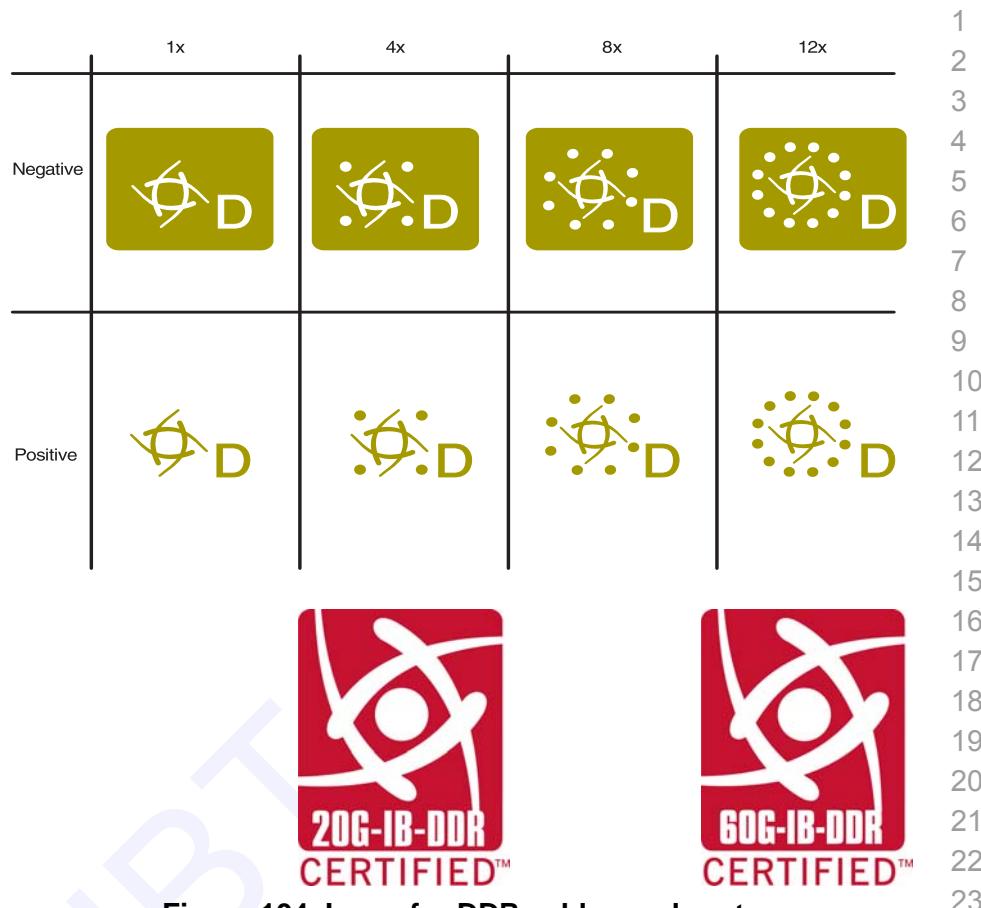


Figure 104 Icons for DDR cables and ports

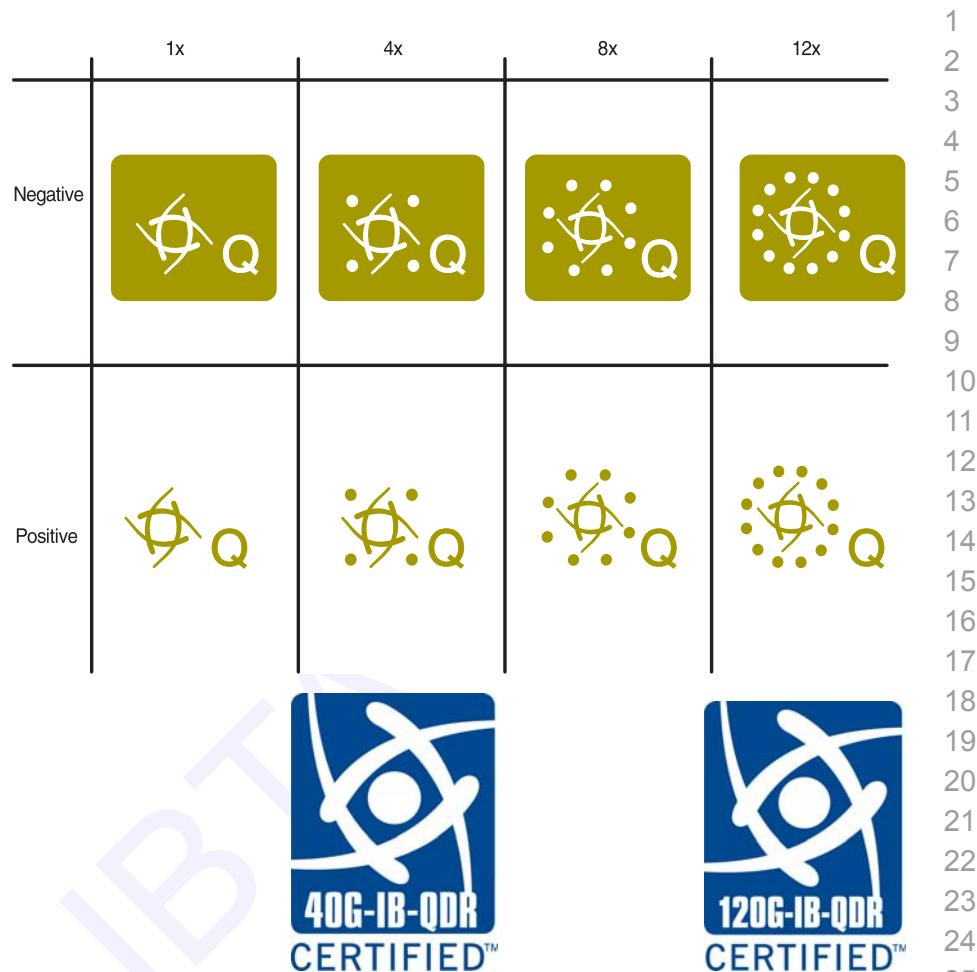


Figure 105 Icons for QDR cables and ports



Figure 106 Icons for 4x and 12x FDR cables and ports

07-1.1.1: If an icon is associated with an InfiniBand port, it shall be a scaled version of the icons in [Figure 103](#), [Figure 104](#), [Figure 105](#), or [Figure 107](#) that represents the in-

terface width and speed of the associated port. The icon shall not be less than 5 mm height and 5 mm in width.

The physical placement (top, bottom, side) and means (label, stamping or molding) as well as the location, relative to the icon, of the DDR and QDR designators (D and Q respectively) is left to the implementation.

o7-1.1.2: If an InfiniBand icon is used to identify a copper or optical port and color coding is used to further enhance identification, Pantone 2738 (Blue) or Pantone 399 C (Green) shall be used.

Recommendation: If the InfiniBand icon is not used for port association and a color scheme is used., it is recommended that Pantone 2738 (Blue) be used.

o7-1.1.3: If an icon is associated with an InfiniBand copper cable or optical fiber, it shall be a scaled version of the icon in [Figure 103](#), [Figure 104](#), or [Figure 105](#) that represents the interface width of the associated cable or fiber. The size and contrast of the icon shall allow it to be visible from a distance of 0.5 m under standard office lighting (500 lux).

Recommendation: If an icon is affixed, stamped or molded on an InfiniBand copper cable, it is recommended that it be placed in the space noted as "Icon Area" in InfiniBand Architecture Specification, [Figure 60: 1x cable plug](#) of Volume 2, Rel. 1.2.1 , [Figure 110 4X cable plug](#) or [Figure 133 12X cable plug](#) or copper cables or [Figure 53 on page 235](#) of InfiniBand Architecture Specification, [Volume 2, Rel. 1.2.1](#) for 1x optical pluggables.

For cables or pluggable transceiver components with pull tabs, the InfiniBand icons from [Section 7.1.2](#) of this specification is recommendd to be molded onto the pull-tabs. For example, a QSFP transceiver or cable capable of DDR speeds would have molded onto the pulltab (in a size that "shall allow it to be visible from a distance of 0.5m under standard office lighting (500 lux)" the patterns shown in [Figure 104](#).

7.2 1X INTERFACE

7.2.1 1X BOARD CONNECTOR

This interface is obsolete. Refer to Volume 2, Release 1.2.1 of the InfiniBand specification for information.

7.3 MICROGIGACN INTERFACE

7.3.1 MECHANICAL REQUIREMENTS

7.3.1.1 PHYSICAL AND MECHANICAL PERFORMANCE REQUIREMENTS

C7-2: Connectors to be used on and for connection to InfiniBand modules using the microGigaCN interface shall meet or exceed the physical and mechanical performance requirements listed in [Table 89](#).

Table 89 microGigaCN cable connector physical requirements

Parameter		Minimum	Maximum	Unit	Conditions/Comments
N	Durability	250		mating cycles	Without physical damage or exceeding low level contact resistance when mated; no more than 1% of contacts with exposed base metal
F_i	Insertion force	1X	30	N	
		4X	56		
		12X	73		
F_w	Withdrawal force	1X	30	N	
		4X	49		
		12X	59		
F_r	Retention force	75		N	Load pull, per EIA-364-38A
F_{ls}	Side load capability	75		N	No damage to cable or board, no opens detected with LLCR test; force applied to the cable in a plane parallel to the I/O plate at a distance of 90 mm, in the direction of the smaller dimension of the receptacle
F_{ll}	Longitudinal load capability	100		N	No damage to cable or board, no opens detected with LLCR test; force applied to the cable in a plane parallel to the I/O plate at a distance of 90 mm, in the direction of the larger dimension of the receptacle
F_{rc}	Housing contact retention force	5		N	
t_{pm}	Contact finish – option 1	0.76 Au over 1.27 Nickel		μm	
t_{pm}	Contact finish – option 2	0.51 PdNi with Au flash over 1.27 Nickel		μm	Min. 75% Pd in PdNi alloy

It is also recommended that connector interfaces meet the parameters defined in [Table 90](#). Unless otherwise noted, successful completion of a given test is indicated by an acceptable Low Level Contact Resistance measurement as indicated in EIA-364-1000-2009 or other equivalent test sequence.

Note that all drawing dimensions in the following sections are in millimeters (mm).

Table 90 Recommended microGigaCN interface connector physical requirements

Symbol	Parameter	Minimum	Maximum	Unit	Conditions/Comment
F_n	Contact normal force	100		cN	per contact beam
S_{hcc}	Contact Hertz stress	170		kpsi	per contact beam
D_{wc}	Contact wipe	1.0		mm	

7.3.2 4X INTRODUCTION

This section defines one connector for the 4X cable interface on InfiniBand boards. The 4X interface provides for simultaneous transmit and receive of four bits of differential data, and is primarily used at SDR and DDR data rates.

The host board connector is surface-mounted, and all signal pins are used. A suitable receptacle, referred to as an eight pair microGigaCN receptacle, is available from Fujitsu Components Ltd. and others, and is shown in [Figure 107](#).

C7-3: All 4X InfiniBand cable plugs using the microGigaCN cable connector shall be intermateable with the board connector shown in [Figure 107](#).

The cable plug to be used on passive InfiniBand 4X cables uses pairs of contacts interspersed with Ground contacts for crosstalk reduction. Eight pairs of signals are used, four each for transmit and receive. A suitable plug, referred to as an eight pair microGigaCN plug, is available from Fujitsu Components Component, Ltd. and others, and is shown in [Figure 110 on page 363](#). Detailed drawings of mating interface dimensions are shown in [Figure 111 on page 364](#).

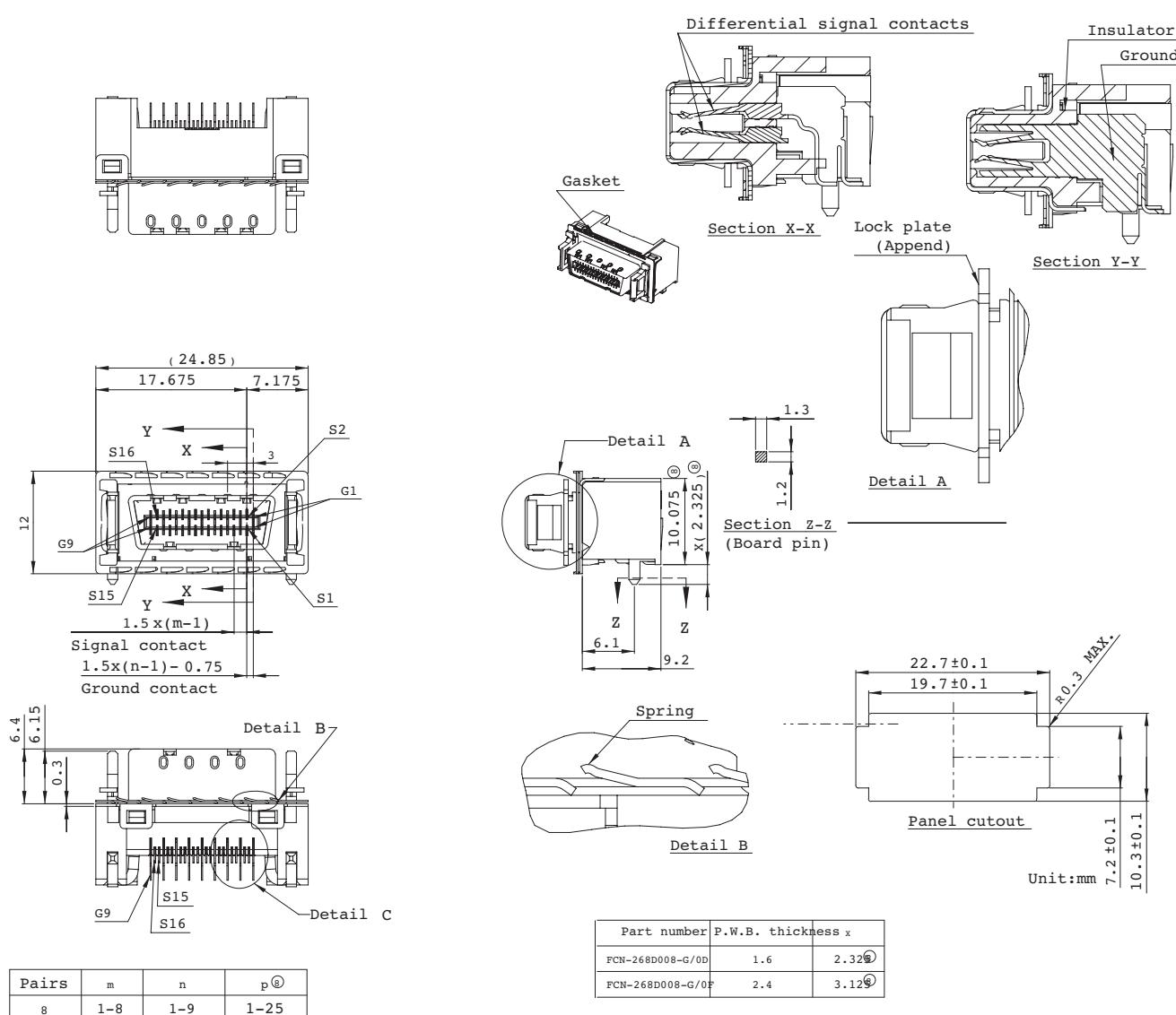
C7-4: 4X Board receptacles used on InfiniBand modules that are not pluggable interfaces shall be intermateable with this connector.

C7-4.1.1: A metal backshell or other means which fully shields the connector shall be bonded to the cable bulk shield through a continuous 360 degree contact to minimize EMI (Electromagnetic Interference).

Side latches are used in conjunction with receptacle features for retention, and are released by pulling on the “lanyard” handle shown in the drawing.

7.3.2.1 HOST BOARD CONNECTOR

This section defines the host board connector for the 4X microGigaCN interface. An exemplary connector is shown in [Figure 107 on page 358](#), with the host board footprint in [Figure 108 on page 359](#). Connector pin numbers are indicated for information. The connector is surface-mounted, so there are no contacts on the secondary side of the board.



Rem1) Unless otherwise specified, dimensions are in mm and tolerances are ±0.5mm.
 Rem2) Coplanarity are 0.1mm

Figure 107 4X cable board connector

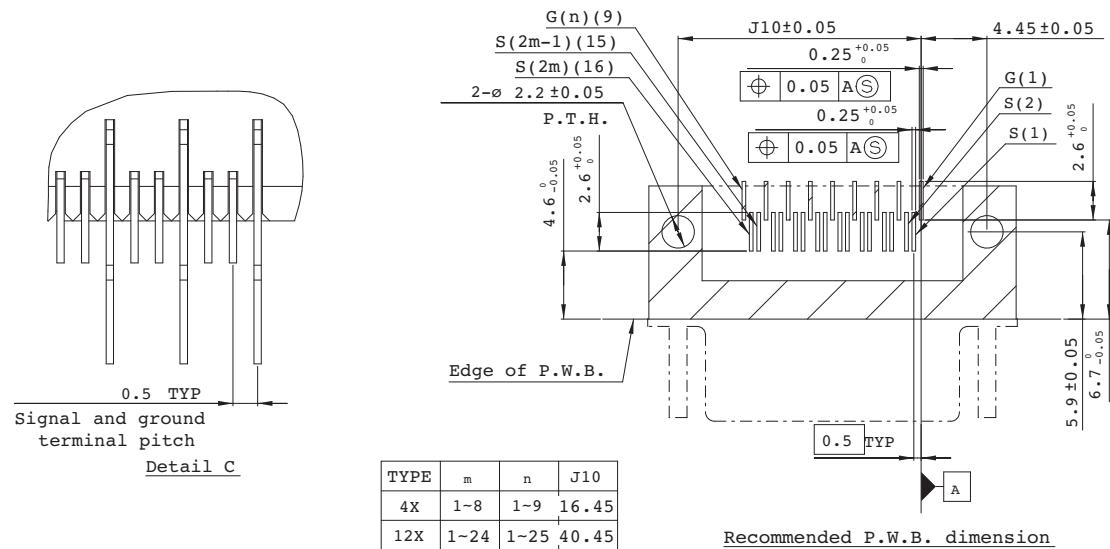


Figure 108 4X cable board connector footprint, top view

The connector uses pairs of contacts separated by Ground contacts to reduce near end crosstalk (NEXT). Detailed drawings of mating interface dimensions are shown in [Figure 109](#).

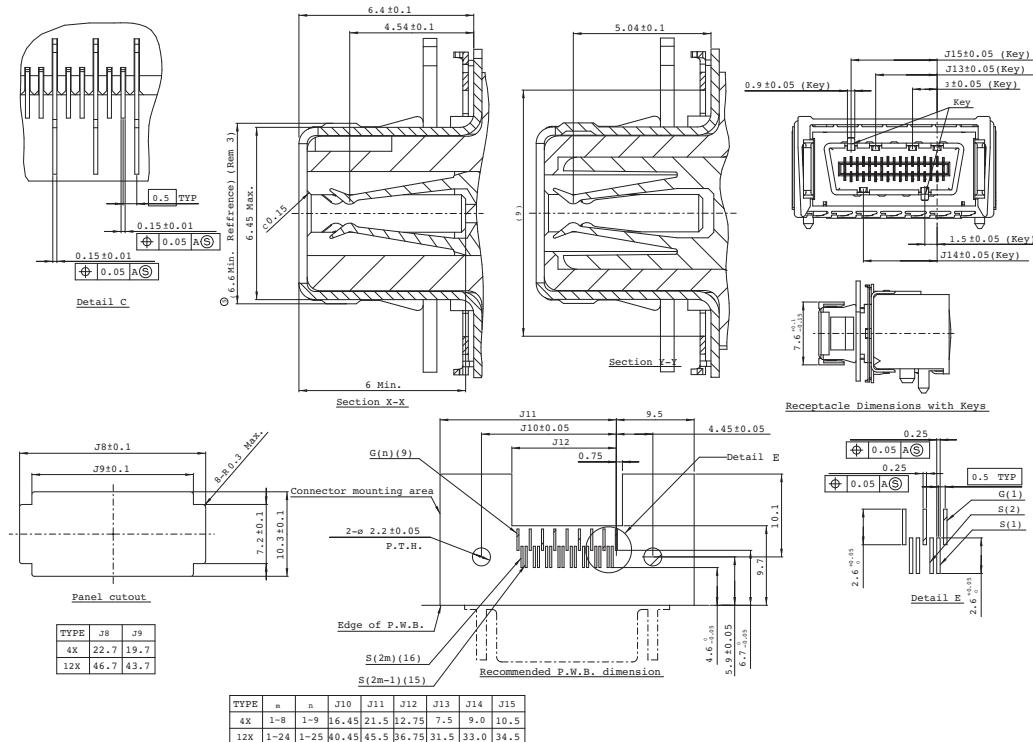


Figure 109 4X/12X cable board connector interface details

7.3.2.2 PIN ASSIGNMENT - 4X PASSIVE CABLE PORTS

C7-5: The pin assignment listed in [Table 91](#) shall be used for the board connector for passive InfiniBand 4X cables using microGigaCN connectors.

The character 'x' in the signal symbol is the port number, as defined in [Section 4.1, "Signal Naming Conventions," on page 71](#).

Table 91 4X passive board connector signal assignment

Pin Number	Signal
G1-G9	Signal Ground
S1	IBtxIp(0)
S2	IBtxIn(0)
S3	IBtxIp(1)
S4	IBtxIn(1)
S5	IBtxIp(2)
S6	IBtxIn(2)
S7	IBtxIp(3)
S8	IBtxIn(3)
S9	IBtxOn(3)
S10	IBtxOp(3)
S11	IBtxOn(2)
S12	IBtxOp(2)
S13	IBtxOn(1)
S14	IBtxOp(1)
S15	IBtxOn(0)
S16	IBtxOp(0)
Housing	Chassis Ground

C7-6: Signal Ground shall be connected to **IB_Sh_Ret** on the module.

C7-7: Signal Ground shall not be connected to Chassis Ground in the connector. This is for meeting EMI and ESD requirements. For discussion, see [Volume 2-DEPR: Section 1.5.4](#).

C7-7.1.1: A continuous ground path from the cable's inner shield(s) through the connector to the board signal ground shall be provided to insure low jitter, low crosstalk and EMI containment.

A cable constructed with only a bulk shield is not likely to meet the electrical requirements of [Chapter 6: High Speed Electrical Interfaces](#).

7.3.2.3 PIN ASSIGNMENT - 4X ACTIVE CABLE PORTS

C7-7.2.2: The pin assignment listed in [Table 92 on page 362](#) shall be used for the board connector for active InfiniBand 4X cables using microGigaCN connectors. Usage requirements including current limitations on the power pins are described in [Section 7.3.3.2, "Active Cable power requirements," on page 367](#). Power return is by way of the Signal Ground pins.

Table 92 4X active board connector signal assignment

Pin Number	Signal
G1	Sense-12V
G2-G6, G9	Signal Ground
S1	IBtxIp(0)
S2	IBtxIn(0)
S3	IBtxIp(1)
S4	IBtxIn(1)
S5	IBtxIp(2)
S6	IBtxIn(2)
S7	IBtxIp(3)
S8	IBtxIn(3)
S9	IBtxOn(3)
S10	IBtxOp(3)
S11	IBtxOn(2)
S12	IBtxOp(2)
G7	Sense-3.3V
S13	IBtxOn(1)
S14	IBtxOp(1)
G8	Vcc
S15	IBtxOn(0)
S16	IBtxOp(0)
Housing	Chassis Ground

The 12 V or 3.3 V Sense signal is used to enable the respective voltage on the Vcc power supply pin used to provide power to the active components in the cable. Both signals are considered active if their voltage level is between 0.9 V and 2.4 V DC.

C7-7.1.3: Both 12 V and 3.3 V Sense signals shall not be active simultaneously.

The character 'x' in the signal symbol is the port number, as defined in [Section 4.1, "Signal Naming Conventions," on page 71](#).

7.3.2.4 CABLE PLUG

This section defines the cable plug for the 4X microGigaCN interface. An exemplary connector is shown in [Figure 110](#) and [Figure 111 on page 364](#).

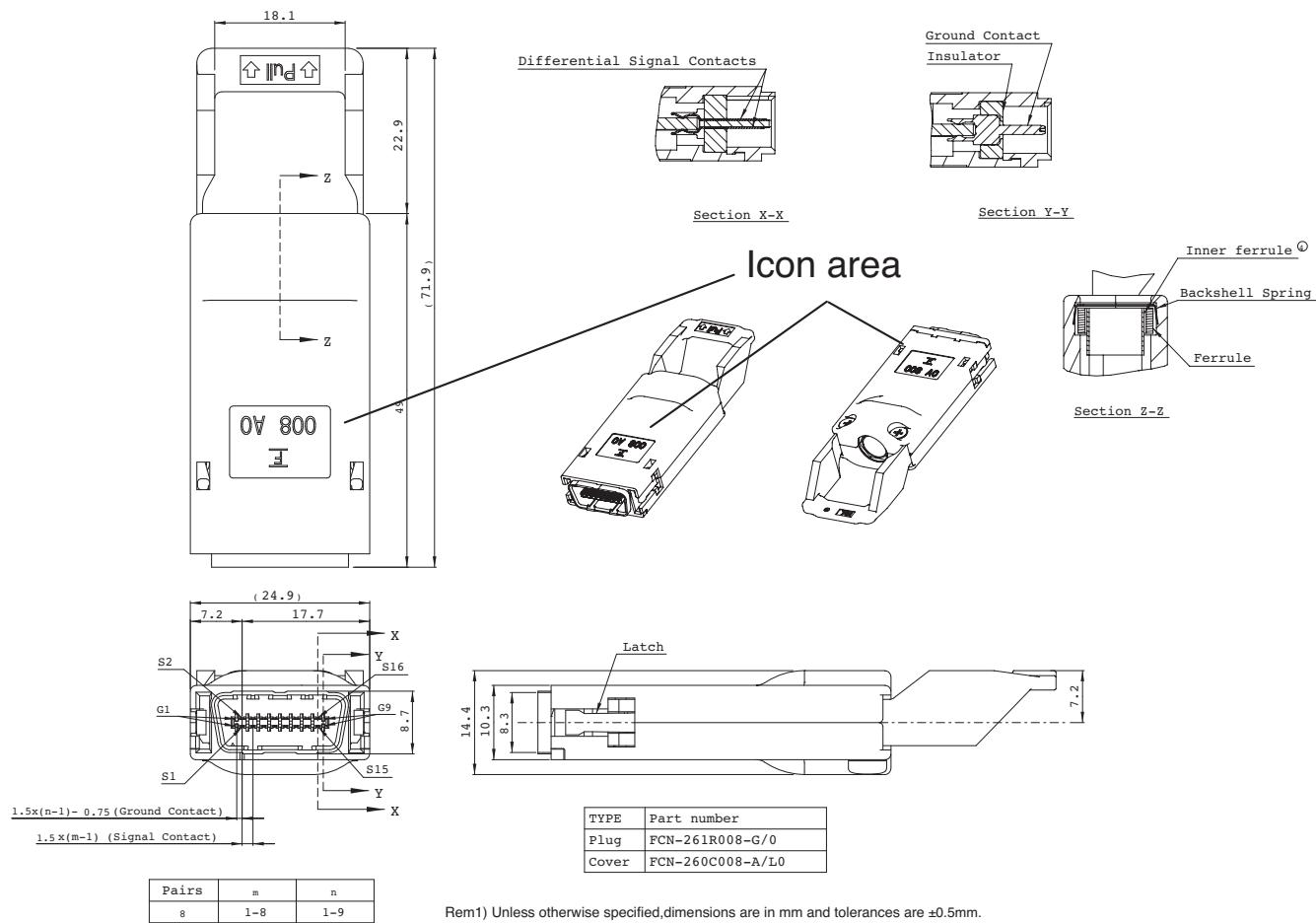


Figure 110 4X cable plug

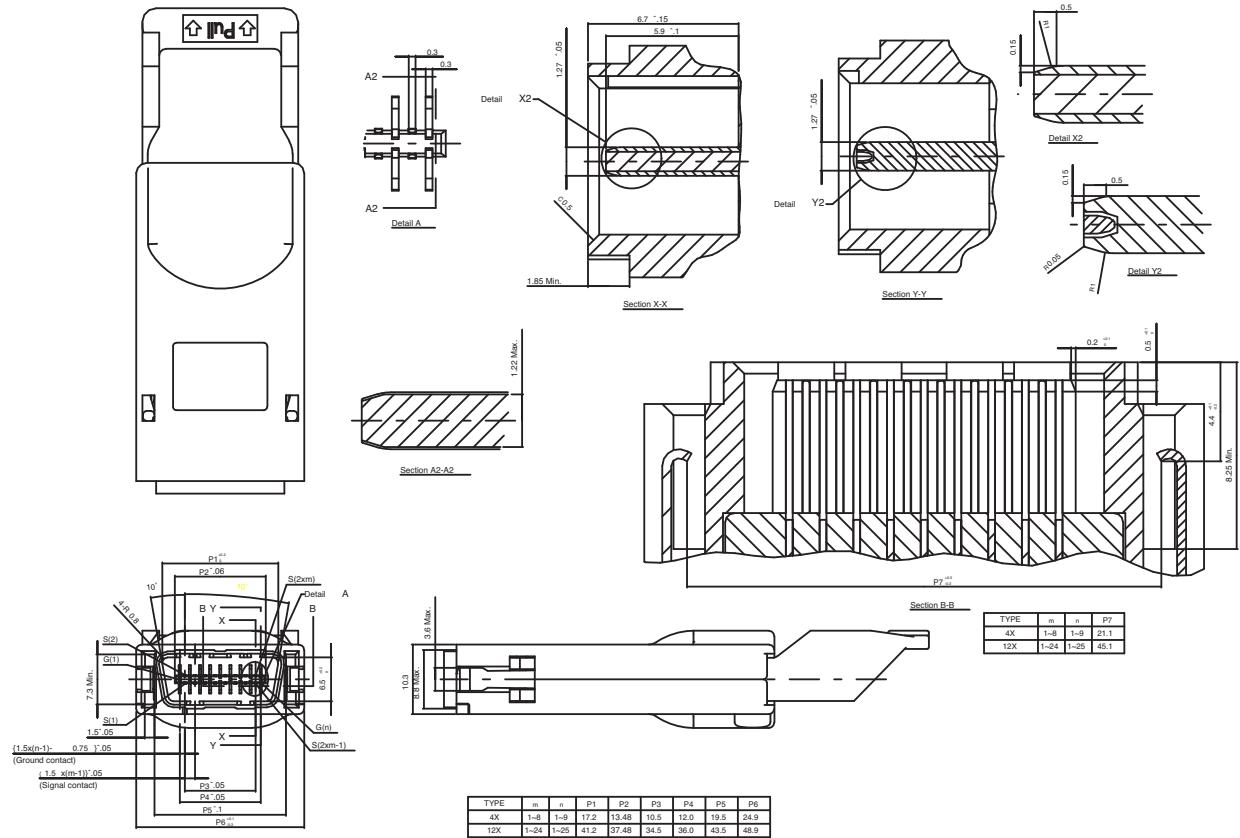
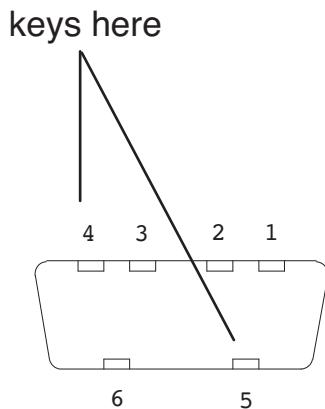


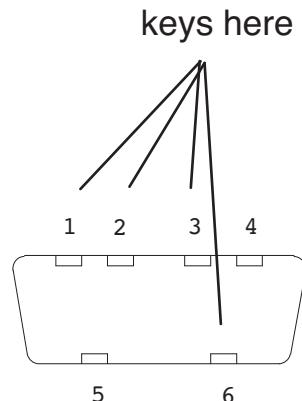
Figure 111 4X/12X cable plug interface details

7.3.2.5 KEYING

It is recommended that the 4X board and cable microGigaCN connectors used for InfiniBand use keys in positions 4 and 5, shown in [Figure 112 on page 365](#) as viewed from the outside of the chassis. This is to prevent misplugging with other interfaces that might have chosen to use the same physical connector.



(a) board connector



(b) cable connector

Figure 112 4X microGigaCN board and cable connector keying

7.3.3 ELECTRICAL REQUIREMENTS

7.3.3.1 MATED CONNECTOR ELECTRICAL PARAMETERS

C7-8: Connectors to be used on and for connection to InfiniBand modules shall meet or exceed the electrical performance requirements for port type 1 which are listed in [Table 93](#).

Table 93 MicroGigaCN Cable connector electrical performance requirements

Symbol	Parameter	Minimum	Maximum	Unit	Conditions/Comment
LLCR	Low level contact resistance - initial		80	mΩ	through testing per EIA-364-23
ΔLLCR	Low level contact resistance - change		20	mΩ	through testing per EIA-364-23, as a result of any test group step
I _{max}	Current rating	0.5		A	per EIA-364-70 or IEC 512-5-1 Test 5a, at 30° C. temperature rise above ambient

Table 93 MicroGigaCN Cable connector electrical performance requirements (Continued)

Z _{dco} (peak)	Differential Impedance (peak)	90	110	Ω	Mated cable and board connector, average value measured over the propagation delay of the connector at (UI/4) ps rise time (at the connector) per EIA-364-108. Includes connector, cable to connector interface, and board termination pads and vias but not equalizer.
Z _{dco} (nom)	Differential Impedance (nominal)	95	105	Ω	
L _{co}	Insertion loss	1.0	dB		Mated cable assembly and board connector, at frequencies up to 1.25 GHz, per EIA-364-101
S _{cop}	Within pair skew	5	ps		per EIA-364-103; by design, measurement not required
J _{co}	Jitter	10	ps		per EIA-364-107 with Fibre Channel CJTPAT stimulus until 10000 hits in max. 20 mV high jitter box or equivalent, with equipment and fixture contribution de-embedded; by design, measurement not required.
NEXT _c	Near end crosstalk	4	%		Mated cable and board connector, measured differentially with all adjacent neighbor pairs driven at (UI/4) ps transition time, per EIA-364-90 and 6.6 on page 282 . A lower value will result in additional design margin. See Note 2.

The signal eye pattern at the cable output must conform to the electrical requirements as listed in [Section 6.7, “Differential Receiver Inputs,” on page 309](#) and shown in [Figure 86](#) or [Figure 94](#) in all cases.

Architecture Notes

1. All rise or transition times referenced in this chapter of the specification are assumed to be measured from the 20% to 80% levels of the waveform base state to top state amplitudes.
2. Crosstalk is calculated by dividing the larger of the amplitudes of the unsigned positive or negative differential crosstalk noise waveforms by the unsigned amplitude of the differential aggressor waveform, the result being multiplied by 100 to obtain percent. In the case of measurement by superposition, the noise voltage is the sum of the maximum noise amplitude values induced by the individual aggressor pair sources.
3. Impedance measurements are not required on passive equalized cables on the signal pins at the cable end containing the equalizer components, nor on signal pins of active cables at the cable end at which the active circuitry is implemented.

7.3.3.2 ACTIVE CABLE POWER REQUIREMENTS

This section defines the use and control of the power provided for microGigaCN active cables. This power is provided to enable the construction of active devices such as repeaters and electro-optical converters which plug into InfiniBand version 1.2 Cable receptacles.

Since the active cable board connectors are intermateable with Release 1.1 InfiniBand cables, the associated power supplies must tolerate having their outputs shorted to ground for an indefinite period. Sense pins are supplied to assist in managing and switching the power supply output.

The IB active cable will generate the design specific voltages from the supplied active cable voltage. These converters will be located in the Active Cable assembly.

7.3.3.2.1 Host Power Supply Filtering

The host board should use the power supply filtering network shown in [Figure 113 on page 368](#), or an equivalent. Any voltage drop across a filter network on the host is counted against the host DC set point accuracy specification. Inductors with DC Resistance of less than 0.1Ω should be used in order to maintain the required voltage at the Host Edge Card Connector. Selection of the filter capacitors (shown as 22 uF in [Figure 113](#)) is left to the implementer, but it is recommended that 22 uF capacitors with approximately 0.22Ω of equivalent series resistance be used to provide adequate filtering

without high frequency ringing. The time constant of the filter circuit is the important quantity.

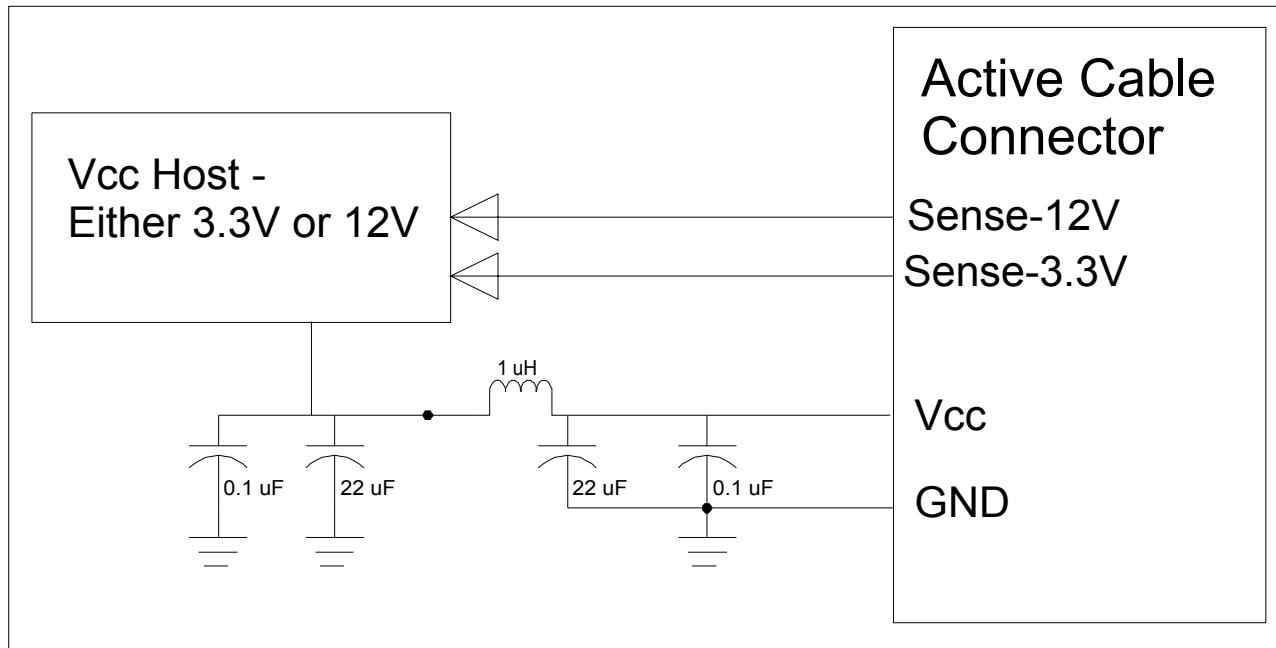


Figure 113 Recommended active cable host board power supply filtering

All active cable power pins, including sense pins shall provide a low impedance to ground at frequencies from 100 MHz to 0.75 times the maximum bit rate supported on the port. This bypassing shall be provided on both the cable assembly and the receptacle.

A host may support power for either 12V, or 3.3V or both.

7.3.3.2.2 HOST POWER SUPPLY SPECIFICATIONS

The specifications for the power supply are listed in [Table 94](#).

Table 94 Power supply specifications

Parameter	Min.	Nominal	Max.	Unit	Condition
Vcc_host		3.3 or 12		V	Measured at Vcc
Vcc set point accuracy	-5		+5	%	Measured at Vcc
Vcc Power supply noise including ripple			50	mVpp	1 kHz to frequency of operation
Module Inrush Current			600	mA	each inductor in the power supply filter

7.3.3.2.3 VOLTAGE

C7-8.2.1: The Active Cable Power (ACP12) voltage shall be a minimum of 10 volts at maximum current and 14 volts maximum at any current from zero to maximum, measured at the Vcc pins at the cable receptacle.

C7-8.2.2: The Active Cable Power (ACP3) voltage shall be a minimum of 3.13 volts at maximum current and 3.47 volts maximum at any current from zero to maximum, measured at the Vcc pins at the cable receptacle.

7.3.3.2.4 POWER

C7-8.2.4: Each end of an active cable assembly using a 12 V nominal power supply shall consume a maximum of 7 watts at maximum voltage for a 4x cable, or 10 watts for an 8x or 12x cable.

C7-8.2.3: Each end of an active cable assembly using a 3.3 V nominal power supply shall consume a maximum of 1.7 watts at maximum draw for a 4x cable, or 5.2 watts for an 8x or 12x cable.

7.3.3.2.5 CURRENT

C7-8.2.5: The Active Cable Power shall be capable of supplying a minimum of 500 mA per ACP (Vcc) power pin in the receptacle. The number of pins varies with the width of the port as defined in the section of this specification on active cables.

C7-8.2.6: The active cable assembly shall draw no more than 500 mA maximum from any power pin at any voltage between the minimum and maximum limits.

7.3.3.2.6 SHORT CIRCUIT PROTECTION

C7-8.2.7: The active cable power supply shall protect itself against indefinite connection to ground and shall limit short circuit current to less than 50 mA when the equivalent load resistance is less than 1 ohm.

7.3.3.3 SHORT CIRCUIT PROTECTION

C7-8.2.8: The active cable power supply shall protect itself against indefinite connection to ground and shall limit short circuit current to less than 50 mA when the equivalent load resistance is less than 1 ohm.

7.3.3.4 LOAD IMPEDANCE

C7-8.2.9: The load shall present a maximum steady state current draw of 500 mA in parallel with a maximum capacitance of 500 microfarads on any Active Cable Power pin. The load shall limit the current drawn to 500 mA from any power pin.

7.3.3.4.1 POWER SENSE

C7-8.2.10: Active cable assemblies shall connect a $5\text{ k}\Omega$ +/- 5% resistor from SENSE-3.3V or SENSE-12V to ground to enable the active cable power. Sense

C7-8.2.11: The active cable assembly shall connect a 5 kW +/- 5% resistor from SENSE-3.3V or SENSE-12V to ground to enable the active cable power.

C7-8.2.12: The ACP circuitry shall only enable power to the receptacle when the presence of the SENSE resistor is detected. If SENSE is connected to ground, or is open (no cable), ACP shall be disabled.

C7-8.2.13: Active Cable Power shall not be active when Bulk Power is not available.

7.3.4 ENVIRONMENTAL REQUIREMENTS

7.3.4.1 ELECTROSTATIC DISCHARGE (ESD)

Cables using the microGigaCN interface shall withstand 2 kV of electrostatic contact discharge to the receptacle housing or plug body using the Human Body Model per JEDEC Standard JESD22-A114:B2000 without damage or non-recoverable error including but not limited to latchup. A recoverable error is one that does not require reset or replacement of the device.

Cables shall meet ESD requirements given in EN 61000-4-2, criterion B test specification such that when installed in a properly grounded housing and chassis the module is subjected to 15 kV air discharges during operation and 8 kV direct contact discharges to the case.

7.3.4.2 HOT INSERTION AND REMOVAL

Cables shall not be damaged by removal or insertion. Removal may occur while the link is operating without damage to either port or the link. Insertion or removal may occur with power on or power off.

The active cable power shall detect the sense pin value and provide full current availability within 50 milliseconds.

7.4 PLUGGABLE INTERFACE CONNECTORS

7.4.1 INTRODUCTION

Unlike the microGigaCN connectors described in the preceding sections, which are designed for passive copper cables (but can support active cables that emulate passive cables), a "pluggable" interface port is designed to support active modules or active cables, using either optical or copper transmission. Pluggable interface ports can also support passive copper cables, with appropriate adjustments in high-speed signaling specifications.

Active copper or optical cables may have either linear transmission characteristics, or may use limiting amplifiers, as described in [Section 6.2, "Electrical Topologies," on page 266](#). Different high-speed interface specifications may be needed for linear vs. limiting transmission technologies, but the mechanical interfaces, as described in the following sections, are independent of transmission technology used.

Due to the need, at SDR signaling rate, to account for the jitter introduced in the assembly between the board and the cable, a specific port type, referred to as Port Type 2, is defined for these devices. The unique electrical requirements for port type 2 are listed in [Table 48 on page 285](#).

Architecture Note

Port type 2 was defined to support pluggable SFP transceivers that are compliant with the SFF-8074 specification and operate at 2.5 Gb/s (SDR). These devices may be used with detachable optical cables, therefore necessitating a tighter jitter specification on the SDR transmitter than for passive copper or active optical or copper cable interfaces.

A pluggable interface port is designed to be used with either copper or optical devices. Pluggable modules are designed to support transmission over passive cables. A pluggable module may either be part of the cable assembly (i.e., pigtailed), or may be detachable from the cable assembly. Pluggable modules will typically incorporate active transceiver components in the module, but may also operate purely passively.

Active cables or modules may contain circuitry to compensate for the jitter introduced in the cable assembly. The active optical device external interface is defined in [Section 9.10, "Optical Pluggable Modules," on page 612](#). The passive copper connector is permanently attached to the cable, so the interface to the host board is the only separable interface.

7.4.2 OBSOLETE PLUGGABLE INTERFACE CONNECTORS

Implementation Note

This section describes a deprecated feature, which may not be compatible with current InfiniBand architecture features and functions.

The following list enumerates previously-defined interfaces that have since been superseded. The following pluggable interfaces are defined:

1x Pluggable interfaces:

- "1x" (HSSDC) - This interface is obsolete. Refer to Volume 2, Release 1.2.1 of the InfiniBand specification for information.

4x interfaces

- "4x Pluggable" - The 4x Pluggable interface is based in part on the XPAK MSA. The mechanical specifications are identical to XPAK, while the High Speed electrical and optical interface specifications are based on the InfiniBand specifications listed in

the InfiniBand Architecture Specification, [Volume 2, Chapters 6 and 8](#). This interface supports 4x InfiniBand electrical interfaces at SDR speeds. The physical size of the module is 42 mm wide x 72 mm deep. Its use for new designs is not recommended.

IBTA

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

7.5 4X QSFP+ INTERFACE CONNECTORS

7.5.1 INTRODUCTION

The 4x QSFP+ interface is based in part on the QSFP/QSFP+ MSA which is specified in SFF-8436, with enhanced versions suitable for FDR in SFF-8685 and for EDR in SFF-8665. The original QSFP specification, INF-8438, “QSFP 4 Gb/s 4X Transceiver (Quad SFP)” specified lower-speed operation, and has been superseded for InfiniBand applications by QSFP+ specifications.

The mechanical specifications for 4x QSFP+ are identical to the QSFP+ SFF specifications, while the electrical interface specifications and some memory map register value specifications are based on the InfiniBand specifications listed in the InfiniBand Architecture Specification, [Volume 2, Chapter 6](#) and [Chapter 8](#). The enhanced QSFP+ module and host board connector are used for InfiniBand FDR and EDR interfaces. Enhanced QSFP+ connectors and modules shall be intermataable with standard QSFP connectors and modules. [Figure 114](#) and [Figure 115](#) illustrate the QSFP+ module and cage assembly. [Figure 117](#) shows the host board connector that is contained inside the EMC cage. Note that this section describes QSFP+ interface connector specifications. [Section 7.9.5 on page 465](#), describes specifications for cables.

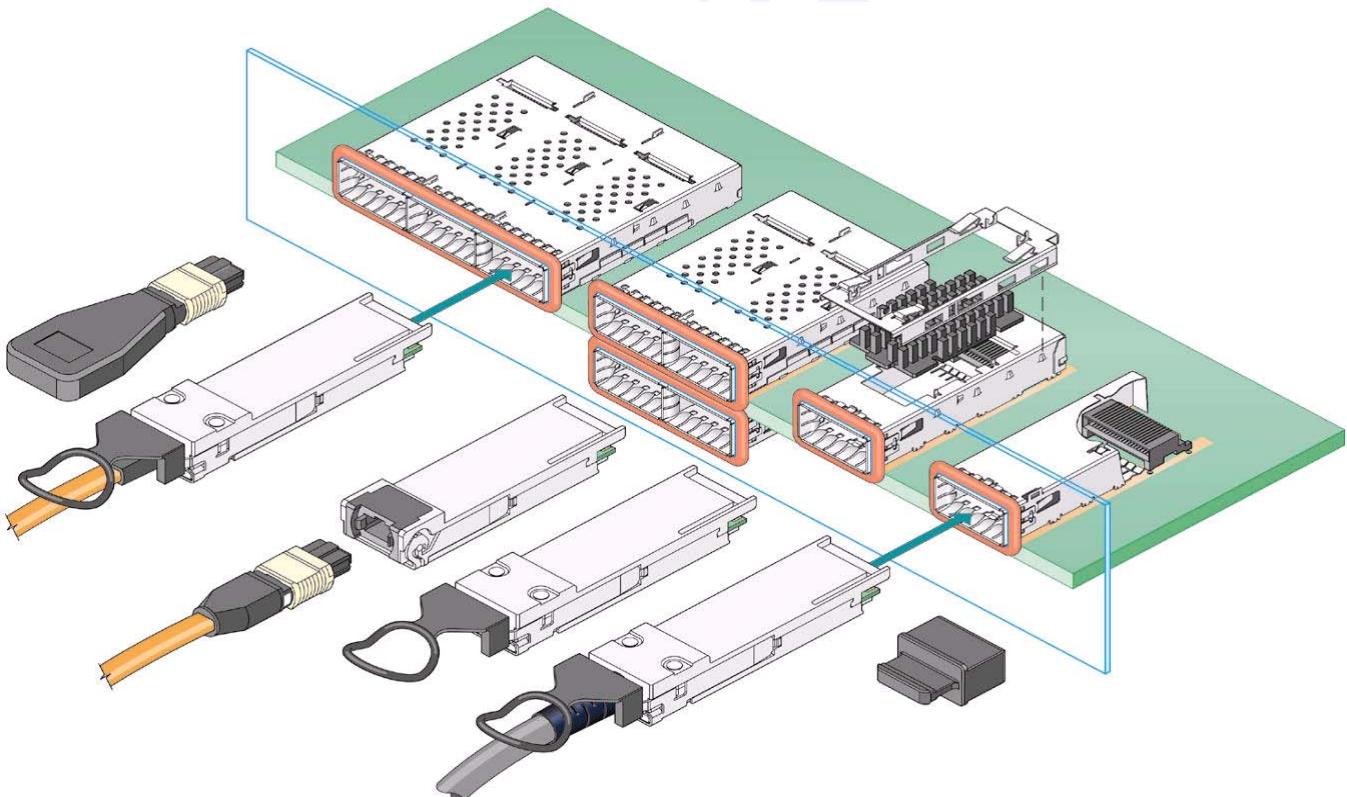


Figure 114 QSFP Conceptual Model

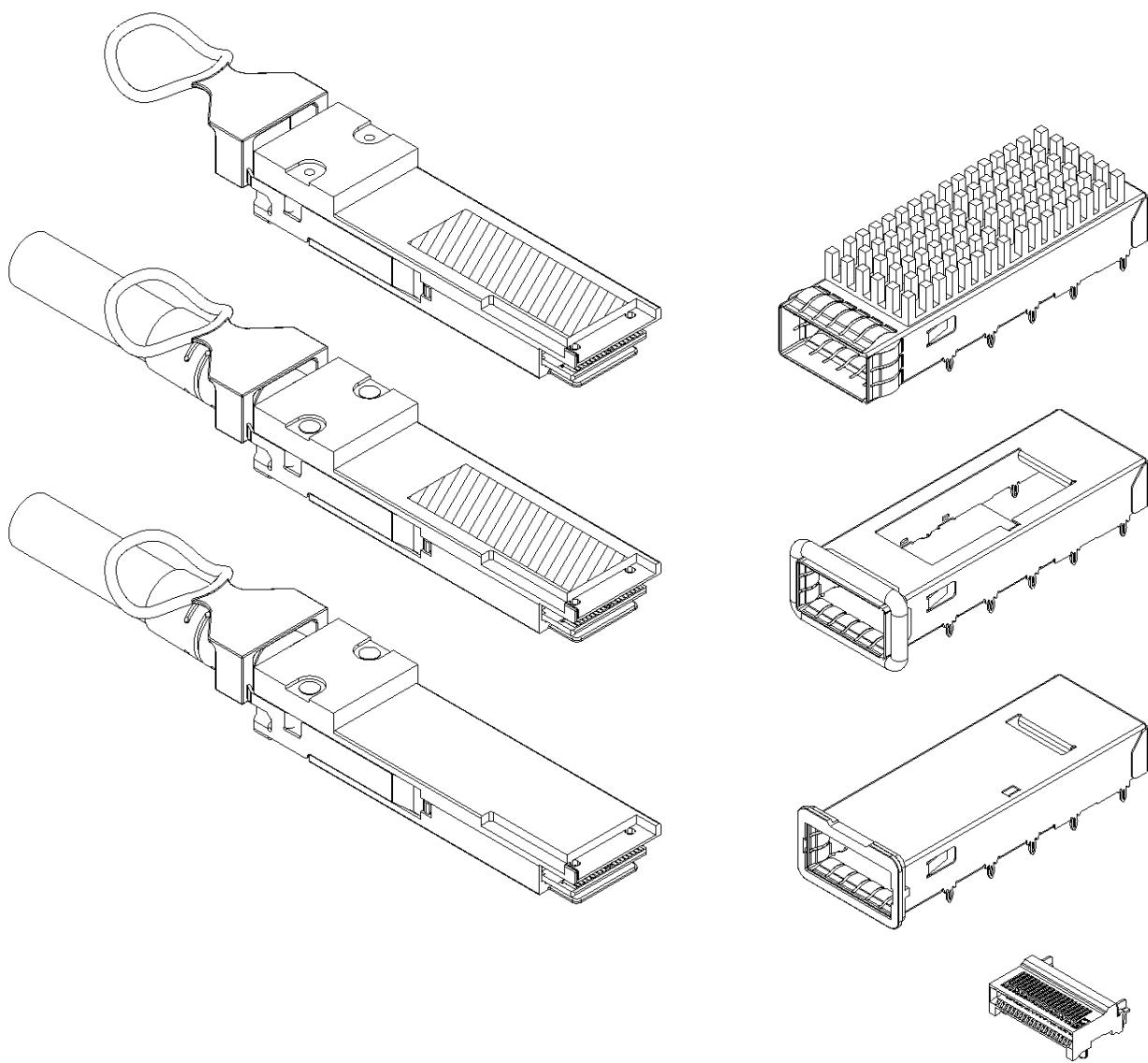


Figure 115 QSFP+ pluggable modules and host board receptacle cages

Architecture Note

Although 8X and 12X pluggable interfaces are defined using the CXP interface below, 8x and 12x links may also be supported using multiple 4x pluggable interface modules, as described in [Section 7.9.5.4, “Two 4x QSFP+ cables for 8x configuration,” on page 469.](#)

7.5.2 MECHANICAL REQUIREMENTS

7.5.2.1 PHYSICAL AND MECHANICAL PERFORMANCE REQUIREMENTS

C7-8.1.4: Connectors to be used for pluggable ports on InfiniBand modules shall meet or exceed the physical and mechanical performance requirements listed in [Table 95](#). See SFF-8436 or SFF-8661 for further details.

The requirements for insertion forces, extraction forces and retention forces are specified in [Table 95](#). QSFP+ modules, connectors, receptacles and receptacle housings should not be damaged by module removal or insertion. If any part is damaged by excessive force, it should be the cable or module, and not the receptacle or receptacle housing which are part of the host system.

Table 95 QSFP+ connector physical requirements

Symbol	Parameter	Minimum	Maximum	Unit	Conditions/Comments
F_i	Insertion force	0	40	N	
F_w	Withdrawal force	0	30	N	
F_{rx}	Transceiver retention force	90		N	no damage to transceiver below 90 N
F_{rcf}	Cage retention (latch strength)	125		N	no damage to latch below 125 N
N_{hc}	Durability, host connector and cage	100		cycles	
N_x	Durability, transceiver	50		cycles	

It is also recommended that the connector interfaces meet the parameters defined in [Table 96](#).

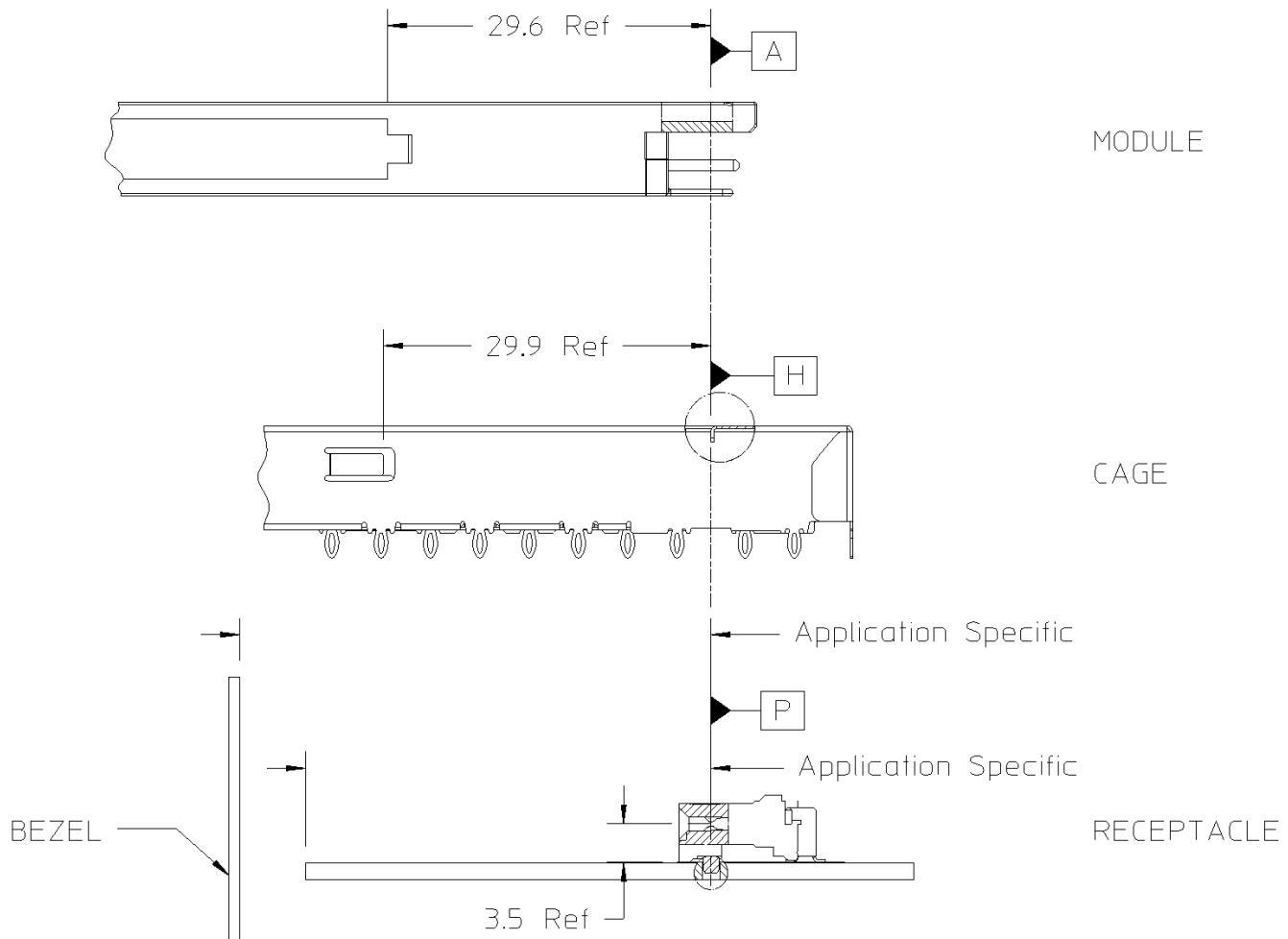
Table 96 Recommended QSFP+ connector physical parameters

Symbol	Parameter	Min	Max	Unit	Comments
t_{pm}	Contact finish	0.76 Au over 1.27 Nickel		μm	As necessary to meet N_x requirements
F_n	Contact normal force	50		cN	per contact
S_{hcc}	Contact Hertz stress	170		kpsi	per contact
D_{wc}	Contact wipe length	0.75		mm	

It is recommended that all components and attach processes used for those components be compliant with RoHS directive 2002/95/EC issued January 27, 2003.

7.5.2.2 QSFP+ DATUMS AND COMPONENT ALIGNMENT

A listing of the datums for the various components is contained in [Figure 116 on page 376](#). The alignments of some of the datums are noted. All dimensions are in millimeters.



Datum	Description	Datum	Description
A	Width of Plug/Module	J	Bottom of Cage
B	Plug/Module Hard Stop	K	Height of Cage Opening
C	Top of Plug/Module	L	Width of Cage Opening
D	Bottom of Plug /Module	M	Width of Card Slot Opening
E	Top of Plug/Module PCB	N	Bottom of Receptacle Housing
F	Leading Edge of Outermost Signal Contact Pads	P	Centerline of Contacts
G	Width of Plug/Module PCB	R	Top of PCB
H	Cage Hard Stop	-	-

Figure 116 QSFP+ Module Datum Definitions

7.5.2.3 HOST BOARD CONNECTOR AND CAGE

This QSFP+ host board connector dimensions are shown in [Figure 117 on page 378](#). The dimensions for the host board cage for the QSFP+ pluggable module are shown in [Figure 118 on page 379](#).

7.5.2.4 CONTACT ASSIGNMENT

[Figure 119 on page 377](#) shows the contact numbering for the QSFP+ module and host board connector.

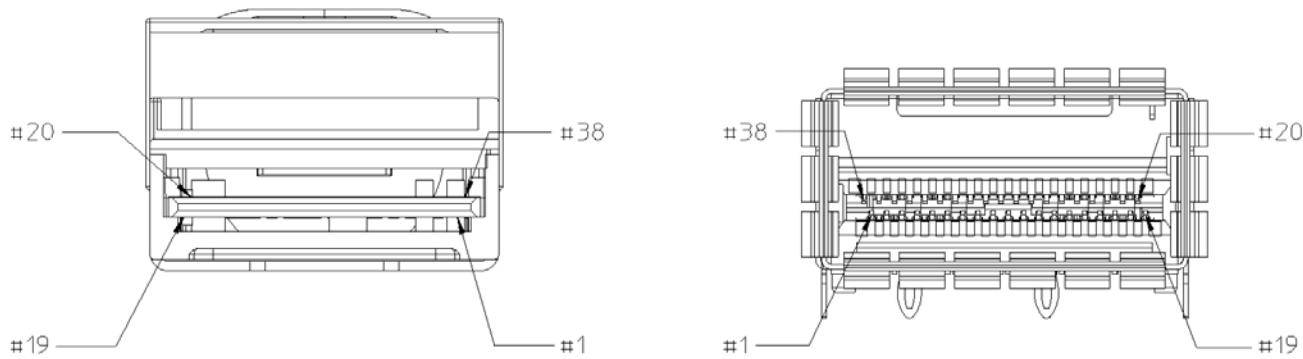
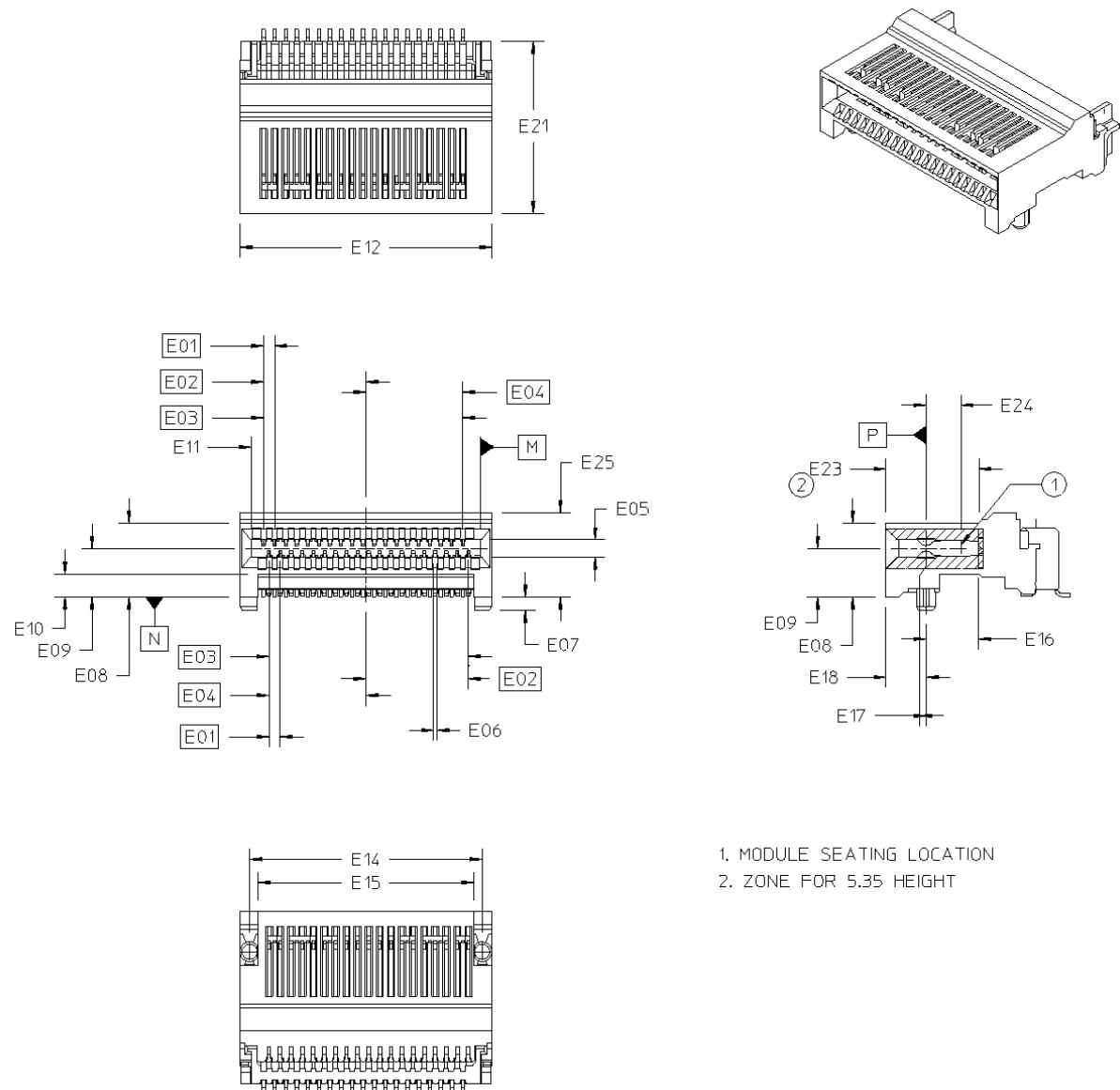


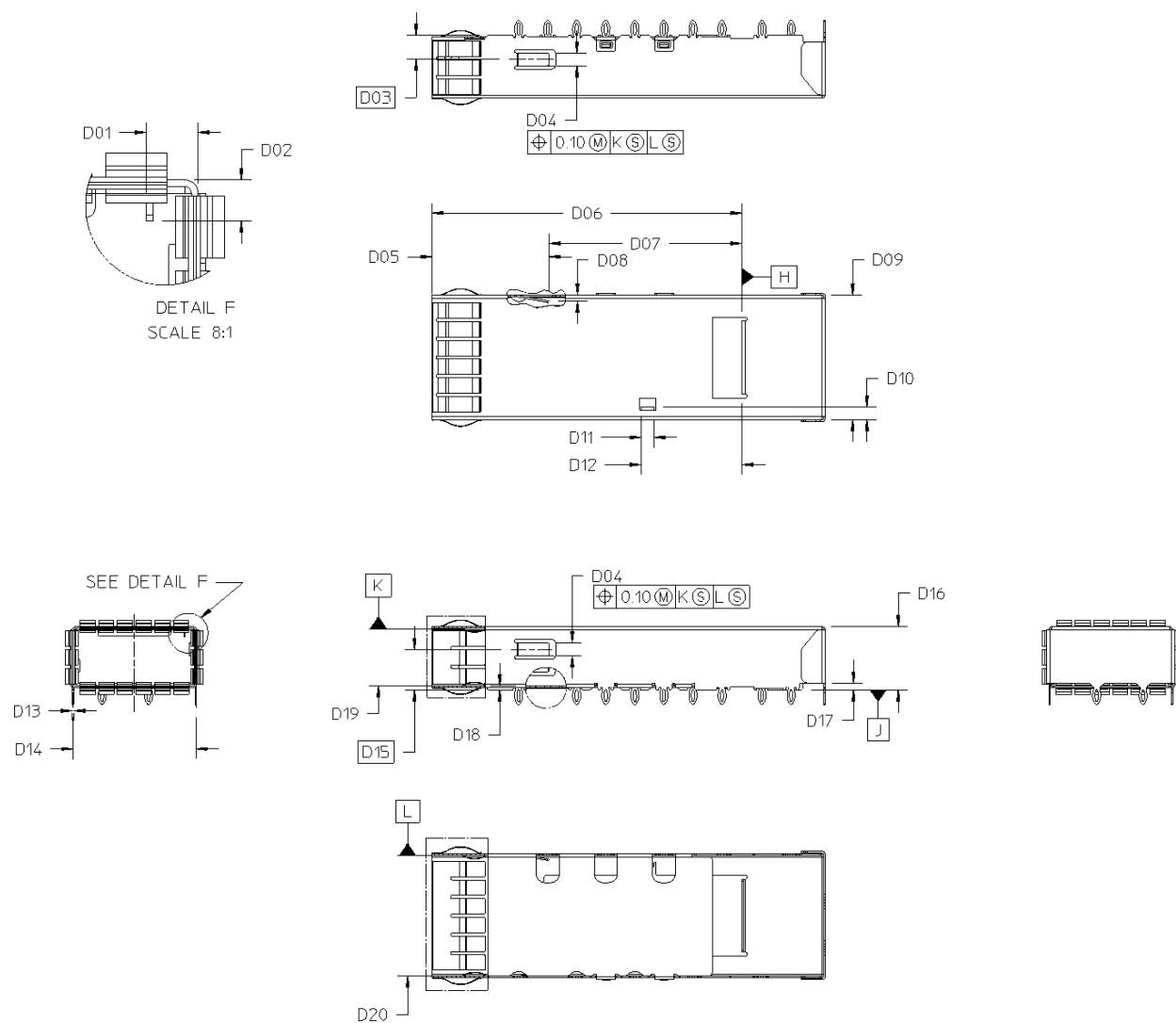
Figure 119 QSFP+ module and host board connector contact assignment



1. MODULE SEATING LOCATION
 2. ZONE FOR 5.35 HEIGHT

ID	Description	Dim	Tol (±)	ID	Description	Dim	Tol (±)
E01	Contact Pitch (within Row)	0.80	Basic	E11	Card Slot Width	16.60	0.10
E02	Centerline to First Contact	7.40	Basic	E12	Receptacle Width	18.20	0.10
E03	First to Last Contact (within row)	14.40	Basic	E13	Receptacle Length	11.50	Max.
E04	Centerline to Last Contact	11.40	Basic	E14	Peg to Peg	16.80	Basic
E05	Card Slot Height	1.26	Min.	E15	Leg to Leg	15.53	0.13
E06	Contact Zone (0.18 wide terminal)	0.30	Max.	E16	Card Slot Depth	3.20	Min.
	Contact Zone (0.20 wide terminal)	0.32	Max.	E17	Peg to Contact Centerline	0.00	0.10
	Contact Zone (0.22 wide terminal)	0.34	Max.	E18	Front Face to Peg	2.90	Basic
	Contact Zone (0.25 wide terminal)	0.37	Max.	E19	Receptacle Length	12.69	Max.
E07	Peg Height	0.95	0.10	E20	Zone 2 Length	6.50	Min.
E08	Mating Zone Height	5.35	0.13	E21	Module Seating Location	2.50	Ref.
E09	PCB to Card Slot Centerline	3.50	0.10	E22	Maximum Height	6.30	Max.
E12	Receptacle Width	18.20	0.10	-	-	-	-

Figure 117 QSFP+ host board connector



ID	Description	Dim	Tol (\pm)	ID	Description	Dim	Tol (\pm)
D01	Polarizing Key Location	2.00	Ref.	D10	Cage Side Wall to Polarizing Key	2.00	0.20
D02	Polarizing Key Height	1.60	0.20	D11	Polarizing Key Length	2.00	Min.
D03	Location of Left Latch	3.75	Basic	D12	Datum to Polarizing Key	15.60	0.30
D04	Latch Width	2.00	0.05	D13	Wall Thickness	0.25	0.03
D05	Cage Front to Latch Tip	18.10	Ref.	D14	Cage Tail to Cage Tail	19.00	Ref.
D06	Datum to Cage Front - Through the Bezel - Finger	48.000	0.20	D15	Location of Right Latch	6.25	Basic
	Datum to Cage Front - Through the Bezel - Elatometer	44.260	0.20	D16	Cage Height	9.760	Ref.
	Datum to Cage Front - Behind the Bezel	42.800	0.25	D17	Cage Notch Cutout Height	1.00	Min.
D07	Datum to Latch Tip	29.90	0.15	D18	Cage Lower Edge to Cage Inside Floor	0.55	0.10
D08	Latch Tip Location	1.00	Min.	D19	Cage Opening Height	8.80	0.10
D09	Cage Width	19.30	Ref.	D20	Cage Opening Width	18.65	0.10

Figure 118 QSFP+ pluggable module cage dimensions

C7-9: The contact assignment shown in [Table 97 on page 380](#) shall be used for the 4X QSFP+ interface.

Table 97 Contact Assignment for 4x QSFP+ Interface^a ^b

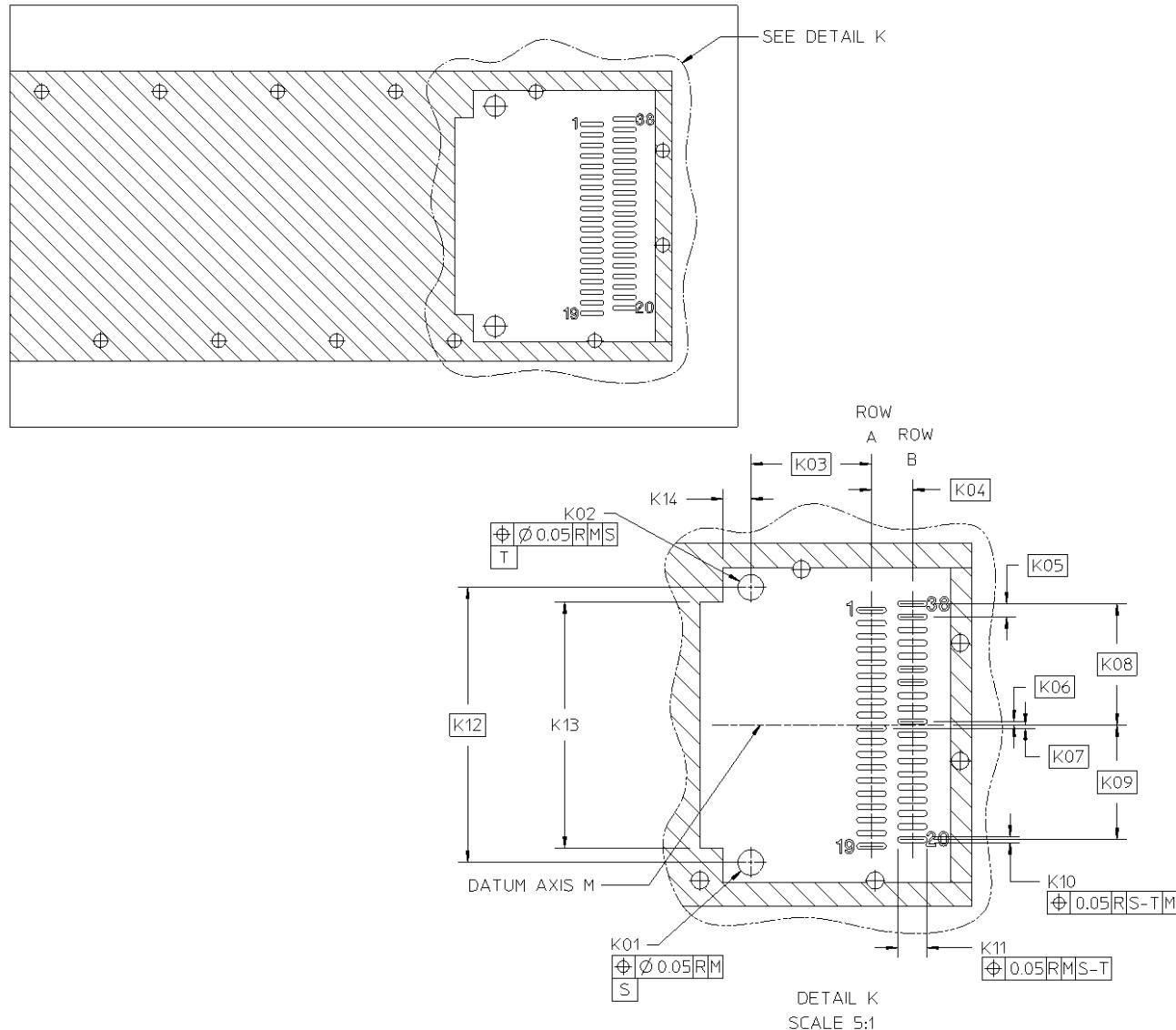
Top side			Bottom Side		
Contact Number	Name	Contact Length	Contact Length	Name	Contact Number
38	GND			GND	1
37	Tx1n			Tx2n	2
36	Tx1p			Tx2p	3
35	GND			GND	4
34	Tx3n			Tx4n	5
33	Tx3p			Tx4p	6
32	GND			GND	7
31	LPMode			ModSelL	8
30	Vcc1			ResetL	9
29	VccTx			VccRx	10
28	IntL			SCL	11
27	ModPrsL			SDA	12
26	GND			GND	13
25	Rx4p			Rx3p	14
24	Rx4n			Rx3n	15
23	GND			GND	16
22	Rx2p			Rx1p	17
21	Rx2n			Rx1n	18
20	GND			GND	19

a. The operation of the low-speed lines (ModSelL, LPMode_Reset_ModPrsL, IntL, SCL, and SDA) for control and status is described in [Section 7.5.3.3 on page 390](#).

b. The several different contact lengths on the card edge allow for hot plug capability and ensure that power and ground connections are not made or broken while high-speed differential connections are active. Actual contact lengths are specified in [Figure 128 on page 389](#).

7.5.2.5 HOST BOARD CONNECTOR FOOTPRINT

Examples of suitable QSFP+ host board connector footprints (informative) are shown in [Figure 120 on page 381](#) (referred to as Style A) and [Figure 121 on page 382](#) (referred to as Style B). The choice of connector (and therefore footprint) is left to the implementer. Note that the two footprints are not interchangeable.



ID	Description	Dim	Tol (±)	ID	Description	Dim	Tol (±)
K01	Locating Hole Diameter	1.55	0.05	K08	Card Centerline To Outer Pad Centerline	7.40	Basic
K02	Locating Hole Diameter	1.55	0.05	K09	Card Centerline To Outer Pad Centerline	7.00	Basic
K03	Locating Hole to Row A - 14 GHz	5.18	Basic	K10	Pad width	0.35	0.03
	Locating Hole to Row G - 28 GHz	7.37	Basic	K11	Pad Length	1.80	0.03
K04	Row A to Row B	2.50	Basic	K12	Locating Hole to Locating Hole	16.80	Basic
K05	Pad Pitch	0.80	Basic	K13	Keep Out Zone	15.02	Max.
K06	Card Centerline to Inner Pad Centerline	0.20	Basic	K14	Keep Out Zone	1.69	Max.
K07	Card Centerline to Outer Pad Centerline	0.20	Basic	K15	-	-	-

Figure 120 Sample QSFP+ host board connector footprint (Style A)

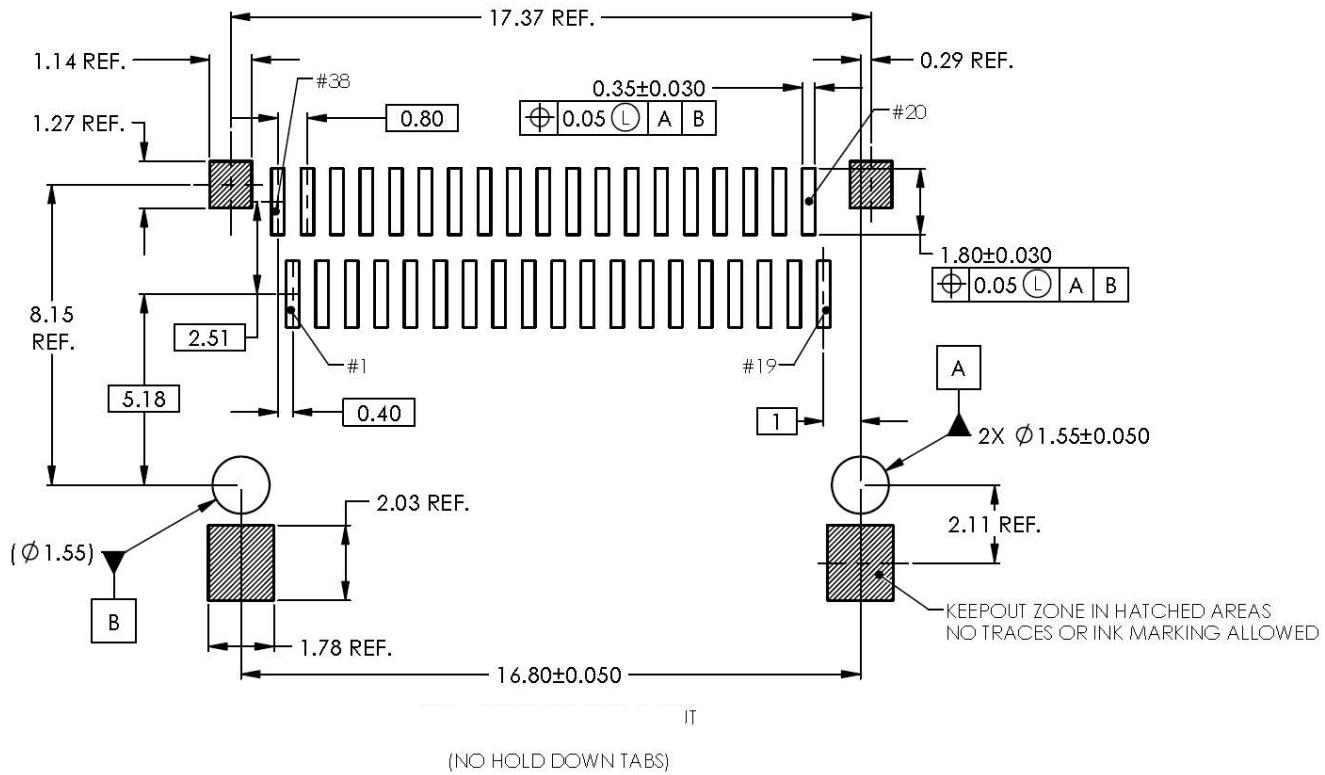
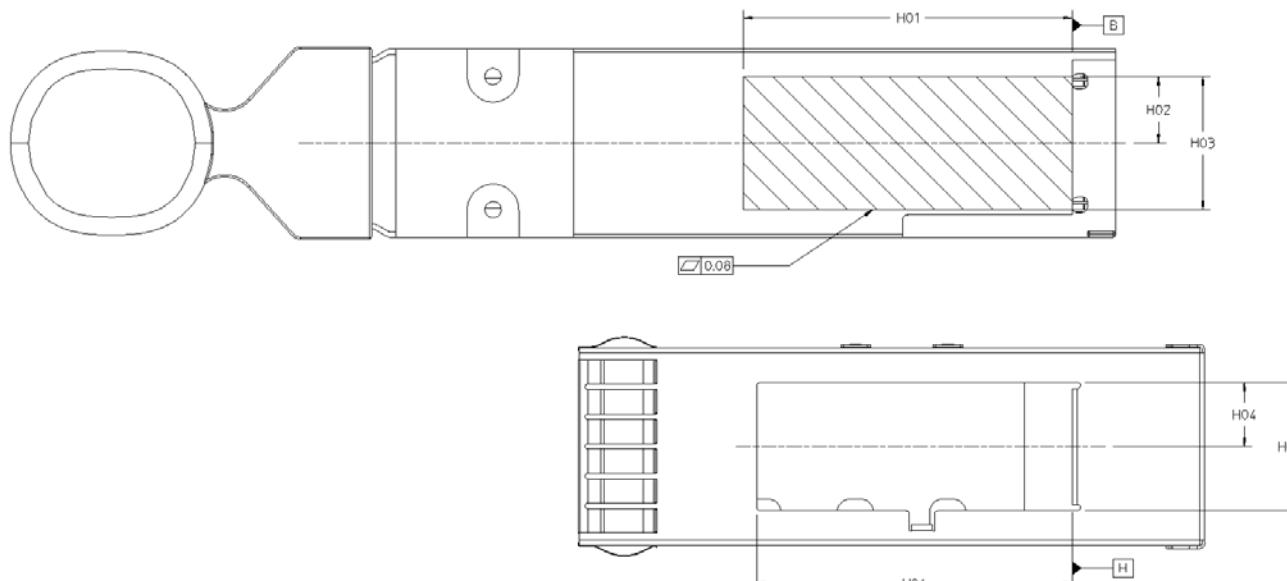


Figure 121 Sample QSFP+ host board connector footprint (Style B)

7.5.2.6 HEAT SINK INTERFACES

Exemplary optional heat sink dimensions for the QSFP+ plug and cage are shown in [Figure 122](#) and [Figure 123 on page 384](#).



ID	Description	Dim	Tol (±)	ID	Description	Dim	Tol (±)
H01	Heat Sink Area Length	32.00	Min.	H04	Top of Heat Sink Area in Cage to Centerline	6.25	Min.
H02	Top of Heat Sink Area to Centerline	6.50	Min.	H05	Heat Sink Area in Cage Width	12.50	Min.
H03	Heat Sink Area Width	13.00	Min.	H06	Heat Sink Area in Cage Length	30.70	Min.

Figure 122 QSFP+ heat sink dimensions (part 1 of 2)

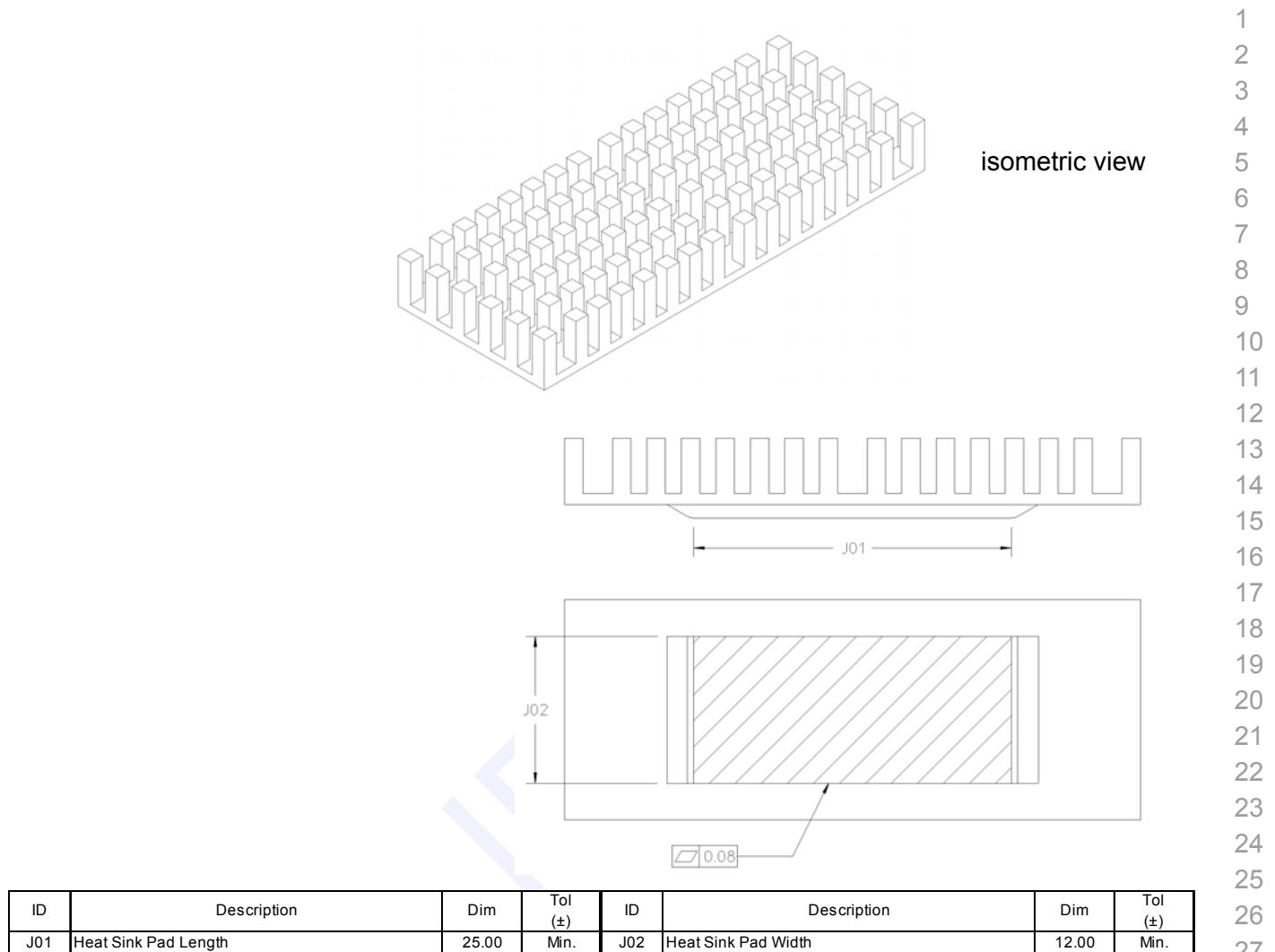
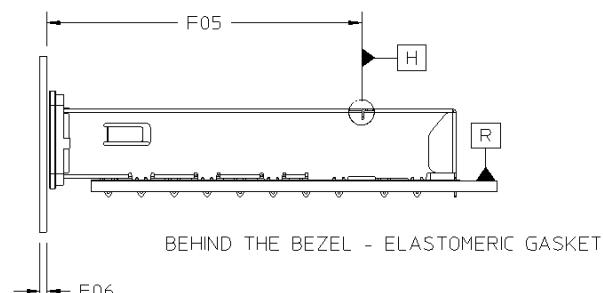
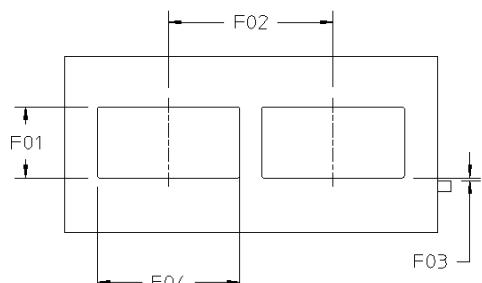
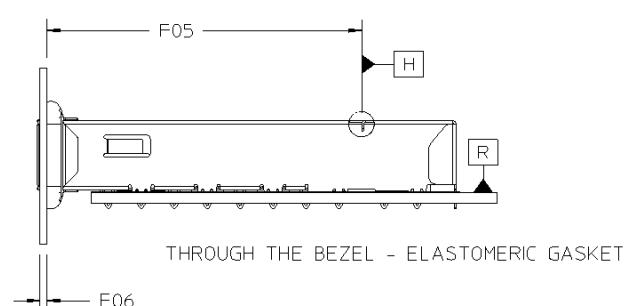
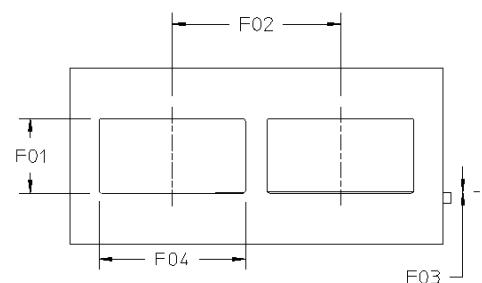
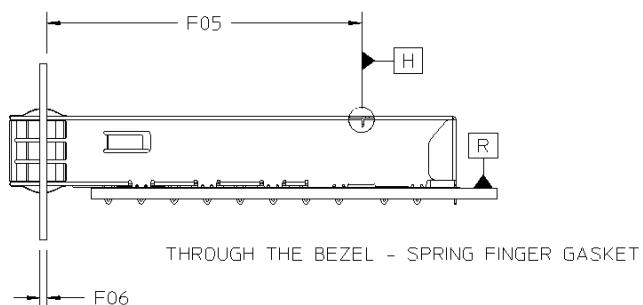
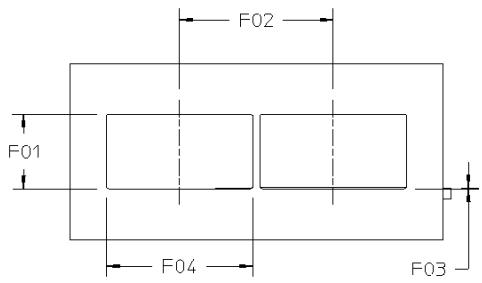


Figure 123 QSFP+ heat sink dimensions (part 2 of 2)

7.5.2.7 HOST BOARD BEZEL

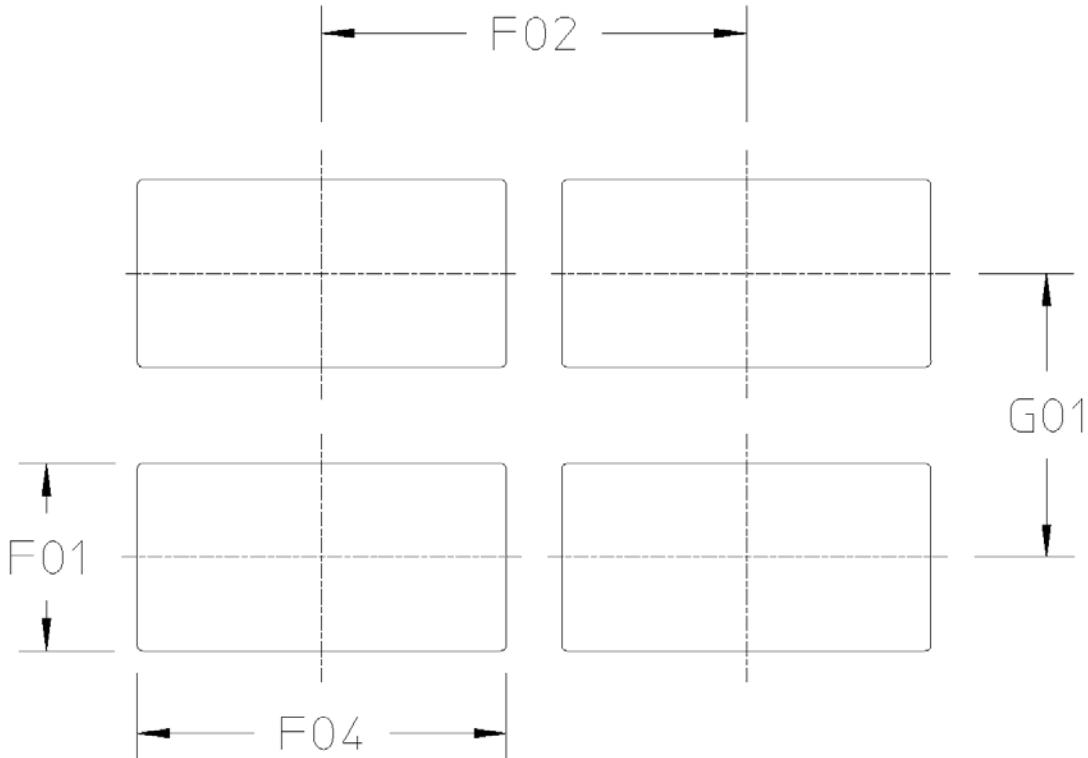
[Figure 124 on page 385](#) shows the QSFP+ cage to bezel dimensions for three different implementations (informative).



ID	Description	Dim	Tol (±)	ID	Description	Dim	Tol (±)
F01	Cutout Height - Through the Bezel - Finger	10.15	0.05	F04	Cutout Length - Through the Bezel - Finger	20.00	0.05
	Cutout Height - Through the Bezel - Elastomeric	10.15	0.05		Cutout Length - Through the Bezel - Elastomeric	20.00	0.05
	Cutout Height - Behind the Bezel	9.70	0.05		Cutout Length - Behind the Bezel	19.50	0.05
F02	Horizontal Pitch - Through the Bezel - Finger	21.00	Min.	F05	Datum H to Panel	43.00	0.30
	Horizontal Pitch - Through the Bezel - Elastomeric	23.00	Min.	F06	Panel Thickness	1.00	Ref.
	Horizontal Pitch - Behind the Bezel	22.50	Min.	-	-	-	-
F03	Cutout Location - Through the Bezel - Finger	0.15	0.10	-	-	-	-
	Cutout Location - Through the Bezel - Elastomeric	0.15	0.10	-	-	-	-
	Cutout Location - Behind the Bezel	0.30	0.20	-	-	-	-

Figure 124 QSFP+ cage to bezel dimensions

[Figure 125 on page 386](#) shows bezel opening dimensions for various implementations (informative).



ID	Description	Dim	Tol (±)	ID	Description	Dim	Tol (±)
G01	Vertical Pitch	15.40	Min.	-	-	-	-

Figure 125 QSFP+ bezel opening dimensions

7.5.2.8 MODULE PACKAGE

The dimensions for the QSFP+ pluggable module are shown in [Figure 126 on page 387](#) and [Figure 127 on page 388](#).

C7-10: 4x QSFP+ interface shall meet or exceed the physical and mechanical performance specifications of [Section 7.5.2.8](#). In cases of conflict with the SFF QSFP+ documentation, [Section 7.5.2.8](#) shall take precedence.

7.5.2.9 MODULE PADDLE CARD

The dimensions for the QSFP+ module paddle card contacts are shown in [Figure 128 on page 389](#).

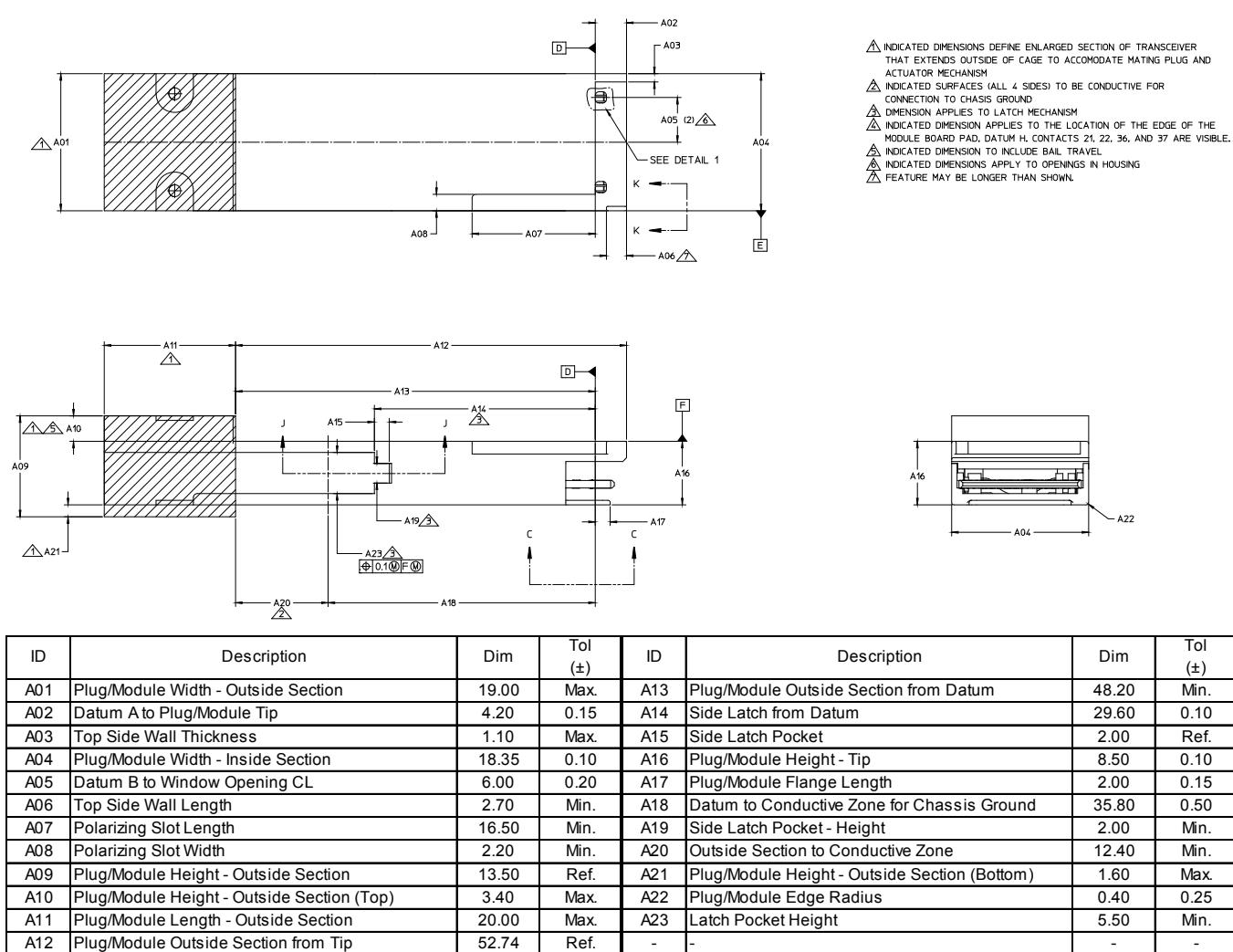
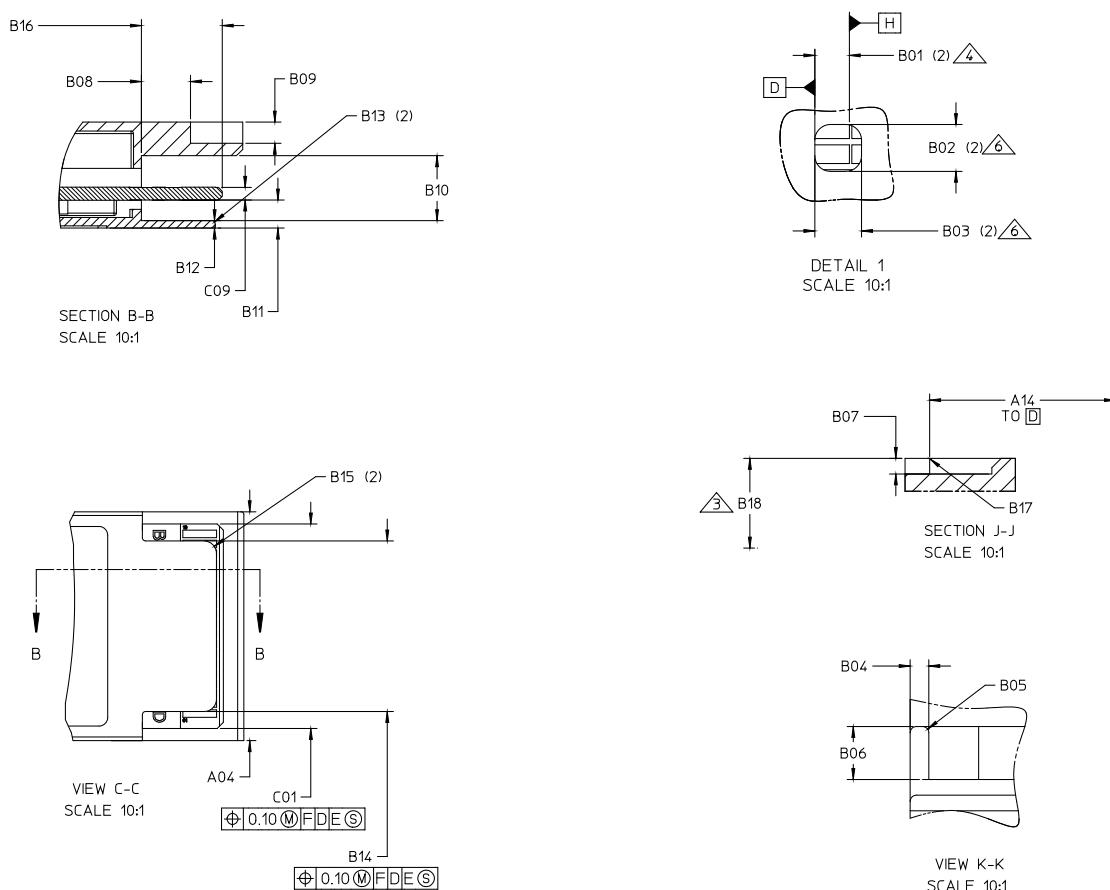
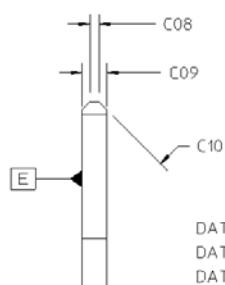
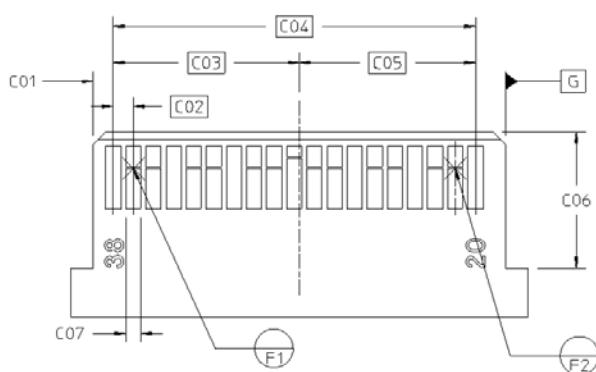


Figure 126 QSFP+ module dimensions (basic)



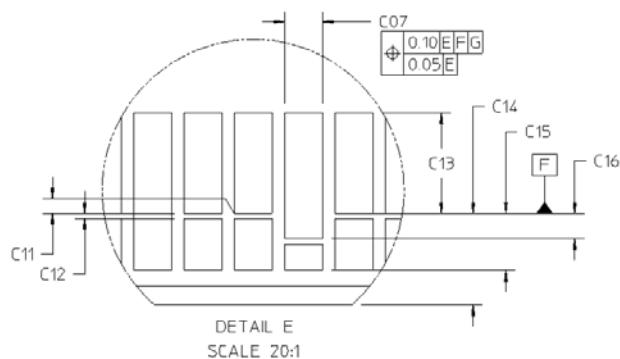
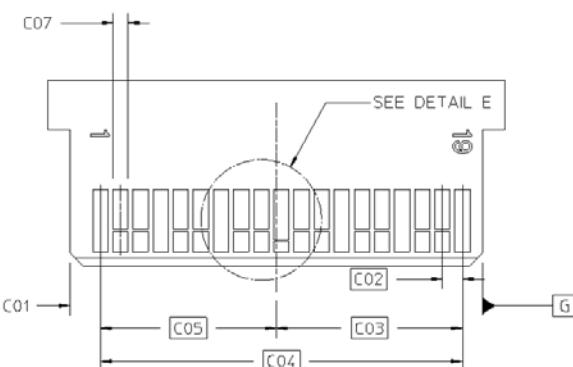
ID	Description	Dim	Tol (±)	ID	Description	Dim	Tol (±)
B01	Datum to Module PCB Pad Datum	1.10	0.18	B10	Inside Height of Plug Opening	5.20	0.15
B02	Window Width	1.50	Min.	B11	Datum F to PCB Bottom	2.25	0.10
B03	Window Length	1.50	Min.	B12	Tongue Thickness	0.60	0.05
B04	Top Side Wall Width	0.80	0.25	B13	Tongue Radius	0.30	Min.
B05	Top Side Wall Radius	0.60	Max.	B14	Tongue Width	13.68	0.10
B06	Top Side Wall Height	1.70	0.10	B15	Tongue Radius or Chamfer	0.30	Min.
B07	Latch Pocket Depth	0.50	Min.	B16	Component Free Area - Both Sides	6.50	Min.
B08	Datum B to Base of Tongue	3.60	Min.	B17	Latch Radius	0.05	Min.
B09	Height of Datum B	1.70	0.10	B18	Width to Opposite Latch Surface	18.45	Max.

Figure 127 QSFP+ module dimensions (detail)



DATUM G - CENTERLINE OF PADDLE CARD
 DATUM E - TOP SURFACE OF PADDLE CARD
 DATUM F - LEADING EDGE OF SHORT PADS
 DEFINED BY OUTER SHORT PADS

NO SOLDER MASK WITHIN 0.05 OF DEFINED PAD LOCATIONS



ID	Description	Dim	Tol (±)	ID	Description	Dim	Tol (±)
C01	Paddle Card Width (Pad contact width 0.54)	16.42	0.08	C08	Lead-in Flat	0.36	Ref.
	Paddle Card Width (Pad contact width 0.60)	16.40	0.10	C09	Paddle Card Thickness	1.00	0.10
C02	Pad Center to Center (Pitch)	0.80	Basic	C10	Lead-in Chamfer x 45 degrees	0.30	0.05
C03	Overall Pad Centers	7.40	Basic	C11	Short Pad to Datum C	0.00	0.03
C04	Card Center to outer Pad, Side B and Side D	14.40	Basic	C12	Pad to Pre-Pad	0.10	0.05
C05	Card Center to outer Pad, Side A and Side C	7.00	Basic	C13	Pad Length	1.60	Min.
C06	Component Keep Out Area	5.40	Min.	C14	Card Edge to Second Pad	1.45	0.10
C07	Pad Width	0.54	0.04	C15	Front Pad Length	0.90	0.05
	Pad Width (SFF-8436)	0.60	0.03	C16	Front Pad Spacing	0.40	0.05

Figure 128 QSFP+ paddle card dimensions

7.5.3 ELECTRICAL REQUIREMENTS

7.5.3.1 MATED CONNECTOR ELECTRICAL REQUIREMENTS

The QSFP/QSFP+ module and receptacle shall comply to the electrical specifications described in [Table 98](#).

Table 98 QSFP/QSFP+ connector electrical performance requirements

Symbol	Parameter	Min	Max	Unit	Comments
LLCR	Low level contact resistance - initial		80	mΩ	through testing per EIA-364-23, measured across interface between paddle card trace and receptacle
ΔLLCR	Low level contact resistance - change		20	mΩ	through testing per EIA-364-23, as a result of any test group setup
I _{max}	Current rating, all contacts simultaneously	0.5		A	per EIA-364-70 or IEC 512-5-1 Test 5a, at 30°C temperature rise above ambient
I _{max,s}	Current rating, single contact	1		A	per EIA-364-70 or IEC 512-5-1 Test 5a, at 30°C temperature rise above ambient
R _{Iso}	Insulation Resistance	1000		MΩ	100 Vdc, between adjacent contacts
V _{Iso}	Dielectric Withstanding Voltage	300		Vdc	No defect or breakdown between adjacent contacts, 300 Vdc minimum for 1 minute
Z _{dco(peak)}	Differential Impedance - peak (connector area)	90	110	Ω	EIA-364-108 Rise/Fall time: 100 ps (20-80%). See Table 69 .
Z _{dco(nom)}	Differential Impedance (nominal)	95	105	Ω	Includes connector cable to connector interface and board termination pads and vias. For QDR and higher speeds, connector return loss specification shall supersede this specification.
S _{cop}	Within-Pair Skew		5	ps	maximum (by design), measured at interface between paddle cards & receptacle. EIA-364-103
NEXT _c	Near End Crosstalk Isolation		-34	dB	EIA-364-90, 50 MHz to 10 GHz. Equivalent to 2% voltage crosstalk, power sum
L _{co}	Insertion Loss		1.0	dB	EIA-364-101, 50 MHz to 5 GHz

7.5.3.2 COMPLIANCE BOARDS

Design requirements for compliance boards may be found in [Annex 1: FDR and EDR Compliance Boards and Test Setups on page 614](#), particularly [Annex 1.3 Compliance Boards - Electrical Specifications on page 618](#). Test setups for devices and cables may be found [Annex 1.2 Test Setups on page 614](#).

7.5.3.3 LOW-SPEED SIGNALING CONTACTS

7.5.3.3.1 SDA, SCL

SCL is the clock of the two-wire serial interface, and SDA is the data for the 2-wire serial interface. Operation of this interface is described in detail in [Chapter 8: Management Interface](#), SCL and SDA must be pulled up in the host, through an pull-up resistor of value appropriate to the overall bus capacitance and the rise and fall time requirements listed

in [Table 133 on page 482](#).

7.5.3.3.2 ModSelL

The ModSelL is an input contact. When held low by the host, the module responds to 2-wire serial communication commands. The ModSelL allows the use of multiple QSFP+ modules on a single 2-wire interface bus. When the ModSelL is “High”, the module shall not respond to or acknowledge any 2-wire interface communication from the host. ModSelL signal input node must be biased to the “High” state in the module.

In order to avoid conflicts, the host system shall not attempt 2-wire interface communications within the ModSelL de-assert time after any modules are deselected. Similarly, the host must wait at least for the period of the ModSelL assert time before communicating with the newly selected module. The assertion and de-asserting periods of different modules may overlap as long as the above timing requirements are met.

7.5.3.3.3 ResetL

The ResetL contact shall be pulled to Vcc in the module. A low level on the ResetL contact for longer than the minimum pulse length (t_{Reset_init}) initiates a complete module reset, returning all user module settings to their default state. Module Reset Assert Time (t_{init}) starts on the rising edge after the low level on the ResetL contact is released. During the execution of a reset (t_{init}) the host shall disregard all status bits until the module indicates a completion of the reset interrupt. The module indicates this by asserting “low” an IntL signal with the Data_Not_Ready bit negated. Note that on power up (including hot insertion) the module should post this completion of reset interrupt without requiring a reset.

7.5.3.3.4 LPMode

The LPMode contact shall be pulled up to Vcc in the module. This function is affected by the LPMode contact and the combination of the Power_override and Power_set software control bits (Address A0h, byte 93 bits 0,1).

The module has two modes a low power mode and a high power mode. The high power mode operates in one of the four power classes.

When the module is in a low power mode it has a maximum power consumption of 1.5W. This protects hosts that are not capable of cooling higher power modules, should such modules be accidentally inserted.

The module's 2-wire serial interface and all laser safety functions shall be fully operational in this low power mode. The module shall still support the completion of reset interrupt in this low power mode.

If the Extended Identifier bits (Page 00h, byte 129 bits 6-7) indicate a power consumption greater than 1.5W and the module is in low power mode it shall reduce its power consumption to less than 1.5W while still maintaining the functionality above. The exact method of accomplishing low power is not specified, however it is likely that either the Tx or Rx or both will not be operational in this state.

If the Extended Identifier bits (Page 00h, byte 129 bits 6-7) indicate that its power consumption is less than 1.5W then the module shall be fully functional independent of whether it is in low power or high power mode.

The module should be in low power mode if the LPMode contact is in the high state, or if the Power_override bit is in the high state and the Power_set bit is also high. The module should be in high power mode if the LPMode contact is in the low state, or the Power_override bit is high and the Power_set bit is low. Note that the default state for the Power_override bit is low.

At Power up, the Power_override and Power_set bits shall be cleared to 0.

Table 99 Power Mode Truth Table

Power_override Byte 93 bit 0	LPMode Contact	Power_set to LPMode Byte 93 bit 1	High Power Class Byte 93 bit 2	Allowed Module Power ^a
0	1	X	X	<1.5 Watts
	0	X	X	<3.5 Watts
1	X	1	X	<1.5 Watts
		0	0	<3.5 Watts
			1	As designated in RO "Device Type" Power Class bits - Upper Page 0, Byte 129, bits 7-6, 1-0

a. Note that if the Allowed Module Power indicated in this column restricts its current power usage to less than the Power Class of the module, then the module may not be fully operational in either Tx or Rx or both, but the 2-wire serial interface and all laser safety functions shall be fully operational, and the module shall still support the completion of reset interrupt.

7.5.3.3.5 ModPrsL

ModPrsL is pulled up to Vcc_Host on the host board and grounded in the module. The ModPrsL is asserted “Low” when inserted and deasserted “High” when the module is physically absent from the host connector.

7.5.3.3.6 INTL

IntL is an output contact. When “Low”, it indicates a possible module operational fault or a status critical to the host system. The host identifies the source of the interrupt using the 2-wire serial interface. The IntL contact is an open collector output and shall be pulled to host supply voltage on the host board. The INTL contact is deasserted “High” after completion of reset, when byte 2 bit 0 (Data Not Ready) is read with a value of ‘0’ and the flag field is read (see SFF-8636).

7.5.3.4 HIGH-SPEED SIGNALING CONTACTS

With the exception of squelch behavior, the high-speed signaling for Tx and Rx (pluggable module data inputs from and outputs to the InfiniBand port, respectively) are as described in [Section 6.8, “Compliant Channels,” on page 323](#) for the relevant speeds. Refer

to the appropriate speed-dependent sub-sections of this section (QDR, FDR, EDR, etc.)
for interface specifications.

For QSFP+ modules, Rx Squelch, for loss of input signal from the link (Rx LOS), is required. In the event that the signal on any channel becomes equal to or less than the level required to assert Rx LOS (Receiver Loss of Signal), the receiver data output for that channel shall be squelched or disabled. In the squelched or disabled state, output impedance levels shall be maintained while the differential voltage swing shall be less than 50 mVpp. Rx squelch may optionally be disabled.

For QSFP+ modules, Tx Squelch, for loss of input signal from the InfiniBand port, is an optional function. Where implemented it shall function as follows. In the event of the differential, peak-to-peak electrical signal on any channel becomes equal to or less than 50 mVpp, then the transmitter optical output for that channel shall be squelched or disabled and the associated Tx LOS flag set.

In InfiniBand links, the use of Rx Squelch and Tx Squelch may interfere with the correct operation of Beacon signaling. As described in [Volume 2, Release 1.2.1, Section 5.6.4.2 Polling States and Section 5.6.4.3 Sleeping States](#), the beaconing sequence is a periodic repeating pattern, with an active TS1 transmission period of 2 ms and a quiescent period of 100 ms. Since Tx and Rx Squelch, if enabled, will activate during the 100 ms quiescent period, the time from resumption of Rx input and Tx output signals until normal signaling condition is reached must be well under 2 ms.

To meet this condition, the Rx Squelch Deassert Time (Time from resumption of Rx input signals until normal Rx output condition is reached) shall be less than 80 µs (as specified in the QSFP+ SFF specification). The Tx Squelch Deassert time, if Tx Squelch is enabled, shall be less than 500 µs. The Rx Squelch Assert Time and Tx Squelch Assert time shall be less than 80 µs and 10 ms, respectively. These specifications are listed in [Table 136, "I/O Timing for Squelch and Disable," on page 485](#). (Note that the QSFP+ SFF timing specifications are less stringent, specifying, for example a much longer Tx Squelch Deassert time, 400 ms, so Tx Squelch may generally be disabled).

7.5.3.4.1 TRANSMIT SIGNALS: Tx[0-3][P/N]

Tx[0-3][p/n] are QSFP+ module transmitter data inputs. They are AC-coupled differential lines with 100 Ohm differential terminations inside the module. The AC coupling is inside the QSFP/QSFP+ module and not required on the Host board.

Output squelch (Tx Squelch), for loss of input signal, (Tx LOS), is an optional function. Where implemented it shall function as follows. In the event of the differential, peak-to-peak electrical signal on any lane becomes equal to or less than 50 mVpp, then the transmitter optical output for that lane shall be squelched and the associated Tx LOS flag set. For an optical transceiver with separable optical connector, the optical modulation amplitude (OMA) squelched shall be less than or equal to -26 dBm, and, where practical, the average output power is recommended to be less than or equal to -26 dBm.

In normal operation, where Tx Squelch is implemented, the default case has Tx Squelch active. Tx Squelch can be deactivated using Tx Squelch Disable through the 2-wire serial interface. Tx Squelch Disable is an optional function.

Implementation Note

Note that Tx Squelch can be implemented in several different ways: (1) No modulation (OMA less than or equal to a limit, but Tx average power within normal operating range, either at '0' power or at average of '0' and '1' power), or (2) No optical output power (Tx average power less than or equal to a limit). Method (2) is clearly more restrictive, since an output power limit also implies an OMA limit.

Revisions of this specification prior to 1.3.1 have only specified the limit on OMA. Revision 1.3.1 has added the recommended limit on Tx average power, since it is a more definitive way of indicating Tx Squelch, especially for QSFP optical transceiver applications beyond the InfiniBand link protocol.

Note that optical transceivers built to Rev. 1.3 specifications may only limit OMA, rather than limiting average power also. Also, for some optical transceiver technologies, such as silicon photonics, it may be impractical or impossible to limit average optical power to -26 dBm (particularly on individual lanes).

Applications using optical transceivers should confirm Tx Squelch implementation method with selected vendors to assure interoperability.

7.5.3.4.2 RECEIVE SIGNALS: Rx[0-3][P/N]

Rx[0-3][p/n] are QSFP+ module receiver data outputs. They are AC-coupled differential lines that should be terminated with 100 Ohm differential at the Host ASIC or SERDES. The AC coupling is inside the QSFP/QSFP+ module and not required on the Host board.

Output squelch for loss of input signal (Rx Squelch), is an optional function. Where implemented it shall function as follows. In the event of the optical or electrical signal on any physical lane becoming equal to or less than the level required to assert loss of signals (Rx LOS), then the receiver data output for that lane shall be squelched or disabled. In the squelched or disabled state, output impedance levels are maintained while the differential voltage swing shall be less than 50 mVpp. This limit on the output voltage swing provides margin vs. the minimum specified host receiver sensitivity — V_{RSD} in [Table 59](#), [“Host Receiver Characteristics for 2.5 Gb/s \(SDR\).” on page 311](#), or V_{Csense} in [Table 61](#), [“Host Receiver Characteristics for 10 Gb/s \(QDR\).” on page 315](#), or Y1, Y2 minimum values in [Table 65](#), [“FDR receiver stressed input signal specifications - limiting cables.” on page 319](#) and [Table 68](#), [“EDR receiver stressed input signal specifications - limiting cables.” on page 322](#).

In normal operation, where Rx Squelch is implemented, the default case has Rx Squelch active. Rx Squelch can be deactivated using Rx Squelch Disable through the 2-wire serial interface. Rx Squelch Disable is an optional function.

Implementation Note

Given that there are several possible implementations of Tx Squelch, applications using optical transceivers should confirm interoperability Tx Squelch and Rx Squelch implementations. See note above.

7.5.3.5 HOST BOARD SCHEMATIC

[Figure 129 on page 396](#) shows an example of a host board schematic for QSFP+, with connections to host SERDES and control logic. For EMI protection the signals to the connector should be turned off when the QSFP+ transceiver is removed. The use of good high speed printed circuit board design practices is recommended. The chassis ground (case common) of the QSFP+ module should not be directly connected to the module's signal ground.

Note that DC blocking capacitors on the high-speed signal wires are implemented inside the module or cable as described in [Section 6.8.1, "DC Blocking," on page 323](#).

7.5.3.6 POWER REQUIREMENTS

Power for the module is supplied through 2 contacts: VccRx and VccTx. Power is applied concurrently to them.

Since different classes of modules exist with pre-defined maximum power consumption limits, it is necessary to avoid exceeding the system power supply limits and cooling capacity when a module is inserted into a system designed to only accommodate lower power modules. See [Section 7.5.3.9 on page 398](#). It is recommended that the host, through the management interface, identify the power consumption class of the module before allowing the module to go into high power mode.

A host board together with the QSFP+ module(s) forms an integrated power system. The host supplies stable power to the module. The module limits electrical noise coupled back into the host system and limits inrush charge/current during hot plug insertion.

All specifications shall be met at the maximum power supply current. No power sequencing of the power supply is required of the host system since the module sequences the contacts in the order of ground, supply and signals during insertion.

7.5.3.7 HOST POWER SUPPLY FILTERING

The host board should use the power supply filtering network shown in [Figure 130 on page 397](#), or an equivalent. Any voltage drop across a filter network on the host is counted against the host DC set point accuracy specification. Inductors with DC Resistance of less than 0.1Ω should be used in order to maintain the required voltage at the Host Edge Card Connector. Selection of the filter capacitors (shown as 22 uF in [Figure 130 on page 397](#)) is left to the implementer, but it is recommended that 22 uF capacitors

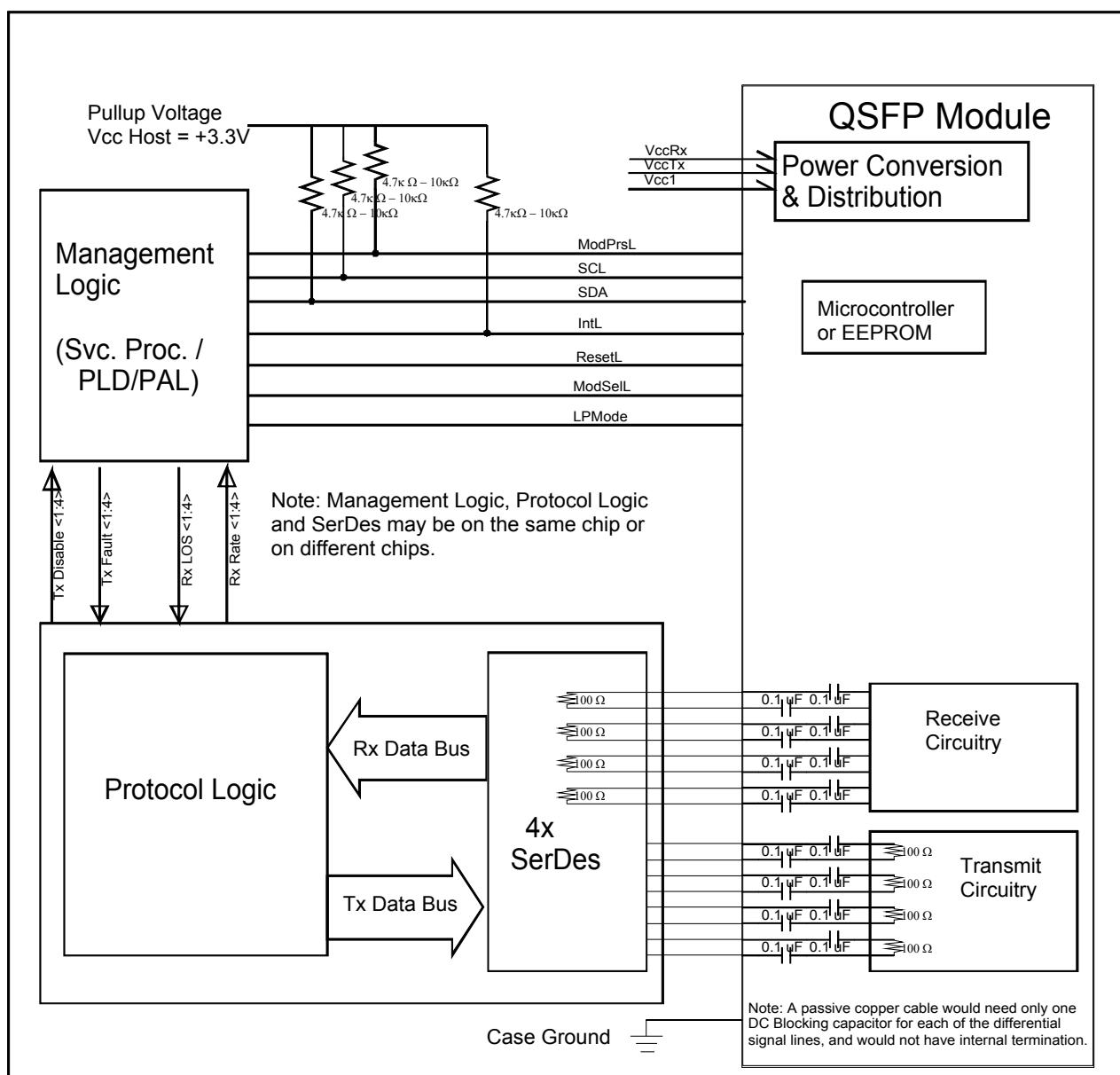


Figure 129 Example QSFP Host board schematic

with approximately 0.22Ω of equivalent series resistance be used to provide adequate filtering without high frequency ringing. The time constant of the filter circuit is the important quantity.

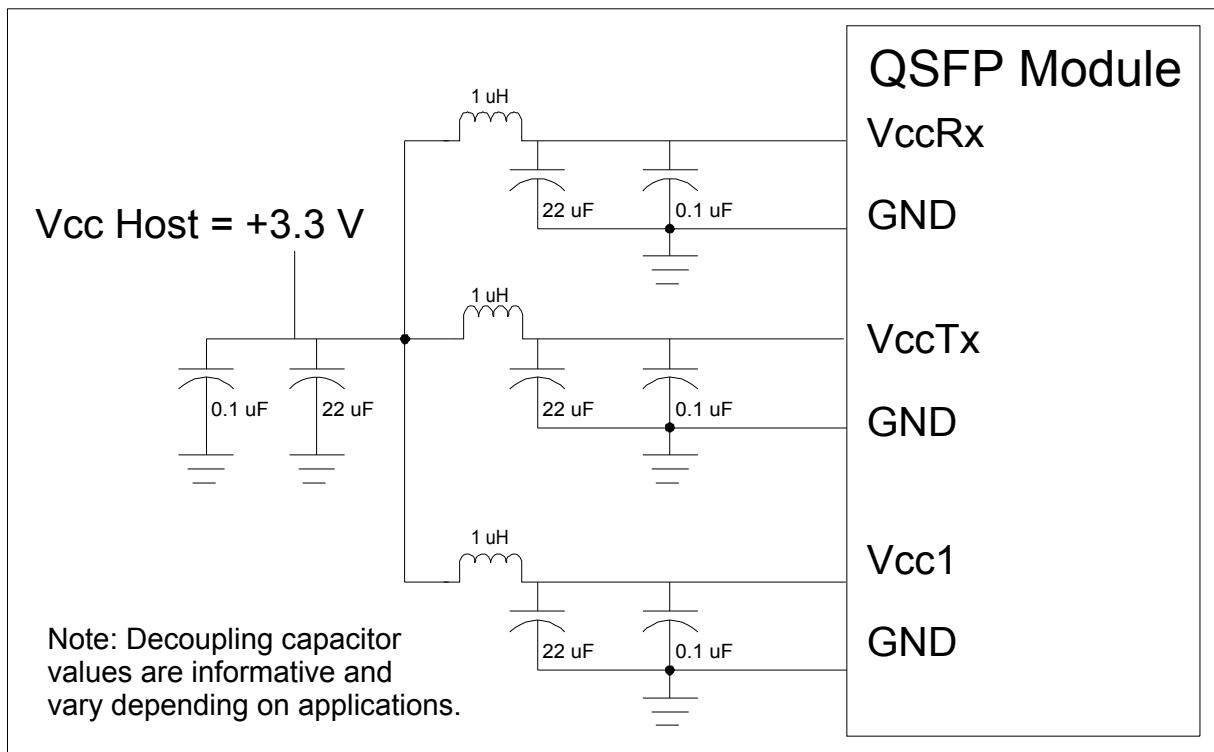


Figure 130 Recommended QSFP Host Board Power Supply Filtering

7.5.3.8 HOST POWER SUPPLY SPECIFICATIONS

The specifications for the power supply are listed in [Table 100](#).

Table 100 Power Supply Specification

Parameter	Min.	Nominal	Max.	Unit	Condition
Vcc_host		3.3		V	Measured at Vcc-Tx and Vcc-Rx
Vcc set point accuracy	-5		+5	%	Measured at Vcc-Tx and Vcc-Rx. Note ^a
Vcc Power supply noise including ripple			50	mVpp	1 kHz to frequency of operation
Module Inrush Current, LPMode asserted			600	mA	each inductor in the power supply filter
Module sustained peak current, LPMode de-asserted			750	mA	each inductor in the power supply filter Note ^b
Module Inrush Current, LPMode de-asserted			900	mA	

a. 5%-accurate power needed for VCSEL laser drivers

b. This limit may be exceeded for up to 50 us. See Figure 5 of SFF-8436.

7.5.3.9 POWER BUDGET CLASSES

Power levels associated with classifications of modules are shown in [Table 101 on page 398](#). In general, the higher power classification levels are associated with higher data rates and longer reach, for a particular technology family.

Table 101 Power Budget Classification

Power Class	Max Power (W)
1	1.5 or less
2	2.0 or less
3	2.5 or less
4	3.5 or less

Power Class 1 supports a management-interface-only power level, for modules or cables such as passive copper cables which require little or no signal power.

Power Classes 2 through 4 describe modules or cables with between 1.5W and 3.5 Watts per end.

The highest maximum power budget is determined by a current limit of 500 mA for each power contact, and by the cooling capability provided by the system. Two power contacts allow power supply of up to 3.3W of power at 3.3V. Generally, cooling capability will limit the amount of power that a module may dissipate. The system designer is responsible for ensuring that the maximum temperature does not exceed the case temperature requirements.

7.5.4 ENVIRONMENTAL AND THERMAL REQUIREMENTS

7.5.4.1 ESD REQUIREMENTS

The module high speed signal contacts shall withstand 1000 V electrostatic discharge using the Human Body Model module and all other contacts shall withstand 2000 V electrostatic discharge using the Human Body Model, per JEDEC Standard JESD22-A114B (March 2006). The module power and ground contacts shall withstand 500 V using the charged device model, per JEDEC Standard JESD22-C101E (Dec. 2009). High speed and management interface contacts do not have charged device model ESD requirements, since they are recessed. The module shall meet ESD requirements given in EN 61000-4-2, criterion B test specification, such that when installed in a properly grounded housing and chassis the module is subjected to 15 kV air discharges during operation and 8 kV direct contact discharges to the case. For all three tests, the module shall withstand these discharge levels without damage.

7.5.4.2 HOT INSERTION AND REMOVAL

Hot insertion and removal requirements for QSFP+ modules and cables are defined in [7.3.4.2 on page 370](#). QSFP+ modules shall not be damaged by removal or insertion while power is applied. Removal may occur while the link is operating without damage to either the port or the link.

7.5.5 MEMORY MAP

The memory map for QSFP+ devices is defined in [Chapter 8: on page 480](#) based on SFF-8436-2011-02-03 rev. 4.2 or a replacement document.

7.6 OPTICAL 4x QSFP+ MODULES

C7-11: 4x Optical QSFP+ modules that have separable optical connectors shall comply with requirements in their respective 4x optical distance classification (SX or LX) in the *InfiniBand Architecture Specification, Volume 2, Chapter 9: Fiber Attachment - 2.5 Gb/s, 5.0 Gb/s, & 10 Gb/s* at the de-mating points.

Implementation Note

Note that while optical pluggable modules with separable interfaces must comply with standard optical link requirements as described in the architecture specification, there is no such requirement on pluggable electrical devices or on optical active cables that interface with the pluggable device interface, which may use any appropriate signaling format within the active cable.

7.7 8X AND 12XMICROGIGACN INTERFACE

7.7.1 INTRODUCTION

This section defines one connector for the 12X and 8X cable interface on InfiniBand boards. The 12X interface is the widest InfiniBand link, providing for simultaneous transmit and receive of twelve bits of encoded differential data. This connector is primarily used at SDR and DDR data rates.

This connector uses pairs of contacts separated by Ground contacts to reduce near end crosstalk (NEXT). The connector is a surface-mounted design similar to that of the 4X connector described in [Section 7.3 on page 355](#), and all signal contacts are used. A suitable receptacle, referred to as a 24 pair microGigaCN receptacle, is available from Fujitsu Components Ltd. and others, and is shown in [Figure 131 on page 400](#). Detailed drawings of mating interface dimensions are shown in [Figure 111 on page 364](#).

The 8X cable interface uses the same board connector as the 12X interface, but with a different contact assignment.

C7-12: All 12X InfiniBand cable plugs that are not pluggable devices shall be intermateable with this connector.

C7-12.2.1: All 8X InfiniBand cable plugs that are not pluggable devices shall be intermateable with this connector.

7.7.2 MECHANICAL REQUIREMENTS

7.7.2.1 PHYSICAL AND MECHANICAL PERFORMANCE REQUIREMENTS

The physical and mechanical performance requirements for the 8X and 12X microGigaCN connectors are defined in [Section 7.3.1.1 on page 355](#).

7.7.2.2 HOST BOARD CONNECTOR

This section defines the host board connector for the 8X and 12X microGigaCN interface. An exemplary connector is shown in [Figure 131](#).

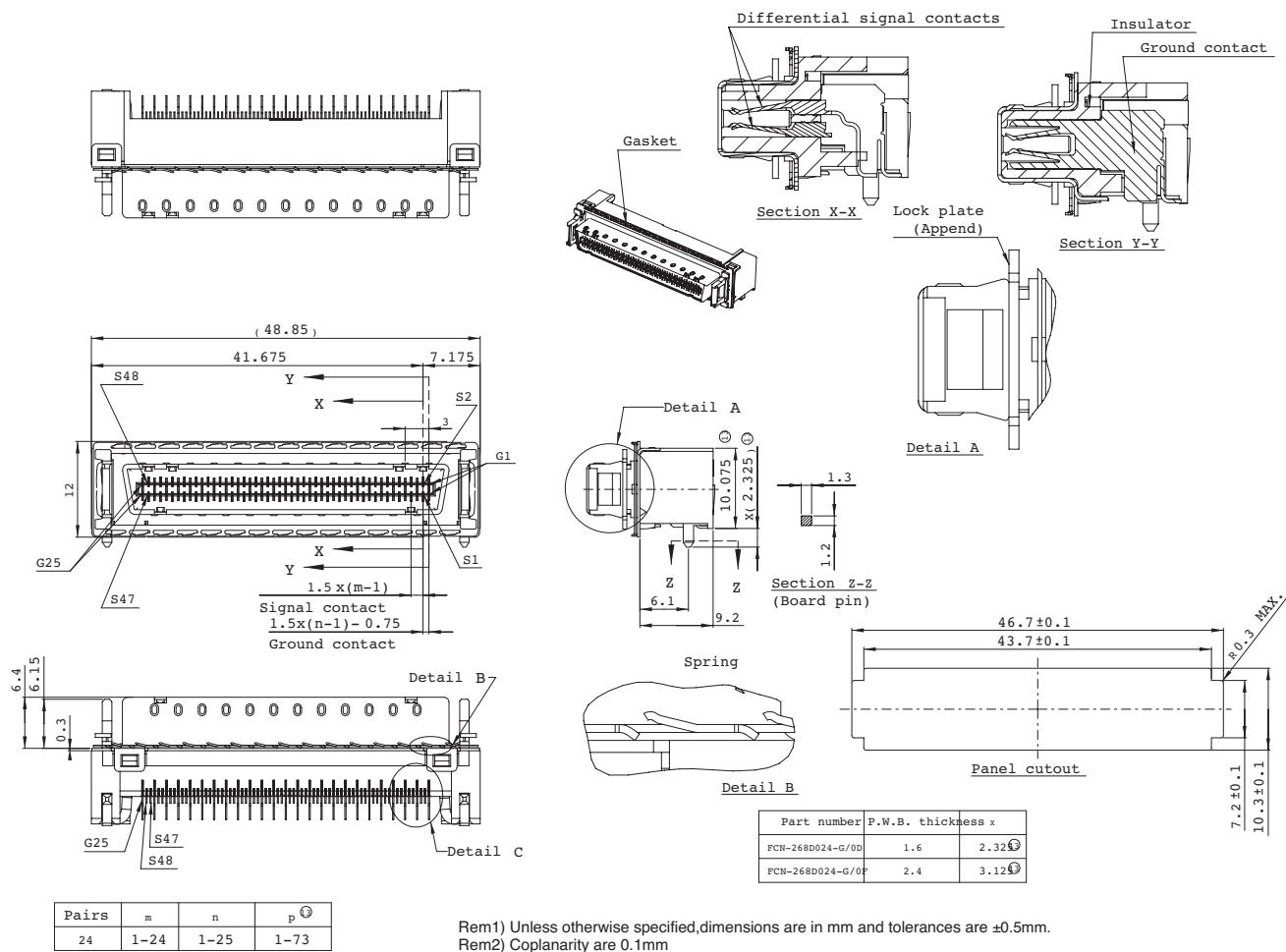


Figure 131 12X cable board connector

The board footprint shown in [Figure 132 on page 401](#) should be used for the 12X cable board connector shown in [Figure 131](#). Connector contact numbers are indicated for information. The connector is surface-mounted, so there are no contacts on the secondary side of the board.

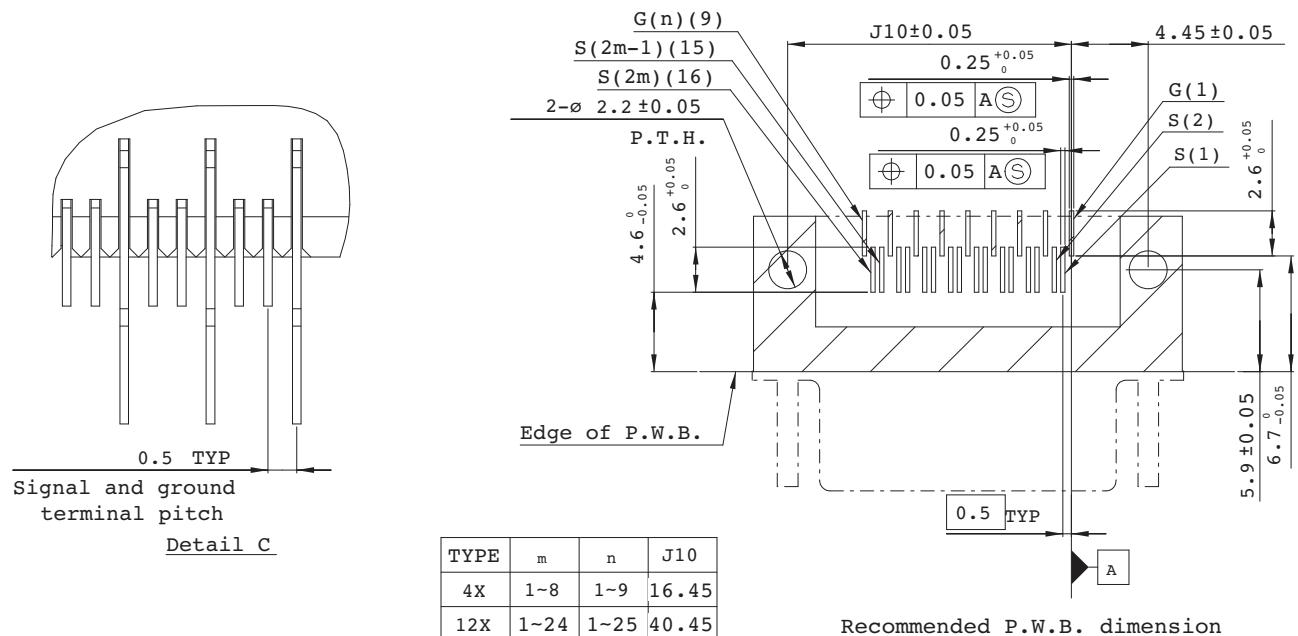


Figure 132 12X cable board connector footprint, top view

7.7.2.3 CABLE PLUG

This cable plug uses a similar design to that of the 4X microGigaCN cable plug. The plug uses pairs of signal contacts interspersed with Ground contacts for crosstalk reduction. Side latches are used in conjunction with receptacle features for retention, and are released by pulling on the “lanyard” handle shown in the drawing. A suitable plug, referred to as a 24 pair microGigaCN plug, is available from Fujitsu Components Ltd. and others, and is shown in [Figure 133 on page 402](#). Detailed drawings of mating interface dimensions are shown in [Figure 111 on page 364](#).

The cable plug may be designed using a dual bulk wire exit to accommodate the large bulk wire diameter while reducing bend radius. In that case, each of the two bulk cables would contain twelve signal pairs of wire.

C7-13: 12X Board receptacles used on InfiniBand modules that are not pluggable devices shall be intermateable with this connector.

C7-13.1.1: A metal backshell or other means which fully shields the connector shall be bonded to the cable bulk shield through a continuous 360 degree contact to minimize EMI (Electromagnetic Interference).

Twenty-four pairs of signals are used, twelve each for transmit and receive.

The same cable plug is used for both passive and active cables but with different contact assignments. Those contact assignments are defined in Sections [7.7.3.2 on page 403](#) through [7.7.3.6 on page 410](#).

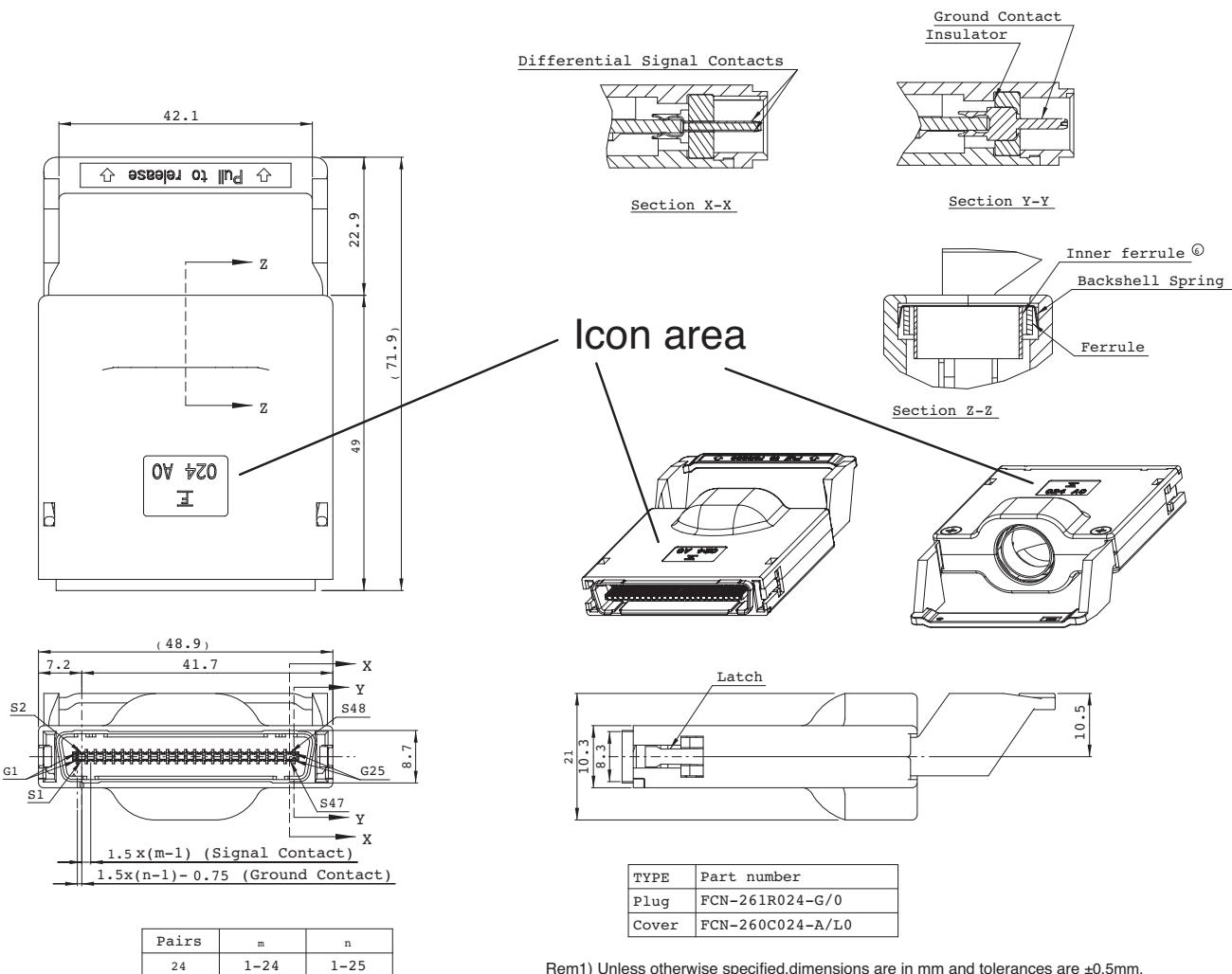
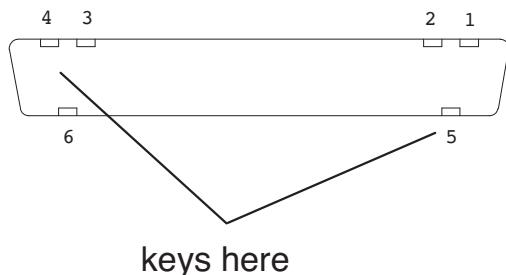


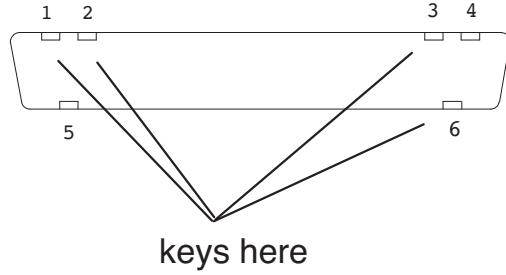
Figure 133 12X cable plug

7.7.2.4 KEYING

It is recommended that the 12X board connectors as used for InfiniBand use keys in positions 4 and 5, shown in [Figure 134 on page 403](#) as viewed from the outside of the chassis. This is to prevent misplugging with other interfaces that might have chosen to use the same physical connector.



(a) board connector



(b) cable connector

Figure 134 12X board and cable connector keying

7.7.3 ELECTRICAL REQUIREMENTS

7.7.3.1 MATED CONNECTOR ELECTRICAL PARAMETERS

The electrical performance requirements for the 8X and 12X microGigaCN connectors are defined in [Section 7.3.3.1 on page 365](#).

7.7.3.2 PIN ASSIGNMENT - 12X PASSIVE CABLE PORTS

C7-14: The contact assignment listed in [Table 102](#) shall be used for the board connector for passive InfiniBand 12X cables.

The character 'x' in the signal symbol is the port number, as defined in [Section 4.1, "Signal Naming Conventions," on page 71](#).

Table 102 12X passive board connector signal assignment

Pin Number	Signal	Pin Number	Signal
G1-G25	Signal Ground	Housing	Chassis Ground
S1	IBtxlp(0)	S25	IBtxOn(11)
S2	IBtxIn(0)	S26	IBtxOp(11)
S3	IBtxlp(1)	S27	IBtxOn(10)
S4	IBtxIn(1)	S28	IBtxOp(10)
S5	IBtxlp(2)	S29	IBtxOn(9)
S6	IBtxIn(2)	S30	IBtxOp(9)
S7	IBtxlp(3)	S31	IBtxOn(8)
S8	IBtxIn(3)	S32	IBtxOp(8)
S9	IBtxlp(4)	S33	IBtxOn(7)
S10	IBtxIn(4)	S34	IBtxOp(7)
S11	IBtxlp(5)	S35	IBtxOn(6)
S12	IBtxIn(5)	S36	IBtxOp(6)
S13	IBtxlp(6)	S37	IBtxOn(5)
S14	IBtxIn(6)	S38	IBtxOp(5)
S15	IBtxlp(7)	S39	IBtxOn(4)
S16	IBtxIn(7)	S40	IBtxOp(4)
S17	IBtxlp(8)	S41	IBtxOn(3)
S18	IBtxIn(8)	S42	IBtxOp(3)
S19	IBtxlp(9)	S43	IBtxOn(2)
S20	IBtxIn(9)	S44	IBtxOp(2)
S21	IBtxlp(10)	S45	IBtxOn(1)
S22	IBtxIn(10)	S46	IBtxOp(1)
S23	IBtxlp(11)	S47	IBtxOn(0)
S24	IBtxIn(11)	S48	IBtxOp(0)

C7-15: Signal Ground shall be connected to **IB_Sh_Ret** on the module.

C7-16: Signal Ground shall not be connected to Chassis Ground in the connector. This is for meeting EMI and ESD requirements. For discussion, see [Volume 2-DEPR, Section 1.5.4](#).

C7-16.1.1: A continuous ground path from the cable's inner shield(s) through the connector to the board signal ground shall be provided to insure low jitter, low crosstalk and EMI containment.

A cable constructed with only a bulk shield is not likely to meet the electrical requirements of [7.9.2.3 on page 449](#).

7.7.3.3 PIN ASSIGNMENT - 12X ACTIVE CABLE PORTS

C7-16.2.2: The contact assignment listed in [Table 103 on page 406](#) shall be used for the board connector for active InfiniBand 12X cables. Usage requirements including current limitations on the power contacts are described in [Section 7.7.3.7, "Active cable power requirements." on page 413](#). Power return is by way of the Signal Ground contacts.

The 12 V or 3.3 V Sense signal is used to enable the respective voltage on the Vcc power supply contact used to provide power to the active components in the cable. Both signals are considered active if their voltage level is between 0.9 V and 2.4 V DC.

C7-16.1.3: Both 12 V and 3.3 V Sense signals shall not be active simultaneously.

The character 'x' in the signal symbol is the port number, as defined in [Section 4.1, "Signal Naming Conventions." on page 71](#).

Table 103 12X active board connector signal assignment

Pin Number	Signal	Pin Number	Signal
G1	Sense-12V	Housing	Chassis Ground
G2, G4-G10, G12-G22, G25	Signal Ground		
S1	IBtxIp(0)	S25	IBtxOn(11)
S2	IBtxIn(0)	S26	IBtxOp(11)
S3	IBtxIp(1)	S27	IBtxOn(10)
S4	IBtxIn(1)	S28	IBtxOp(10)
G3	Vcc	S29	IBtxOn(9)
S5	IBtxIp(2)	S30	IBtxOp(9)
S6	IBtxIn(2)	S31	IBtxOn(8)
S7	IBtxIp(3)	S32	IBtxOp(8)
S8	IBtxIn(3)	S33	IBtxOn(7)
S9	IBtxIp(4)	S34	IBtxOp(7)
S10	IBtxIn(4)	S35	IBtxOn(6)
S11	IBtxIp(5)	S36	IBtxOp(6)
S12	IBtxIn(5)	S37	IBtxOn(5)
S13	IBtxIp(6)	S38	IBtxOp(5)
S14	IBtxIn(6)	S39	IBtxOn(4)
S15	IBtxIp(7)	S40	IBtxOp(4)
S16	IBtxIn(7)	S41	IBtxOn(3)
S17	IBtxIp(8)	S42	IBtxOp(3)
S18	IBtxIn(8)	S43	IBtxOn(2)
G11	Vcc	S44	IBtxOp(2)
S19	IBtxIp(9)	G23	Sense-3.3V
S20	IBtxIn(9)	S45	IBtxOn(1)
S21	IBtxIp(10)	S46	IBtxOp(1)
S22	IBtxIn(10)	G24	Vcc
S23	IBtxIp(11)	S47	IBtxOn(0)
S24	IBtxIn(11)	S48	IBtxOp(0)

7.7.3.4 PIN ASSIGNMENT - 12X MICROGIGACN PORT USED AS 3-4X PORTS

The 12X to 3-4X cables are used for connecting to devices which may be configured to operate with a single 12X port, or with three separate 4X ports, using the same contacts. The 12X to 3-4X copper cable provides an interface to a 12X interface board connector, operating as either a single 12X port or three 4X ports. The opposite side of the cable provides three separate 4X cable connectors.

C7-16.2.4: The contact assignment listed in [Table 104](#) shall be used for the board connector for InfiniBand microGigaCN 12X used as three 4X ports with 12X to 3-4x microGigaCN cables.

The character 'x' in the signal symbol is the port number, as defined in [4.1 Signal Naming Conventions](#).

Table 104 12X board connector signal assignment for 12X to 3-4X cables

Pin Number	Signal (single 12X port)	Signal (three 4X ports)	Pin Number	Signal (single 12X port)	Signal (three 4X ports)
G1-G25	Signal Ground	Signal Ground	Housing	Chassis Ground	Chassis Ground
S1	IBtxIp(0)	IBtx.1Ip(0)	S25	IBtxOn(11)	IBtx.3On(3)
S2	IBtxIn(0)	IBtx.1In(0)	S26	IBtxOp(11)	IBtx.3Op(3)
S3	IBtxIp(1)	IBtx.1Ip(1)	S27	IBtxOn(10)	IBtx.3On(2)
S4	IBtxIn(1)	IBtx.1In(1)	S28	IBtxOp(10)	IBtx.3Op(2)
S5	IBtxIp(2)	IBtx.1Ip(2)	S29	IBtxOn(9)	IBtx.3On(1)
S6	IBtxIn(2)	IBtx.1In(2)	S30	IBtxOp(9)	IBtx.3Op(1)
S7	IBtxIp(3)	IBtx.1Ip(3)	S31	IBtxOn(8)	IBtx.3On(0)
S8	IBtxIn(3)	IBtx.1In(3)	S32	IBtxOp(8)	IBtx.3Op(0)
S9	IBtxIp(4)	IBtx.2Ip(0)	S33	IBtxOn(7)	IBtx.2On(3)
S10	IBtxIn(4)	IBtx.2In(0)	S34	IBtxOp(7)	IBtx.2Op(3)
S11	IBtxIp(5)	IBtx.2Ip(1)	S35	IBtxOn(6)	IBtx.2On(2)
S12	IBtxIn(5)	IBtx.2In(1)	S36	IBtxOp(6)	IBtx.2Op(2)
S13	IBtxIp(6)	IBtx.2Ip(2)	S37	IBtxOn(5)	IBtx.2On(1)
S14	IBtxIn(6)	IBtx.2In(2)	S38	IBtxOp(5)	IBtx.2Op(1)
S15	IBtxIp(7)	IBtx.2Ip(3)	S39	IBtxOn(4)	IBtx.2On(0)
S16	IBtxIn(7)	IBtx.2In(3)	S40	IBtxOp(4)	IBtx.2Op(0)
S17	IBtxIp(8)	IBtx.3Ip(0)	S41	IBtxOn(3)	IBtx.1On(3)
S18	IBtxIn(8)	IBtx.3In(0)	S42	IBtxOp(3)	IBtx.1Op(3)
S19	IBtxIp(9)	IBtx.3Ip(1)	S43	IBtxOn(2)	IBtx.1On(2)
S20	IBtxIn(9)	IBtx.3In(1)	S44	IBtxOp(2)	IBtx.1Op(2)
S21	IBtxIp(10)	IBtx.3Ip(2)	S45	IBtxOn(1)	IBtx.1On(1)
S22	IBtxIn(10)	IBtx.3In(2)	S46	IBtxOp(1)	IBtx.1Op(1)
S23	IBtxIp(11)	IBtx.3Ip(3)	S47	IBtxOn(0)	IBtx.1On(0)
S24	IBtxIn(11)	IBtx.3In(3)	S48	IBtxOp(0)	IBtx.1Op(0)

7.7.3.5 PIN ASSIGNMENT - 8X PASSIVE CABLE PORTS

This section defines the connector for the 8X passive cable interface on InfiniBand boards. This interface provides for simultaneous transmit and receive of eight bits of en-

coded differential data. It uses the same board connector as the 12X copper cable interface.

All 8X InfiniBand cable plugs shall be intermateable with this connector.

C7-16.2.5: The contact assignment listed in [Table 105 on page 410](#) shall be used for the board connector for passive InfiniBand 8X cables.

The character 'x' in the signal symbol is the port number, as defined in [Section 4.1, "Signal Naming Conventions," on page 71](#).

IBTA

Table 105 8X passive board connector signal assignment

Pin Number	Signal	Pin Number	Signal
G1-G25	Signal Ground	Housing	Chassis Ground
S1	IBtxlp(0)	S25	reserved
S2	IBtxln(0)	S26	reserved
S3	IBtxlp(1)	S27	reserved
S4	IBtxln(1)	S28	reserved
S5	IBtxlp(2)	S29	reserved
S6	IBtxln(2)	S30	reserved
S7	IBtxlp(3)	S31	reserved
S8	IBtxln(3)	S32	reserved
S9	IBtxlp(4)	S33	IBtxOn(7)
S10	IBtxln(4)	S34	IBtxOp(7)
S11	IBtxlp(5)	S35	IBtxOn(6)
S12	IBtxln(5)	S36	IBtxOp(6)
S13	IBtxlp(6)	S37	IBtxOn(5)
S14	IBtxln(6)	S38	IBtxOp(5)
S15	IBtxlp(7)	S39	IBtxOn(4)
S16	IBtxln(7)	S40	IBtxOp(4)
S17	reserved	S41	IBtxOn(3)
S18	reserved	S42	IBtxOp(3)
S19	reserved	S43	IBtxOn(2)
S20	reserved	S44	IBtxOp(2)
S21	reserved	S45	IBtxOn(1)
S22	reserved	S46	IBtxOp(1)
S23	reserved	S47	IBtxOn(0)
S24	reserved	S48	IBtxOp(0)

7.7.3.6 PIN ASSIGNMENT - 8X ACTIVE CABLE PORTS

This section defines the connector for the 8X active cable interface on InfiniBand boards. The active cable interface provides for simultaneous transmit and receive of eight bits of

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
encoded differential data with power available on the board connector for fiber or copper transponder devices. It uses the same board connector as the 8X and 12X passive copper cable interfaces, but with a different contact assignment for the signal ground contacts.

All 8X InfiniBand cable plugs that are not pluggable devices shall be intermateable with this connector.

C7-16.2.6: The contact assignment listed in [Table 106](#) shall be used for the board connector for active InfiniBand 8X cables. Usage requirements including current limitations on the power contacts are described in [Section 7.8.7.2, "Power Requirements," on page 442](#). Power return is by way of the Signal Ground contacts.

The 12 V or 3.3 V Sense signal is used to enable the respective voltage on the Vcc power supply contact used to provide power to the active components in the cable. Both signals are considered active if their voltage level is between 0.9 V and 2.4 V DC.

C7-16.1.7: Both 12 V and 3.3 V Sense signals shall not be active simultaneously.

The character 'x' in the signal symbol is the port number, as defined in [Section 4.1, "Signal Naming Conventions," on page 71](#).

Table 106 8X active board connector signal assignment

Pin Number	Signal	Pin Number	Signal
G1	Sense-12V	Housing	Chassis Ground
G2, G4-G10, G12-G22, G25	Signal Ground		
S1	IBtxIp(0)	S25	reserved
S2	IBtxIn(0)	S26	reserved
S3	IBtxIp(1)	S27	reserved
S4	IBtxIn(1)	S28	reserved
G3	Vcc	S29	reserved
S5	IBtxIp(2)	S30	reserved
S6	IBtxIn(2)	S31	reserved
S7	IBtxIp(3)	S32	reserved
S8	IBtxIn(3)	S33	IBtxOn(7)
S9	IBtxIp(4)	S34	IBtxOp(7)
S10	IBtxIn(4)	S35	IBtxOn(6)
S11	IBtxIp(5)	S36	IBtxOp(6)
S12	IBtxIn(5)	S37	IBtxOn(5)
S13	IBtxIp(6)	S38	IBtxOp(5)
S14	IBtxIn(6)	S39	IBtxOn(4)
S15	IBtxIp(7)	S40	IBtxOp(4)
S16	IBtxIn(7)	S41	IBtxOn(3)
S17	reserved	S42	IBtxOp(3)
S18	reserved	S43	IBtxOn(2)
G11	Vcc	S44	IBtxOp(2)
S19	reserved	G23	Sense-3.3V
S20	reserved	S45	IBtxOn(1)
S21	reserved	S46	IBtxOp(1)
S22	reserved	G24	Vcc
S23	reserved	S47	IBtxOn(0)
S24	reserved	S48	IBtxOp(0)

It is recommended that the 8X board and copper cable connectors as used for InfiniBand use the same keying as that for a 12X cable, as defined in [Figure 134 on page 403](#).

7.7.3.7 ACTIVE CABLE POWER REQUIREMENTS

7.7.3.7.1 HOST POWER SUPPLY FILTERING

Host board power supply filtering requirements for active 8X and 12X cables are defined in [7.3.3.2.1 on page 367](#).

7.7.3.7.2 HOST POWER SUPPLY SPECIFICATIONS

Host board power supply requirements for active 8X and 12X cables are defined in [7.3.3.2.2 on page 368](#).

7.7.3.7.3 SHORT CIRCUIT PROTECTION

Short circuit protection requirements for active 8X and 12X cables are defined in [7.3.3.2.6 on page 369](#).

7.7.3.7.4 POWER SENSE

Power sense requirements for active 8X and 12X cables are defined in [7.3.3.4.1 on page 369](#).

7.7.4 ENVIRONMENTAL REQUIREMENTS

7.7.4.1 ESD REQUIREMENTS

ESD requirements for 8X and 12X cables are defined in [7.3.4.1 on page 370](#).

7.7.4.2 HOT INSERTION AND REMOVAL

Hot insertion and removal requirements for 8X and 12X cables are defined in [7.3.4.2 on page 370](#).

7.8 CXP INTERFACE

7.8.1 INTRODUCTION²

This section describes a 12X form factor pluggable active device interface with 12 transmit and 12 receive lanes capable of supporting bit rates up to 26 Gb/s per lane on a variety of electrical and optical transmission technologies. The CXP module described in this document is illustrated in [Figure 135 on page 414](#) and [Figure 136 on page 415](#).

The CXP and CXP28 specifications are also described in SFF documents SFF-8648 "Mini Multilane 12x 28 Gb/s Shielded Cage/Connector (CXP28)", with mechanical specifications described in SFF-8617 "Mini Multilane 12x Shielded Cage/Connector (CXP)". The specifications are intended to be maintained as identical specifications, but in cases of conflict between this document and those documents, this one shall have preference.

2. Terminology note: The name CXP was derived from:

C = the Roman numeral for 100, indicating a form-factor targeted for >100 Gb/s per direction transmission.
C is also the hexadecimal character for 12, indicating an interface with 12 lanes per direction.

XP = eXtended-capability Pluggable form factor.

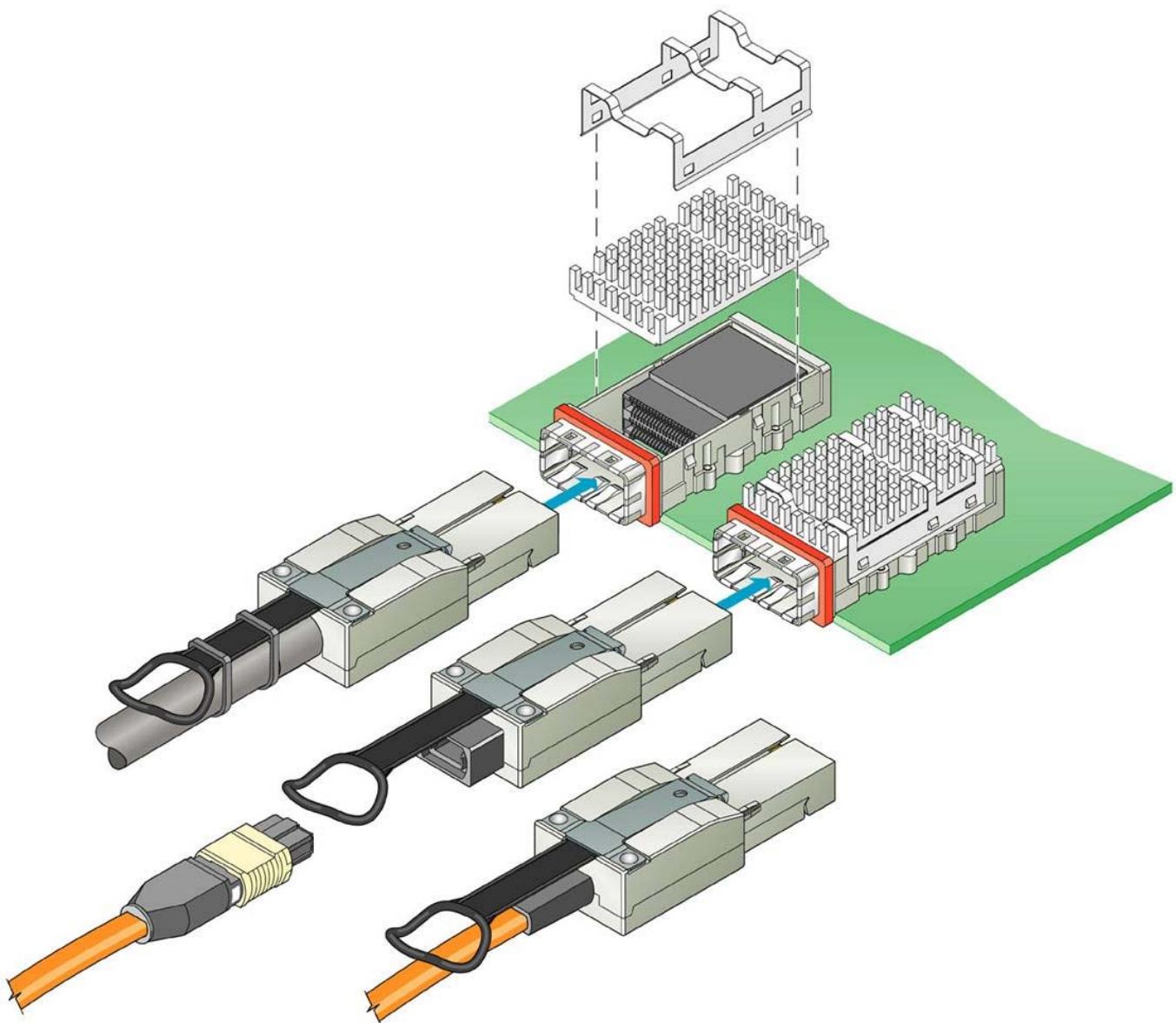


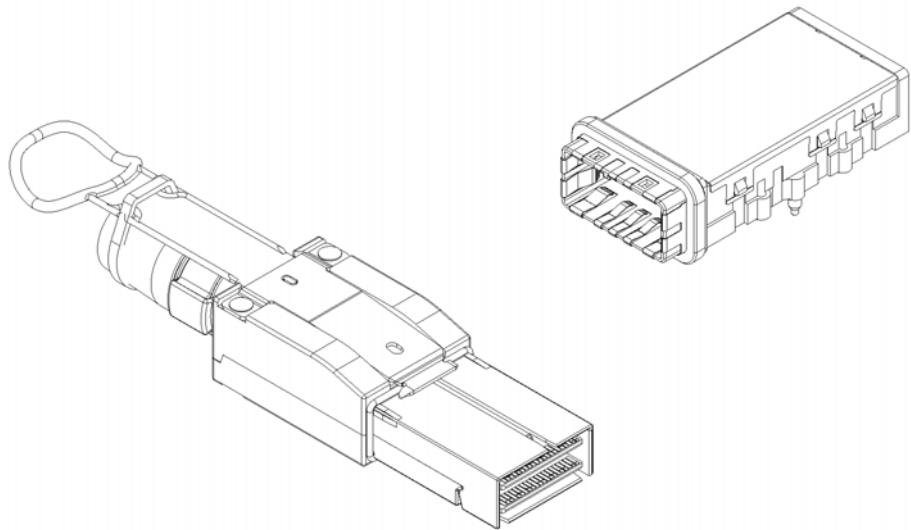
Figure 135 CXP Conceptual Model

A single host interface receptacle supporting a variety of transmission technologies, cost-effectively supporting a variety of link lengths. Cooling requirements will be dependent on device technology. For devices and cables that require cooling inside the host, the integrated receptacle may require a thermal solution to augment the host solution. One example is a riding heat sink as shown in [Figure 143 on page 425](#). Dimensions not specifically called out may be modified, subject to intermateability and interchangeability constraints of the individual application.

This specification describes the form factor, electrical, mechanical, power, and thermal interfaces between the devices or cables and the systems. The transmission technology

(e.g., optical or electrical), transmission medium (e.g., single-mode, multi-mode fiber, or copper), form factor (i.e., with cable attached to the pluggable device, or detachable with a separable connector), and physical layer definition for the communication between transceivers are not explicitly specified. However, the specifications are intended to support several different technologies, including VCSEL/MMF parallel ribbon fiber links.

This variety of transmission technologies allow cost-effective implementation across a wide range of link lengths using the same host receptacle and electrical interfaces. Note also that explicit transmission technology choices within an active cable assembly (e.g., single-mode vs. multi-mode optical transmission, glass vs. plastic optical fiber, or equalization and coding techniques) are not addressed by this specification. This document specifies the electrical, mechanical and thermal interfaces between “module” (cable plug, or transceiver) and the host. Any transmission technology which transports data transmission between two interfaces at the specified speed with good signal integrity at each end is compliant.



**Figure 136 CXP Cable Plug/module and Receptacle/Cage
(without heat sink)**

7.8.2 MECHANICAL REQUIREMENTS

7.8.2.1 PHYSICAL AND MECHANICAL PERFORMANCE REQUIREMENTS

The requirements for insertion forces, extraction forces and retention forces are specified in [Table 107 on page 416](#). CXP modules, connectors, receptacles and receptacle housings should not be damaged by module removal or insertion. If any part is damaged by

excessive force, it should be the cable or module, and not the receptacle or receptacle housing which are part of the host system.

Table 107 CXP connector physical requirements

Symbol	Parameter	Min	Max	Unit	Comments
F_i	CXP module insertion force		80	N	EIA-364-13
F_w	CXP module extraction force		50	N	EIA-364-13
F_r	CXP module retention	90	170	N	Load pull, per EIA-364-38A No damage to transceiver below 90 N
F_{rcl}	Cage retention (latch strength)	180		N	No damage to latch below 180 N
F_{rhb}	Cage retention in host board	114		N	Force to be applied in a vertical direction, no damage to cage
N_{hc}	Insertion / removal cycles, connector/receptacle	100		Cycles	Number of cycles for the connector and receptacle with multiple transceivers
N_x	Insertion / removal cycles, CXP module	50		Cycles	Number of cycles for an individual module

It is also recommended that the connector interfaces meet the parameters defined in [Table 108](#).

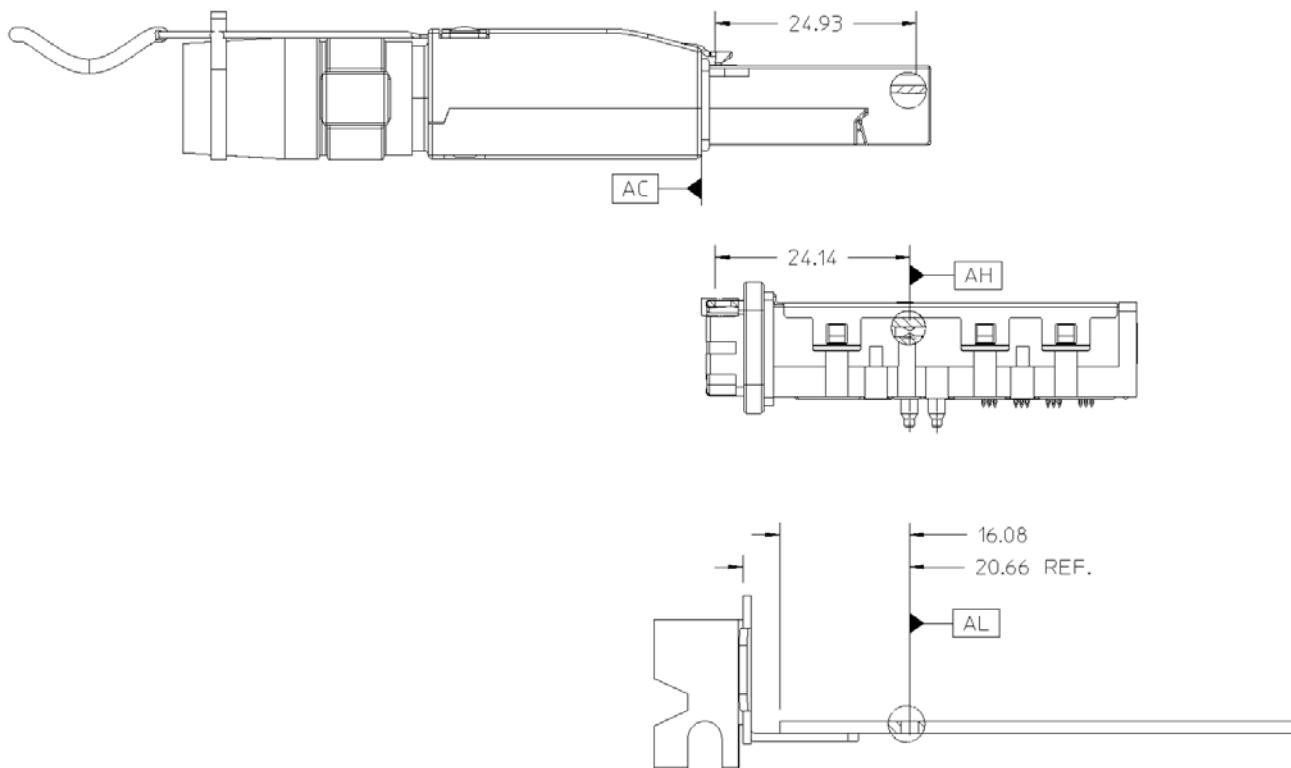
Table 108 Recommended CXP connector physical parameters

Symbol	Parameter	Min	Max	Unit	Comments
t_{pm}	Contact finish	0.76 Au over 1.27 Nickel		μm	As necessary to meet N_x requirements
F_n	Contact normal force	50		cN	per contact
S_{hcc}	Contact Hertz stress	170		kpsi	per contact
D_{wc}	Contact wipe length	0.75		mm	

It is recommended that all components and attach processes used for those components be compliant with RoHS directive 2002/95/EC issued January 27, 2003.

7.8.2.2 DATUMS AND COMPONENT ALIGNMENT

The datums for the various components are shown in [Figure 137 on page 417](#). The alignments of some of the datums are noted. All dimensions are in millimeters.



Datum	Description	Datum	Description
AA	Bottom of Plug /Module	AG	Bottom of Cage
AB	Width of Plug/Module Snout	AH	Centerline of Contacts
AC	Plug/Module Hard Stop	AJ	Front of Cage
AD	Width of Plug.Module PCB	AK	Width of Cage Opening
AE	Top of Plug/Module PCB	AL	Locating Hole in PCB
AF	Leading Edge of Outermost Signal Contact Pads	AM	-

Figure 137 CXP interface datum definitions

7.8.2.3 PIN ASSIGNMENT

[Figure 138](#) shows the contact numbering for the CXP module. The signal contact assignment is listed in [Table 109 on page 419](#).

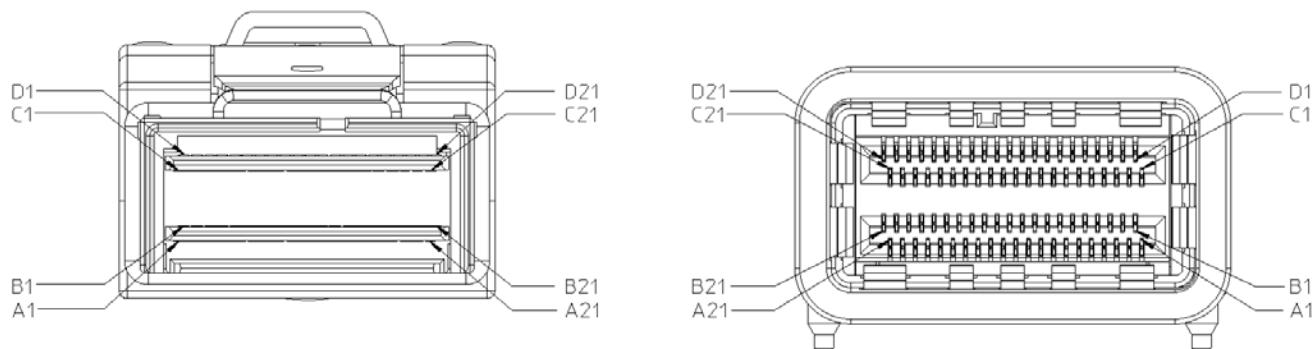


Figure 138 CXP module and host board connector pin assignments

Table 109 Contact Assignments for 12x Pluggable-CXP Interface

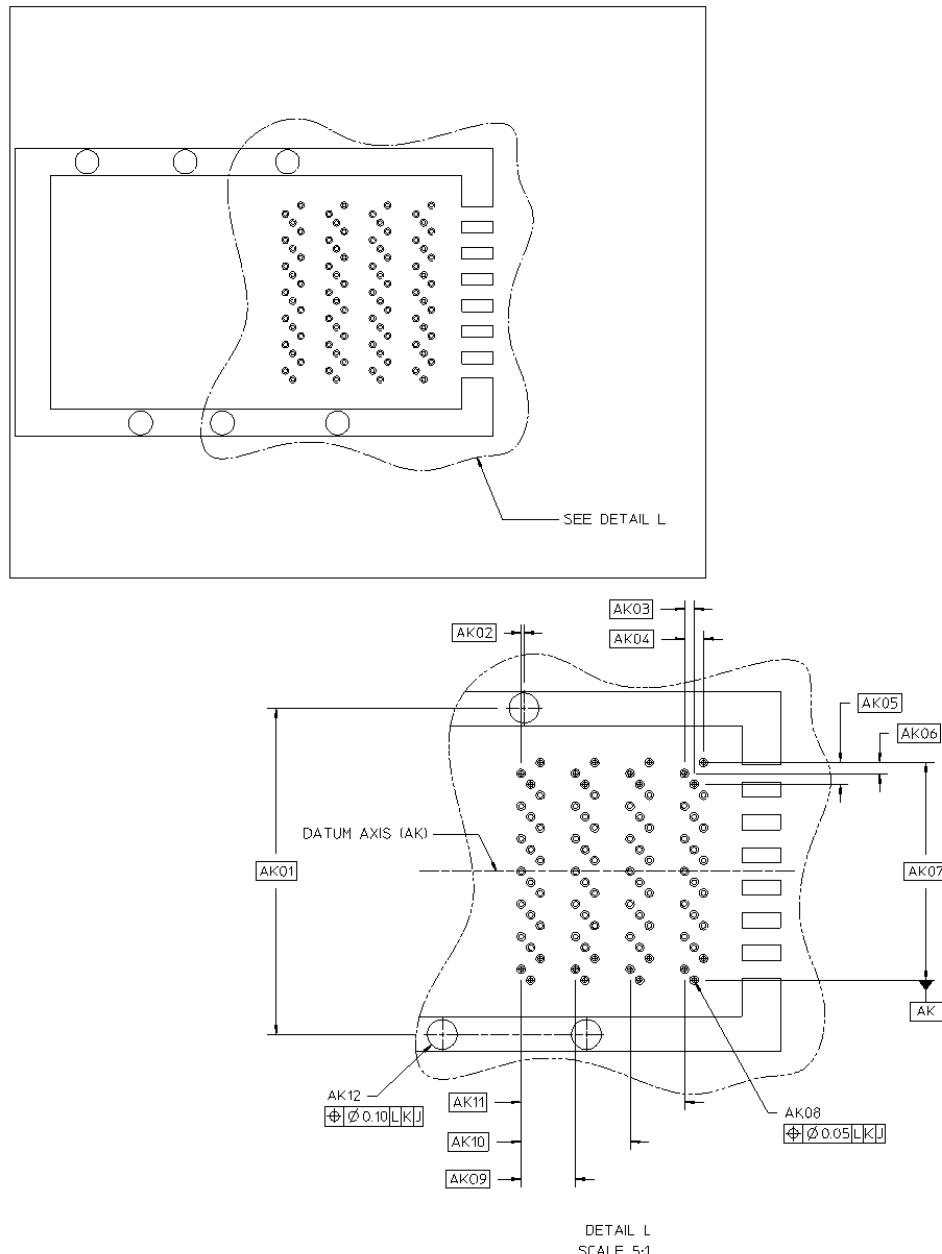
Bottom side			Top Side		
I/O #	Name	Contact Length	Contact Length	Name	I/O #
Receiver -- Top Card					
C1	GND			GND	D1
C2	Rx1p			Rx0p	D2
C3	Rx1n			Rx0n	D3
C4	GND			GND	D4
C5	Rx3p			Rx2p	D5
C6	Rx3n			Rx2n	D6
C7	GND			GND	D7
C8	Rx5p			Rx4p	D8
C9	Rx5n			Rx4n	D9
C10	GND			GND	D10
C11	Rx7p			Rx6p	D11
C12	Rx7n			Rx6n	D12
C13	GND			GND	D13
C14	Rx9p			Rx8p	D14
C15	Rx9n			Rx8n	D15
C16	GND			GND	D16
C17	Rx11p			Rx10p	D17
C18	Rx11n			Rx10n	D18
C19	GND			GND	D19
C20	PRSNT_L			Vcc3.3-Rx	D20
C21	Int_L/Reset_L			Vcc12-Rx	D21
Transmitter -- Bottom Card					
A1	GND			GND	B1
A2	Tx1p			Tx0p	B2
A3	Tx1n			Tx0n	B3
A4	GND			GND	B4
A5	Tx3p			Tx2p	B5
A6	Tx3n			Tx2n	B6
A7	GND			GND	B7
A8	Tx5p			Tx4p	B8
A9	Tx5n			Tx4n	B9
A10	GND			GND	B10
A11	Tx7p			Tx6p	B11
A12	Tx7n			Tx6n	B12
A13	GND			GND	B13
A14	Tx9p			Tx8p	B14
A15	Tx9n			Tx8n	B15
A16	GND			GND	B16
A17	Tx11p			Tx10p	B17
A18	Tx11n			Tx10n	B18
A19	GND			GND	B19
A20	SCL			Vcc3.3-Tx	B20
A21	SDA			Vcc12-Tx	B21

7.8.2.4 HOST BOARD CONNECTOR

This section defines the CXP host board connector. A suitable CXP host board connector footprint (informative) is shown in [Figure 139](#). Package drawings and dimensions for the host board connector housing are shown in [Figure 140](#) and [Figure 141 on page 423](#). Note that some dimensions of the receptacle housing relate specifically to the heat sink clip attachment, and are shown separately in the following section.

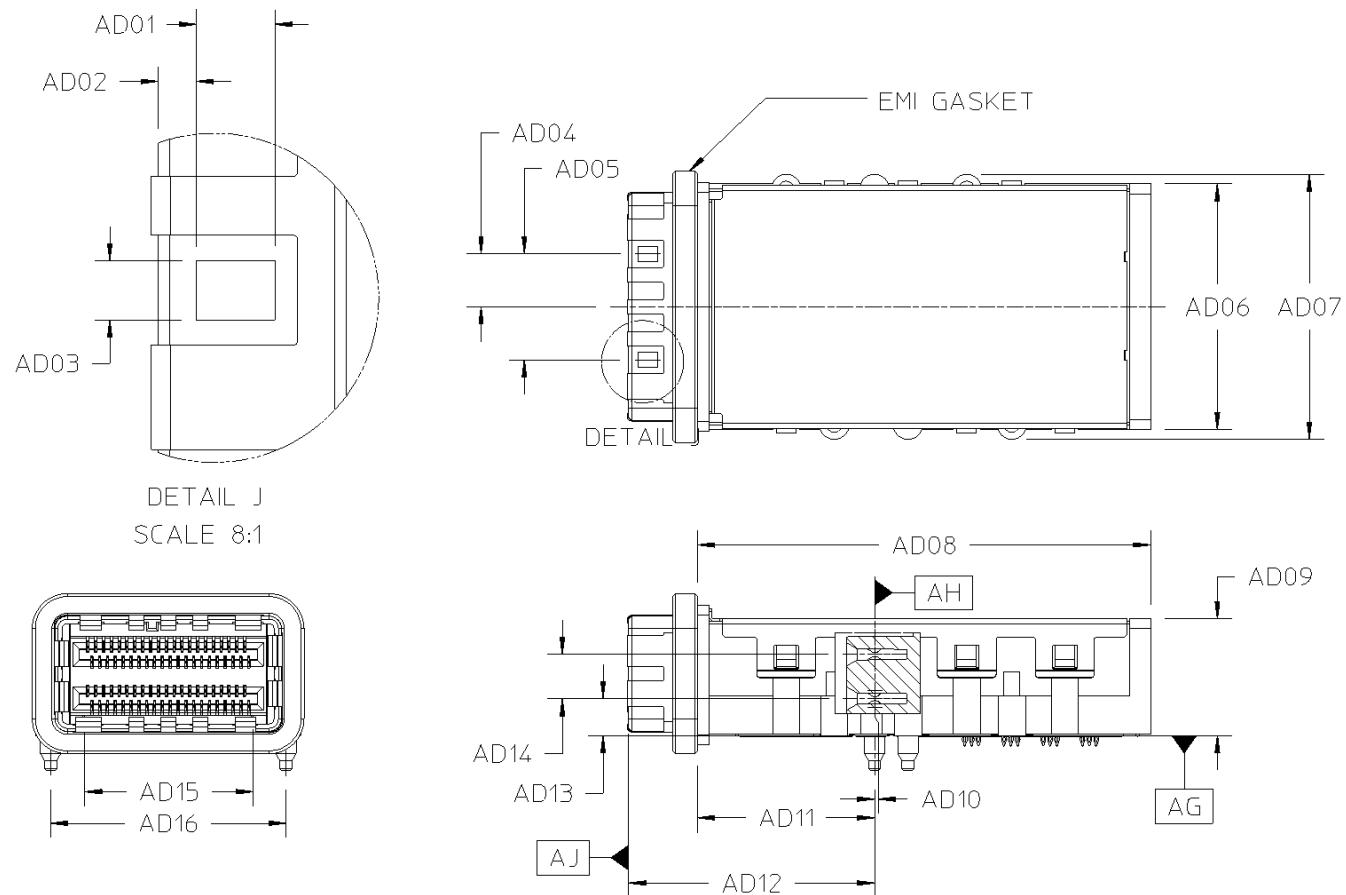
IBTA

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42



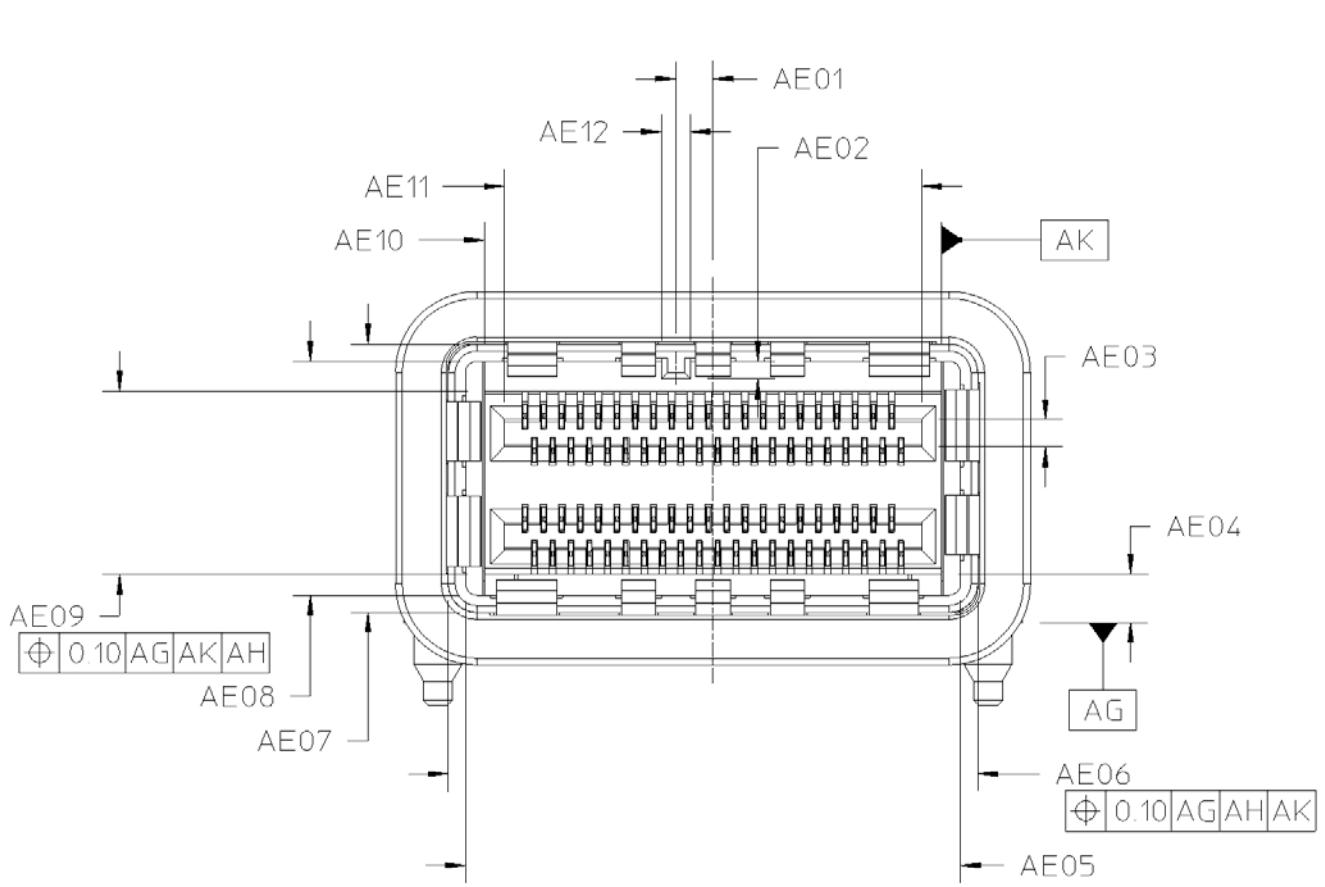
ID	Description	Dim	Tol (±)	ID	Description	Dim	Tol (±)
AK01	Locating Hole to Locating Hole	24.00	0.05	AK07	First to Last Column	16.00	Basic
AK02	Locating Hole to First Row of Signal Holes	0.20	Basic	AK08	Contact Hole Diameter (Finished PTH)	0.37	0.05
AK03	First Row to Second Row of Signal Holes	0.70	Basic	AK09	Row A to Row B	4.00	Basic
AK04	First Row to Third Row of Signal Holes	1.40	Basic	AK10	Row A to Row C	8.00	Basic
AK05	Column to Column Pitch	1.60	Basic	AK11	Row A to Row D	12.00	Basic
AK06	Column to Column Pitch	0.80	Basic	AK12	Locating Hole Diameter (Finished PTH)	2.20	0.05

Figure 139 CXP host board connector footprint



ID	Description	Dim	Tol (\pm)	ID	Description	Dim	Tol (\pm)
AD01	Latch Hole Length	2.00	0.10	AD09	Shell Height	11.88	0.13
AD02	Latch Hole from Face	0.97	0.05	AD10	Locating Post Centerline to Centerline of Receptacle	0.05	0.05
AD03	Latch Hole Width	1.50	0.10	AD11	Locating Post to EMI Shell Base	18.06	0.13
AD04	Datum to Latch Hole	5.40	0.10	AD12	Locating Post to Face	25.06	0.08
AD05	Latch Hole to Hole	10.80	0.05	AD13	PCB to Lower Card Slot Centerline	3.75	0.10
AD06	Shell Width	25.05	0.25	AD14	Lower Card Slot to Upper Card Slot Centerline	4.50	0.10
AD07	Shell Width at screw attach features	27.00	0.25	AD15	Card Slot Rib to Rib	17.18	0.10
AD08	EMI Shell Base to Back	46.22	0.25	AD16	Peg Centerlin to Peg Centerline	24.00	0.08

Figure 140 CXP host board connector dimensions (part 1 of 2)

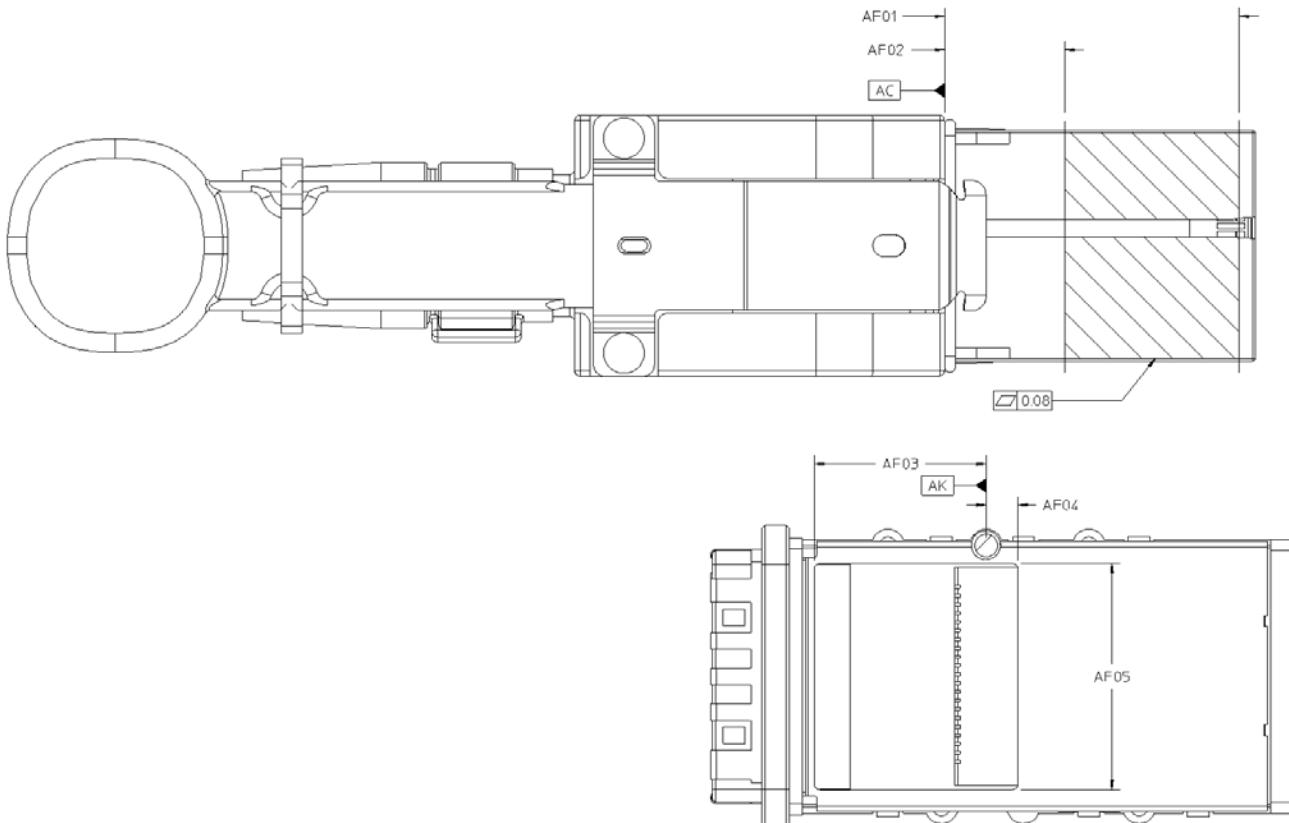


ID	Description	Dim	Tol (±)	ID	Description	Dim	Tol (±)
AE01	Orientation Key Location	1.63	0.13	AE07	Snout Height	11.70	0.08
AE02	Orientation Key Location, Depth	20.75	Basic	AE08	Snout Opening Height	10.20	0.05
AE03	Card Slot Height	1.18	0.05	AE09	Receptacle Body Height	8.00	0.08
AE04	Datum to Bottom of Receptacle Housing	2.09	0.10	AE10	Receptacle Body Width	19.89	0.05
AE05	Snout Opening Width	21.60	0.05	AE11	Card Slot Width	18.20	0.05
AE06	Snout Width	23.10	0.08	AE12	Orientation Key Width	1.25	0.13

Figure 141 CXP host board connector dimensions (part 2 of 2)

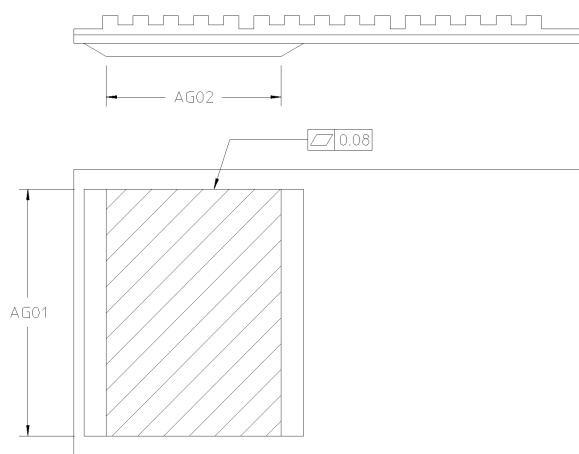
7.8.2.5 HEAT SINK INTERFACES

[Figure 142](#) and [Figure 143 on page 425](#) show exemplary dimensions for heat sink attachment to the CXP host board connector cage.



ID	Description	Dim	Tol (\pm)	ID	Description	Dim	Tol (\pm)
AF01	Heat Sink Interface Zone	27.00	Min.	AF04	Locator Pin to Heat Sink Cover Opening	2.91	0.10
AF02	Plug Body to Heat Sink Interface Start	11.00	Max.	AF05	Heat Sink Cover Opening Width	20.75	0.10
AF03	Flange to Locator Pin	0.10	Basic	-	-	-	-

Figure 142 CXP module and host board cage heat sink dimensions



ID	Description	Dim	Tol (±)	ID	Description	Dim	Tol (±)
AG01	Heat Sink Pad Width	20.25	0.10	AG02	Heat Sink Pad Length	14.47	0.10

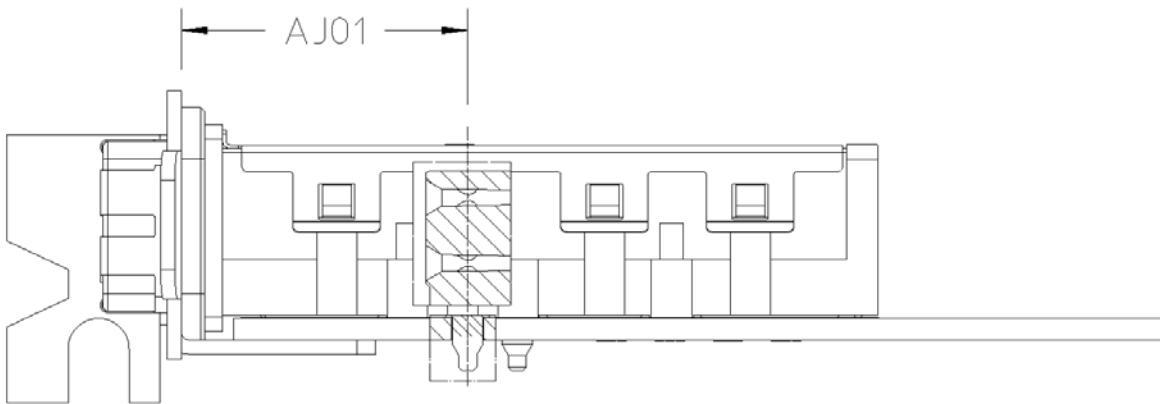
Figure 143 CXP host board cage riding heat sink dimensions

7.8.2.6 HOST BOARD BEZEL

Host enclosures that use CXP devices should provide appropriate clearances between the CXP transceivers to allow insertion and extraction without the use of special tools and a bezel enclosure with sufficient mechanical strength. The relationship of the transceiver/connector and receptacle housing relative to the host board and bezel is illustrated in [Figure 144](#) by the location of the key datums of each of the components. [Figure 145 on page 426](#) shows the bezel opening dimensions for the host board connector.

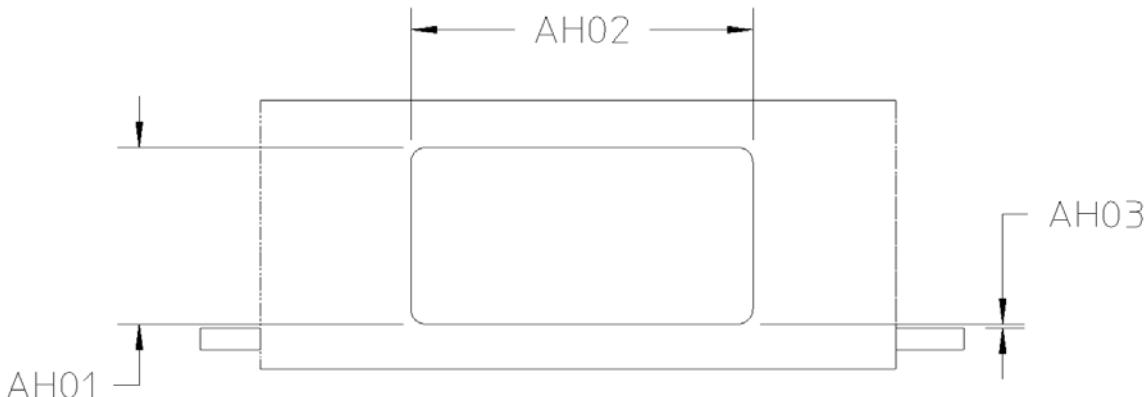
The front surface of the receptacle housing may pass through the bezel. If EMI spring fingers are used, they shall make contact to the inside of the bezel cutout. If an EMI gasket is used, it shall make contact to the inside surface of the bezel or to the inside of the bezel cutout. To accept all housing designs, both bezel surfaces must be conductive and connected to chassis ground.

The CXP transceiver insertion slot should be clear of nearby moldings and covers that might block convenient access to the latching mechanisms, the CXP transceiver, or the cables connected to the CXP transceiver.



ID	Description	Dim	Tol (±)	ID	Description	Dim	Tol (±)
AJ01	Centerline of Receptacle Contacts to Base of EMI Sh	19.66	Basic	-	-	-	-

Figure 144 CXP cage to bezel dimensions



ID	Description	Dim	Tol (±)	ID	Description	Dim	Tol (±)
AH01	Cutout Height	12.10	Basic	AH03	Bottom of Cutout to Peg	0.28	Basic
AH02	Cutout Length	23.50	Basic	-	-	-	-

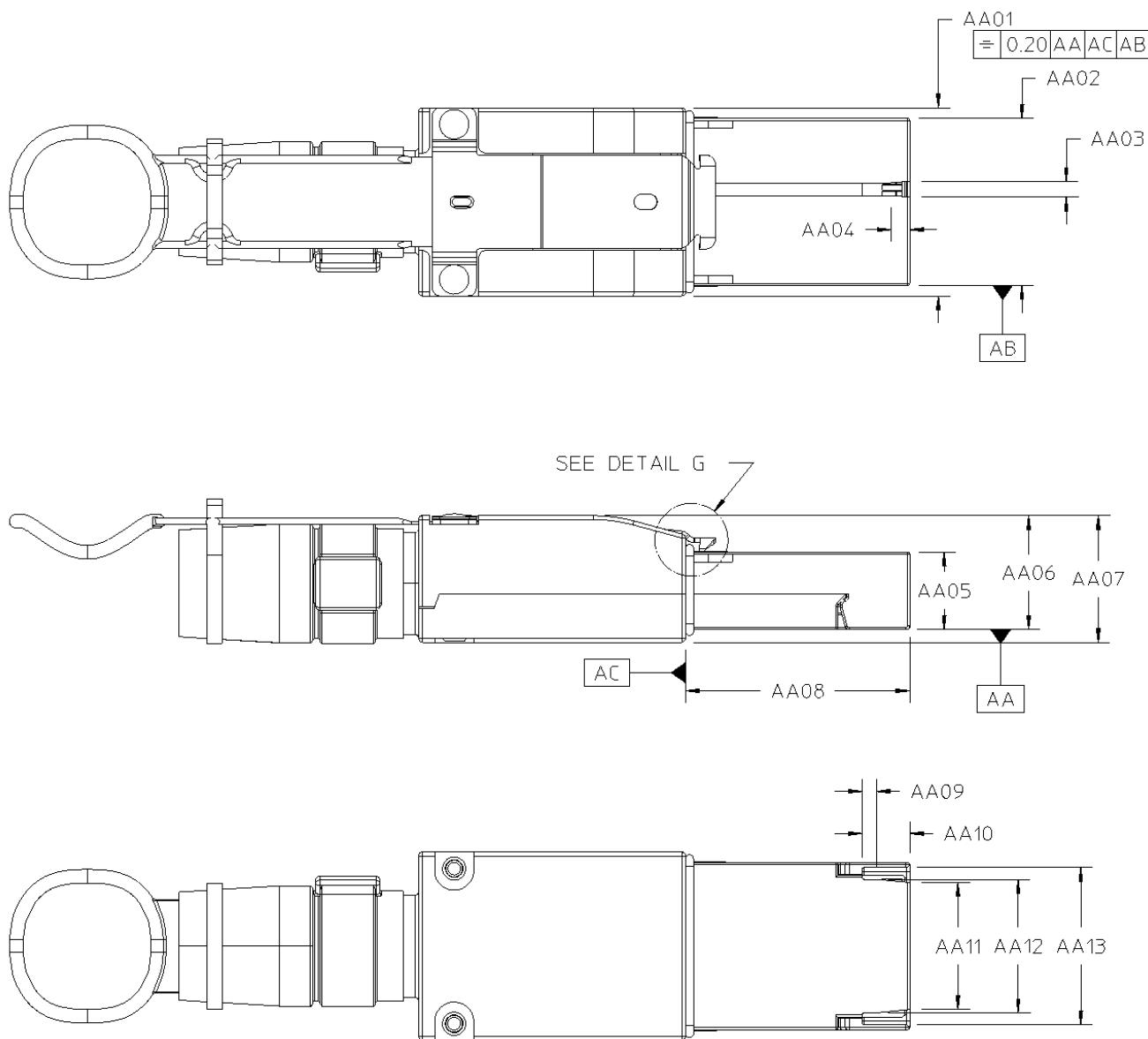
Figure 145 CXP bezel opening dimensions

7.8.2.7 MODULE PACKAGE

A common mechanical outline is used for all CXP cable plugs and CXP Modules. The package drawing and dimensions for the CXP module and cable plug are defined in [Figure 146 on page 428](#) and [Figure 147 on page 429](#).

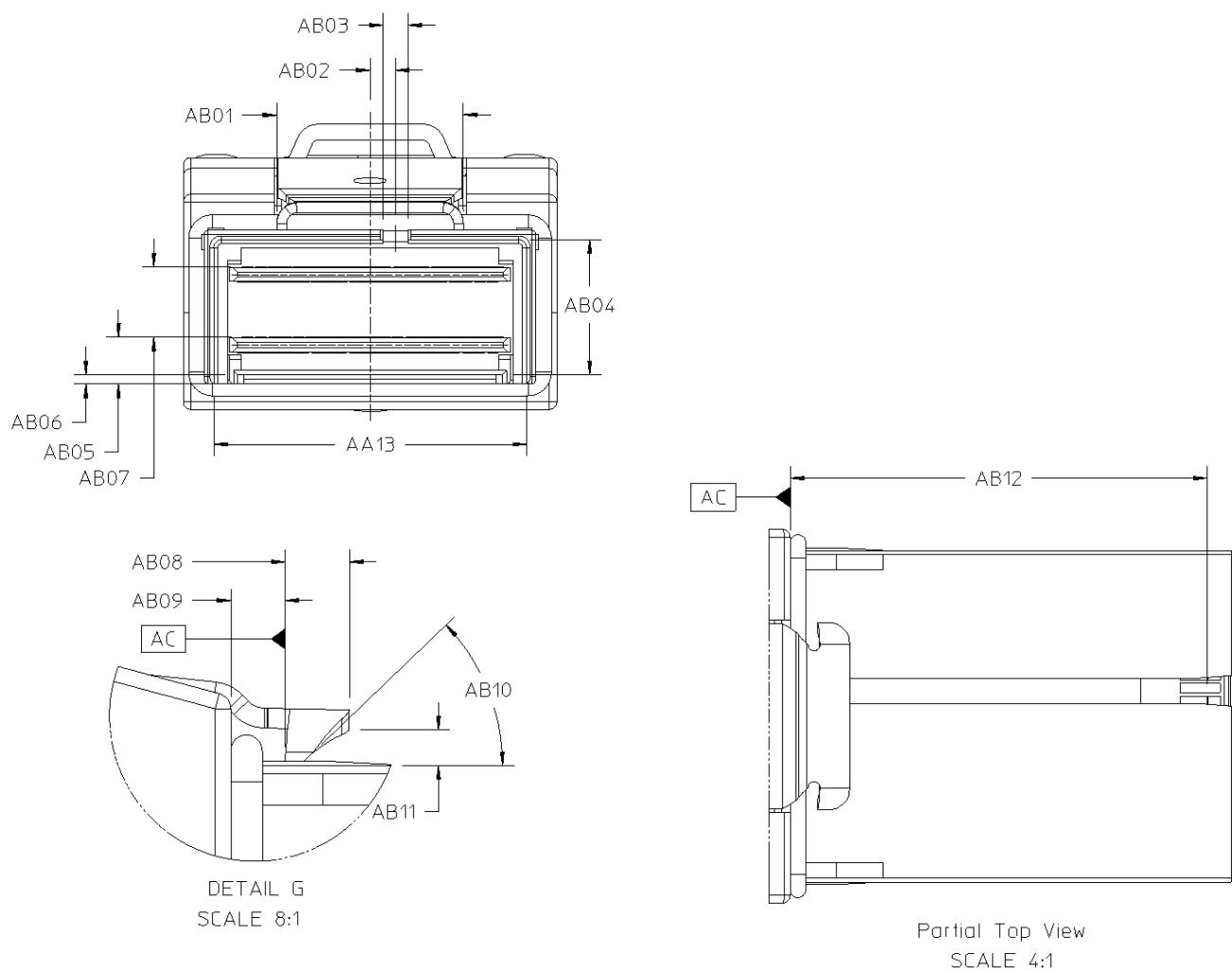
IBTA

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42



ID	Description	Dim	Tol (\pm)	ID	Description	Dim	Tol (\pm)
AA01	Body Width	24.05	Max.	AA08	Snout Length	28.45	0.13
AA02	Snout Width	21.20	0.13	AA09	Length of Tongue - Straight Section	1.80	0.10
AA03	Orientation Key Lead-in Width	2.00	0.25	AA10	Length of Tongue	6.00	Min.
AA04	Orientation Key Lead-in Length	2.40	0.25	AA11	Tongue Width - Tip	16.10	0.20
AA05	Snout Thickness	9.81	0.13	AA12	Tongue Width - Base	16.95	0.10
AA06	Snout Bottom to Plug Top	14.46	Max.	AA13	Inside Width of Snout	20.00	0.05
AA07	Plug Body Thickness	16.21	Max.	-	-	-	-

Figure 146 CXP module dimensions (basic)

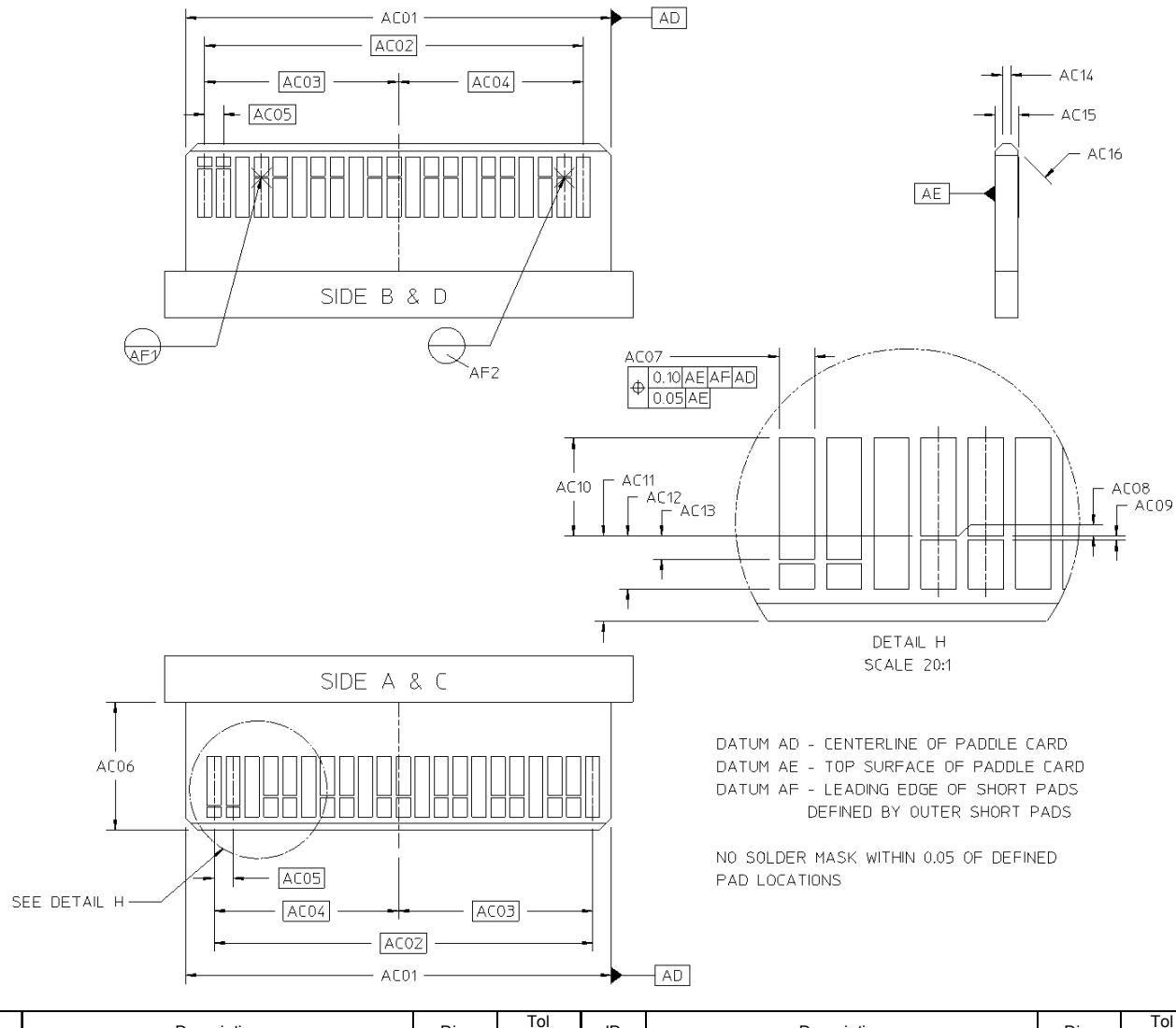


ID	Description	Dim	Tol (±)	ID	Description	Dim	Tol (±)
AB01	Latch Width	11.90	0.10	AB07	Top of 1st to top of 2nd Card	4.50	0.10
AB02	Orientation Key Location	1.625	0.13	AB08	Latch Barb Length	2.05	0.10
AB03	Orientation Key Width	1.60	0.10	AB09	Plug Body (Datum AC) to Latch	1.74	0.13
AB04	Inside Height of Snout	8.60	0.25	AB10	Barb Lead-in Angle	45°	1°
AB05	Base of Snout to top of 1st Paddle Card	2.99	0.20	AB11	Barb Lead-in Height	1.14	0.10
AB06	Tongue Thickness	0.60	Ref.	AB12	Plug Body to Short Pad	26.67	0.20

Figure 147 CXP module dimensions (detail)

7.8.2.8 MODULE PADDLE CARD / ELECTRICAL INTERFACE

[Figure 148](#) shows the dimensions of the paddle card which forms the plug contacts inside the connector plug.



ID	Description	Dim	Tol (±)	ID	Description	Dim	Tol (±)
AC01	Paddle Card Width	18.00	0.08	AC09	Pad to Pre-Pad	0.08	0.05
AC02	First to Last Pad Centers	16.00	Basic	AC10	Pad Length - Third Mate	1.65	Min.
AC03	Card Center to Outer Pad Center	8.20	Basic	AC11	Third Mate to Card Edge	1.45	0.10
AC04	Card Center to Outer Pad Center	7.80	Basic	AC12	Third Mate to Second Mate	0.90	0.05
AC05	Pad Center to Pad Center (Pitch)	0.80	Basic	AC13	Third Mate to First Mate	0.40	0.05
AC06	Component Keep Out Area	5.40	Min.	AC14	Lead-in Flat	0.36	Ref.
AC07	Pad Width	0.60	0.03	AC15	Paddle Card Thickness	1.00	0.10
AC08	Datum to Third Mate Pad	0.00	0.03	AC16	Lead-in Chamfer x 45°	0.30	0.05

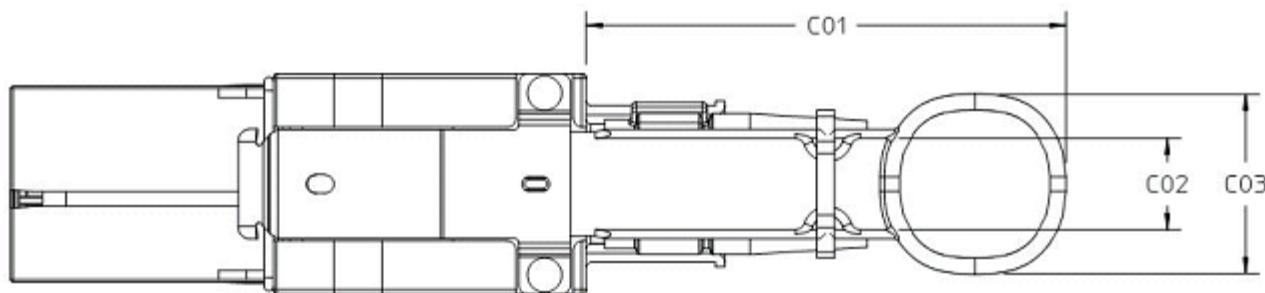
Note regarding AC01 Paddle Card Width: Tolerance of 0.08 mm applies to FDR devices. For QDR/DDR/SDR devices, the tolerance is 0.10 mm, as specified in InfiniBand Specification Release 1.2.1, Annex 6.

Figure 148 CXP module paddle card dimensions

7.8.2.9 LATCH RELEASE

CXP modules and cables require a releasable latch for retention into the receptacle housing assembly. Since a latch release mechanism is not an interface that affects interoperability, the design of the latch release mechanism may be media-dependent and vendor-dependent, and is beyond the scope of this specification.

[Figure 149](#) shows an exemplary pull-tab type latch release mechanism for a CXP module.



	Description	Dim.	Tol.
C01	Pull Tab Length	50.00	Ref
C02	Pull Tab Inner Diameter	9.91	Ref
C03	Pull Tab Width	19.51	Ref

Figure 149 Exemplary CXP module latch release mechanism

7.8.3 ELECTRICAL REQUIREMENTS

7.8.3.1 MATED CONNECTOR ELECTRICAL PARAMETERS

The CXP module and Receptacle shall comply to the electrical specifications described in [Table 110](#).

Table 110 CXP Module & Receptacle Connector Electrical Performance Requirements

Symbol	Parameter	Min	Max	Unit	Comments
LLCR	Low level contact resistance - initial		80	mΩ	through testing per EIA-364-23, measured across interface between paddle card trace and receptacle
ΔLLCR	Low level contact resistance - change		20	mΩ	through testing per EIA-364-23, as a result of any test group setup
I _{max}	Current rating, all contacts simultaneously	0.5		A	per EIA-364-70 or IEC 512-5-1 Test 5a, at 30°C temperature rise above ambient
I _{max,s}	Current rating, single contact	1		A	per EIA-364-70 or IEC 512-5-1 Test 5a, at 30°C temperature rise above ambient
R _{Iso}	Insulation Resistance	1000		MΩ	100 Vdc, between adjacent contacts
V _{Iso}	Dielectric Withstanding Voltage	300		Vdc	No defect or breakdown between adjacent contacts, 300 Vdc minimum for 1 minute
Z _{dco(peak)}	Differential Impedance - peak (connector area)	90	110	Ω	EIA-364-108 Rise time: 1.4*UI ps (20-80%). See Table 69 .
Z _{dco(nom)}	Differential Impedance (nominal)	95	105	Ω	Includes connector cable to connector interface and board termination pads and vias.
S _{cop}	Within-Pair Skew		5	ps	maximum (by design), measured at interface between paddle cards & receptacle. EIA-364-103
NEXT _c	Near End Crosstalk Isolation		-34	dB	EIA-364-90, 50 MHz to 10 GHz. Equivalent to 2% voltage crosstalk, power sum
L _{co}	Insertion Loss		1.0	dB	EIA-364-101, 50 MHz to 5 GHz

7.8.3.2 COMPLIANCE BOARDS

Specifications for host and module compliance boards are defined in [Annex 1: FDR and EDR Compliance Boards and Test Setups on page 614](#).

7.8.3.3 PIN ASSIGNMENT

[Table 109 on page 419](#) shows the signal assignment and module PCB edge as a top and bottom view, for both circuit cards in the two-level connector. There are 21 pads per level, for a total of 84, with 48 pads allocated for (12+12) differential pairs, 28 for Signal Common or Ground (GND), 4 for power connections, 4 for control/service.

The operation of the low-speed control and status lines (PRSNT_L, Int_L/Reset_L, SCL, & SDA) is described in [Section 7.8.3.4, “Low-Speed Electrical Contact/Signal Definitions,” on page 434](#), and operation of the high-speed lines for receiving and transmitting data is described in [Section 7.8.3.5, “High-Speed Electrical Contact/Signal Definitions,” on page 435](#).

The contact assignment listed in [Table 111](#) shall be used for the host board connector for InfiniBand 12X CXP cables.

Table 111 CXP Board Connector Signal Assignment

CXP Pin Number	Signal
B2	Tx0p
B3	Tx0n
A2	Tx1p
A3	Tx1n
B5	Tx2p
B6	Tx2n
A5	Tx3p
A6	Tx3n
B8	Tx4p
B9	Tx4n
A8	Tx5p
A9	Tx5n
B11	Tx6p
B12	Tx6n
A11	Tx7p
A12	Tx7n
B14	Tx8p
B15	Tx8n
A14	Tx9p
A15	Tx9n
B17	Tx10p
B18	Tx10n
A17	Tx11p
A18	Tx11n

A1, A4, A7, A10, A13, A16, A19, B1, B4, B7, B10, B13, B16, B19, C1, C4, C7, C10, C13, C16, C19, D1, D4, D7, D10, D13, D16, D19 on each plug are connected to local Signal Ground. They are not connected through cable.

A20, A21, C20, and C21 are connected to local management interface. They are not connected through cable.

B20, B21, D20, and D21 provide local power. They are not connected through cable.

CXP Pin Number	Signal
D2	Rx0p
D3	Rx0n
C2	Rx1p
C3	Rx1n
D5	Rx2p
D6	Rx2n
C5	Rx3p
C6	Rx3n
D8	Rx4p
D9	Rx4n
C8	Rx5p
C9	Rx5n
D11	Rx6p
D12	Rx6n
C11	Rx7p
C12	Rx7n
D14	Rx8p
D15	Rx8n
C14	Rx9p
C15	Rx9n
D17	Rx10p
D18	Rx10n
C17	Rx11p
C18	Rx11n

A1, A4, A7, A10, A13, A16, A19, B1, B4, B7, B10, B13, B16, B19, C1, C4, C7, C10, C13, C16, C19, D1, D4, D7, D10, D13, D16, D19 on each plug are connected to local Signal Ground. They are not connected through cable.

A20, A21, C20, and C21 are connected to local management interface. They are not connected through cable.

B20, B21, D20, and D21 provide local power. They are not connected through cable.

The contact assignment listed in [Table 112](#) shall be used for the host board connector for InfiniBand 12X CXP to 3-4X QSFP cables.

Table 112 12X Board Connector Signal Assignment for 12X CXP to 3-4X QSFP Cables

CXP Pin Number	Signal (single 12x port)	Signal (three 4X ports) ^{a,b}
B2	Tx0p	IBtx.1Op(0)
B3	Tx0n	IBtx.1On(0)
A2	Tx1p	IBtx.1Op(1)
A3	Tx1n	IBtx.1On(1)
B5	Tx2p	IBtx.1Op(2)
B6	Tx2n	IBtx.1On(2)
A5	Tx3p	IBtx.1Op(3)
A6	Tx3n	IBtx.1On(3)
B8	Tx4p	IBtx.2Op(0)
B9	Tx4n	IBtx.2On(0)
A8	Tx5p	IBtx.2Op(1)
A9	Tx5n	IBtx.2On(1)
B11	Tx6p	IBtx.2Op(2)
B12	Tx6n	IBtx.2On(2)
A11	Tx7p	IBtx.2Op(3)
A12	Tx7n	IBtx.2On(3)
B14	Tx8p	IBtx.3Op(0)
B15	Tx8n	IBtx.3On(0)
A14	Tx9p	IBtx.3Op(1)
A15	Tx9n	IBtx.3On(1)
B17	Tx10p	IBtx.3Op(2)
B18	Tx10n	IBtx.3On(2)
A17	Tx11p	IBtx.3Op(3)
A18	Tx11n	IBtx.3On(3)

A1, A4, A7, A10, A13, A16, A19, B1, B4, B7, B10, B13, B16, B19, C1, C4, C7, C10, C13, C16, C19, D1, D4, D7, D10, D13, D16, D19 on each plug are connected to local Signal Ground. They are not connected through cable.

A20, A21, C20, and C21 are connected to local management interface. They are not connected through cable.

B20, B21, D20, and D21 provide local power. They are not connected through cable.

CXP Pin Number	Signal (single 12x port)	Signal (three 4X ports)
D2	Rx0p	IBtx.1Ip(0)
D3	Rx0n	IBtx.1In(0)
C2	Rx1p	IBtx.1Ip(1)
C3	Rx1n	IBtx.1In(1)
D5	Rx2p	IBtx.1Ip(2)
D6	Rx2n	IBtx.1In(2)
C5	Rx3p	IBtx.1Ip(3)
C6	Rx3n	IBtx.1In(3)
D8	Rx4p	IBtx.2Ip(0)
D9	Rx4n	IBtx.2In(0)
C8	Rx5p	IBtx.2Ip(1)
C9	Rx5n	IBtx.2In(1)
D11	Rx6p	IBtx.2Ip(2)
D12	Rx6n	IBtx.2In(2)
C11	Rx7p	IBtx.2Ip(3)
C12	Rx7n	IBtx.2In(3)
D14	Rx8p	IBtx.3Ip(0)
D15	Rx8n	IBtx.3In(0)
C14	Rx9p	IBtx.3Ip(1)
C15	Rx9n	IBtx.3In(1)
D17	Rx10p	IBtx.3Ip(2)
D18	Rx10n	IBtx.3In(2)
C17	Rx11p	IBtx.3Ip(3)
C18	Rx11n	IBtx.3In(3)

a. Nomenclature for signals is described in Vol. 2, Table 4.

b. Note that Pin Numbers / Contact Numbers are not specified, since they will be different for MicroGigaCN and QSFP connectors. Please refer to relevant specifications for contact numbers used for signals listed in this column.

7.8.3.4 Low-Speed Electrical Contact/Signal Definitions

7.8.3.4.1 SDA, SCL

SCL is the clock of the two-wire serial interface, and SDA is the data for the 2-wire serial interface. Operation of this interface is described in detail in [Chapter 8: Management Interface](#), SCL and SDA must be pulled up in the host, through an pull-up resistor of value appropriate to the overall bus capacitance and the rise and fall time requirements listed in [Table 133 on page 482](#).

7.8.3.4.2 INT_L/RESET_L

Int_L/Reset_L is a bidirectional signal. When driven from the host, it operates logically as a Reset signal. When driven from the module, it operates logically as an Interrupt signal. In both cases, the signal is asserted low, as indicated by the '_L' suffix. The Int_L/Reset_L

signal requires open collector outputs in both the host and module, and must be pulled up on the host board, as described for SDA and SCL. Int_L and Reset_L indications are distinguished from each other by timing - a shorter assertion, driven by the module, indicates an interrupt, and a longer assertion of the signal, driven by the host, indicates a reset, as listed in [Table 135 on page 483](#).

Int_L operation: When Int_L/Reset_L is pulled “Low” by the module for longer than the minimum interrupt pulse width ($t_{Int_L,PW-min}$) and shorter than the maximum interrupt pulse width ($t_{Int_L,PW-max}$) the signal signifies an interrupt. An interrupt indicates a possible module operational fault or a module status critical to the host system. The host identifies the cause of the interrupt using the 2-wire serial interface. Int_L must operate in pulse mode (vs. static mode), in order to distinguish a short interrupt signal from a longer reset signal, so the module must de-assert Int_L/Reset_L after the interrupt has been signaled.

Reset_L operation: When the Int_L/Reset_L signal is pulled “Low” by the host for longer than the minimum reset pulse length ($t_{reset_L,PW-min}$), it initiates a complete module reset, returning all user module settings to their default state. There is no maximum reset pulse length. Module Reset Assert Time (t_{init}) starts on the rising edge after the low level on the Reset_L signal is released. During the execution of a reset (t_{init}) the host shall disregard all status bits until the module indicates a completion of the reset interrupt. The module indicates this by posting an Int_L signal with the Data_Not_Ready bit (Memory Map, Byte 2, bit 0) negated (set to 0). Note that on power up (including hot insertion) the module should post this completion of reset interrupt without requiring a reset from the host.

7.8.3.4.3 PRSNT_L

PRSNT_L is used to indicate when the module is plugged into the host receptacle. PRSNT_L is pulled up to Vcc3.3 on the host board through ≥ 50 kOhm. It is pulled down to signal common through 5.2 kOhm in modules requiring 12V power, and tied down to signal common either directly or through resistor of up to 100 Ohm, in modules requiring 3.3V power only. The PRSNT_L signal is asserted “Low” when inserted and deasserted “High” when the module is physically absent from the host connector.

7.8.3.5 HIGH-SPEED ELECTRICAL CONTACT/SIGNAL DEFINITIONS

With the exception of squelch behavior, the high-speed signaling for Tx and Rx (pluggable module data inputs from and outputs to the InfiniBand port, respectively) are as described in Section 6.6.4, “Host Driver Output Characteristics for QDR,” on page 283 and Section 6.7.7, “Host Receiver input characteristics for QDR,” on page 308.

For CXP modules, Rx Squelch, for loss of input signal from the link (Rx LOS), is required. In the event that the signal on any channel becomes equal to or less than the level required to assert Rx LOS (Receiver Loss of Signal), the receiver data output for that channel shall be squelched or disabled. In the squelched or disabled state, output impedance levels shall be maintained while the differential voltage swing shall be less than 50 mVpp. Rx squelch may optionally be disabled.

For CXP modules, Tx Squelch, for loss of input signal from the InfiniBand port, is an optional function. Where implemented it shall function as follows. In the event of the differential, peak-to-peak electrical signal on any channel becomes equal to or less than 50 mVpp, then the transmitter optical output for that channel shall be squelched or disabled and the associated Tx LOS flag set.

In InfiniBand links, the use of Rx Squelch and Tx Squelch may interfere with the correct operation of Beacon signaling. As described in Volume 2, Release 1.2.1, Section 5.6.4.2 Polling States and Section 5.6.4.3 Sleeping States, the beaconing sequence is a periodic repeating pattern, with an active TS1 transmission period of 2 ms and a quiescent period of 100 ms. Since Tx and Rx Squelch, if enabled, will activate during the 100 ms quiescent period, the time from resumption of Rx input and Tx output signals until normal signaling condition is reached must be well under 2 ms.

To meet this condition, the Rx Squelch Deassert Time (Time from resumption of Rx input signals until normal Rx output condition is reached) shall be less than 80 μ s (as specified in the QSFP+ SFF specification). The Tx Squelch Deassert time, if Tx Squelch is enabled, shall be less than 500 μ s. The Rx Squelch Assert Time and Tx Squelch Assert time shall be less than 80 μ s and 10 ms, respectively. These specifications are listed in [Table 136, “I/O Timing for Squelch and Disable,” on page 485](#). (Note that the QSFP+ SFF timing specifications are less stringent, specifying, for example a much longer Tx Squelch Deassert time, 400 ms, so Tx Squelch may generally be disabled).

7.8.3.5.1 TRANSMIT SIGNALS: Tx[0-11][P/N]

Tx[0-11][p/n] are CXP module transmitter data inputs. They are AC-coupled differential lines with 100 Ohm differential terminations inside the CXP module. The AC coupling is inside the CXP module and not required on the Host board.

Output squelch (Tx Squelch) for loss of input signal Tx LOS is an optional function. Where implemented it shall function as follows: In the event of the differential, peak-to-peak electrical signal on any lane becomes equal to or less than 50 mVpp, then the transmitter optical output for that lane shall be squelched and the associated Tx LOS flag set. For an optical transceiver with separable optical connector, the optical modulation amplitude (OMA) when squelched shall be less than or equal to -26 dBm, and, where practical, the average output power is recommended to be less than or equal to -26 dBm.

In normal operation, where Tx Squelch is implemented the default case has Tx Squelch active. Tx Squelch can be deactivated using Tx Squelch Disable through the 2-wire serial interface. Tx Squelch Disable is an optional function.

Implementation Note

Note that Tx Squelch can be implemented in several different ways: (1) No modulation (OMA less than or equal to a limit, but Tx average power within normal operating range, either at '0' power or at average of '0' and '1' power), or (2) No optical output power (Tx average power less than or equal to a limit). Method (2) is clearly more restrictive, since an output power limit also implies an OMA limit.

Revisions of this specification prior to 1.3.1 have only specified the limit on OMA. Revision 1.3.1 has added the recommended limit on Tx average power, since it is a more definitive way of indicating Tx Squelch, especially for CXP optical transceiver applications beyond the InfiniBand link protocol.

Note that optical transceivers built to Rev. 1.3 specifications may only limit OMA, rather than limiting average power also. Also, for some optical transceiver technologies, such as silicon photonics, it may be impractical or impossible to limit average optical power to -26 dBm (particularly on individual lanes).

Applications using optical transceivers should confirm Tx Squelch implementation method with selected vendors to assure interoperability.

7.8.3.5.2 RECEIVE SIGNALS: Rx[0-11][P/N]

Rx[0-11][p/n] are CXP module receiver data outputs. They are AC-coupled differential lines that should be terminated with 100 Ohm differential at the Host ASIC or SerDes. The AC coupling is inside the CXP module and not required on the Host board.

Output squelch for loss of input signal Rx Squelch is an optional function. Where implemented it shall function as follows: In the event of the optical or electrical signal on any physical lane becoming equal to or less than the level required to assert loss of signal (Rx LOS), then the receiver data output for that lane shall be squelched or disabled. In the squelched or disabled state, output impedance levels are maintained while the differential voltage swing shall be less than 50 mVpp. This voltage swing limit provides margin vs. the value specified for V_{RSD} , Signal Threshold for receiver signal detection listed in [Section 6.6.4. "Host Driver Output Characteristics for QDR." on page 290](#).

In normal operation, where Rx Squelch is implemented the default case has Rx Squelch active. Rx Squelch can be deactivated using Rx Squelch Disable through the 2-wire serial interface. Rx Squelch Disable is an optional function.

Implementation Note

Given that there are several possible implementations of Tx Squelch, applications using optical transceivers should confirm interoperability Tx Squelch and Rx Squelch implementations. See note above.

Implementation Note: LANE USAGE FOR 10-LANE INTERFACE:

If a CXP module or cable is being used for an interface other than InfiniBand which uses 10 lanes, e.g., for a 100 Gb Ethernet interface or proprietary 10-lane interface, the middle lanes should be active, as they are less likely to see stress than the outer lanes. That is, logical lanes 0 through 9 in a 10-lane interface should be implemented on physical contacts corresponding to lanes 1 through 10, with the outer two lanes (0 and 11) in each direction left unused.

Correspondingly, the mapping of control bits in [Chapter 8: Management Interface](#) to the lanes should remain invariant, e.g. with the Tx LOS flag registers, Lower Page Bytes 7 & 8, a 10 lane application should use Byte 7 bit 2 through Byte 8 bit 1, leaving locations Byte 7 bit 3 and Byte 8 bit 0 unused.

7.8.4 ENVIRONMENTAL AND THERMAL REQUIREMENTS

The CXP module and receptacle shall comply to the mechanical and environmental specifications described in [Table 113 on page 439](#). Connectors shall meet or exceed the environmental performance requirements of EIA-364-1000-2009, including exposure to Mixed Flowing Gas consistent with the required product life

7.8.5 THERMAL PERFORMANCE RANGES

The CXP module shall operate within one or more of the case temperatures ranges defined in [Table 114 on page 439](#). The temperature ranges are applicable between 60 m below sea level and 1800 m above sea level, (Ref. NEBS GR-63) utilizing the host system's designed airflow. CXP is designed to allow for up to 16 adjacent transceivers in a 19-inch rack-mount design using individual or ganged receptacles, with the appropriate thermal design for cooling / airflow. (Ref. NEBS GR-63).

Table 113 Module and Receptacle Mechanical and Environmental Requirements

Parameter	Specification	Test Condition
Vibration	No damage No discontinuity longer than 1 μ sec allowed. 20 mOhms maximum change from initial (baseline) contact resistance	EIA-364-28
Mechanical Shock	No damage 20 mOhms maximum change from initial (baseline) contact resistance	EIA-364-27
Thermal Shock	No Damage 20 mOhms maximum change from initial (baseline) contact resistance	EIA-364-32C, Condition 1 -55°C to +85°C
Temperature Life	No Damage 20 mOhms maximum change from initial (baseline) contact resistance	EIA-364-17, Method A Test Condition 2, Test Time Condition C Subject mated specimens to 70°C for 500 hours
Humidity-Temperature Cycling	No Damage 20 mOhms maximum change from initial (baseline) contact resistance	EIA-364-31, Method III Subject unmated specimens to 10 cycles (10 days) between 25°C and 65° at 80-100% RH
Mixed Flowing Gas	No Damage 20 mOhms maximum change from initial (baseline) contact resistance	EIA-364-65, Class 2A Subject specimens to environments Class 2A, 7 days unmated and 7 days mated
Thermal Disturbance	No Damage 20 mOhms maximum change from initial (baseline) contact resistance	EIA-364-32 Cycle the connector between 15±3°C and 85±3°C as measured on part. Temperature ramps should be a minimum of 2°C per minute and dwell times should ensure that the contacts reach the temperature extremes (a minimum of 5 minutes). Humidity is not controlled. Perform 10 such cycles.

Table 114 Temperature Classification of Module Case

Class	Case Temperature Range During Operation
Standard	0 through 70 C
Extended	-5 through 85 C
Industrial	-40 through 85 C

7.8.6 CXP HOUSING ASSEMBLY THERMAL INTERFACES

Cooling requirements will be dependent on device technology. For devices or cables that require cooling inside the host, a receptacle housing with riding heat sink is shown in

[Figure 150.](#)

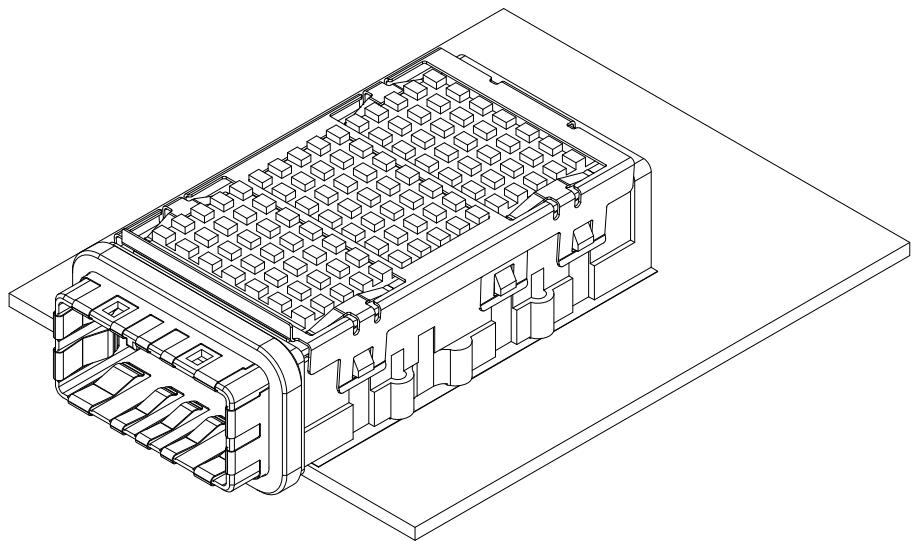


Figure 150 Receptacle Housing with Integrated Riding Heat Sink

The thermal interface locations of the connector plug are shown in [Figure 142](#). The dimensions for an example heat sink are shown in [Figure 143 on page 425](#). Dimensions of the heat sink clip attachment points on the receptacle housing and for a heat sink clip are shown in [Figure 142 on page 424](#) and [Figure 143 on page 425](#), so that heat sinks appropriate to specific thermal environments or transmission media can be built. Use of a modified heat sink would require the use of a suitably-modified clip.

7.8.7 MATING OF CXP MODULE & Host PCBs TO CXP ELECTRICAL CONNECTORS

The cards and other components inside the CXP modules will require careful design to support 10Gb/s signaling on (12+12) differential pairs. Similarly, high-speed traces in host PCBs must be carefully designed to minimize impedance discontinuities and reduce losses to acceptable levels. These designs are outside the bounds of this specification.

7.8.7.1 HOST BOARD SCHEMATIC

[Figure 151](#) shows an example of a host board schematic for CXP, with connections to host SERDES and control logic. For EMI protection the signals to the connector should be turned off when the CXP transceiver is removed. The use of good high speed printed circuit board design practices is recommended. The chassis ground (case common) of the CXP module should not be directly connected to the module's signal ground.

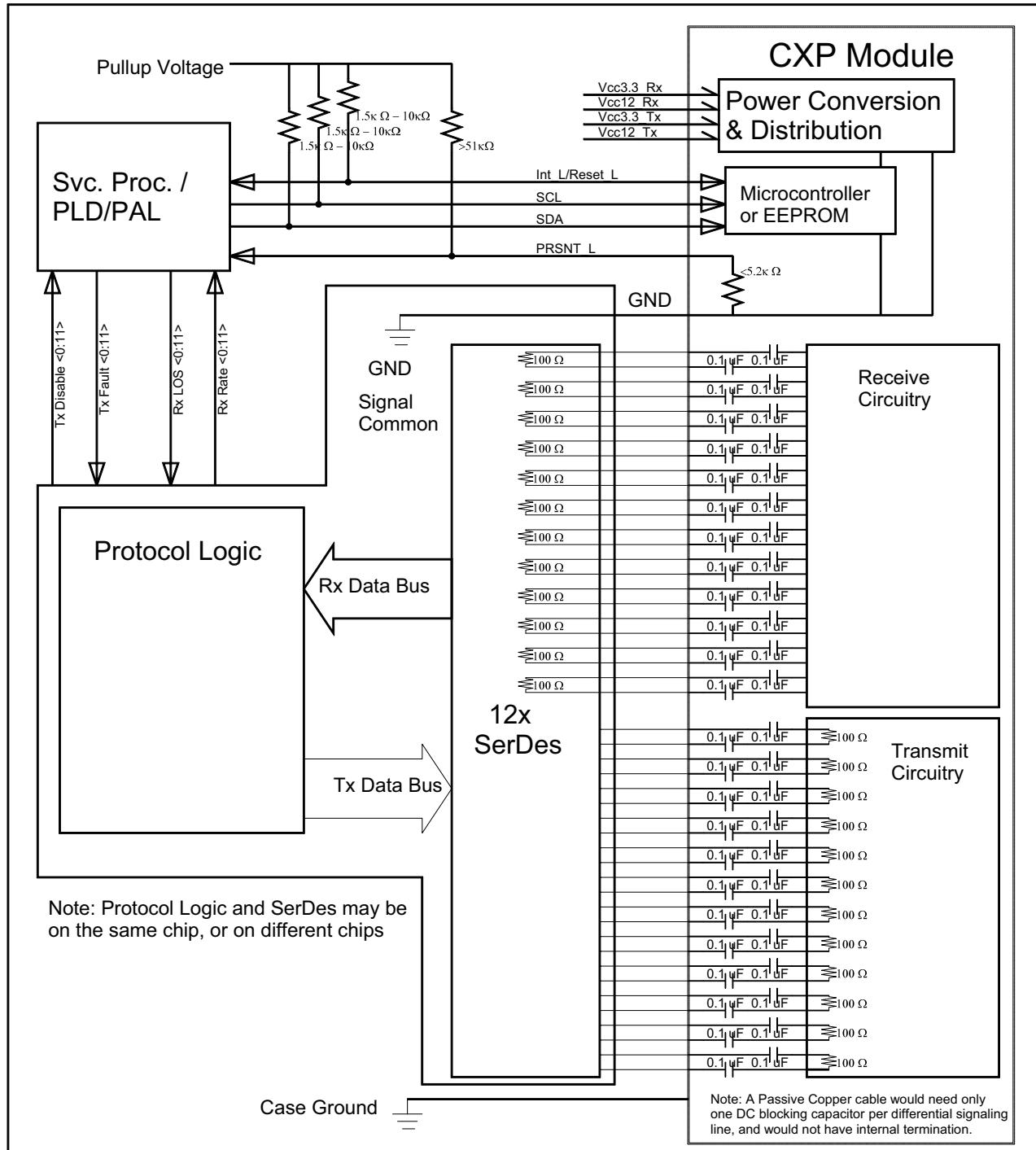


Figure 151 Example CXP host board schematic

Note that DC blocking capacitors on the high-speed signal wires are implemented inside the module or cable as described in [Section 6.8.1, “DC Blocking,” on page 323.](#)

7.8.7.2 POWER REQUIREMENTS

Power for the module is supplied through 4 contacts: Vcc3.3-Rx, Vcc12-Rx, Vcc3.3-Tx and Vcc12-Tx. Power is applied concurrently to them.

Since different classes of modules exist with pre-defined maximum power consumption limits, it is necessary to avoid exceeding the system power supply limits and cooling capacity when a module is inserted into a system designed to only accommodate lower power modules. It is recommended that the host, through the management interface, identify the power consumption class of the module before allowing the module to go into high power mode. See [Section 7.8.7.5 on page 444.](#)

A host board together with the CXP module(s) forms an integrated power system. The host supplies stable power to the module. The module limits electrical noise coupled back into the host system and limits inrush charge/current during hot plug insertion.

All specifications shall be met at the maximum power supply current. No power sequencing of the power supply is required of the host system since the module sequences the contacts in the order of ground, supply and signals during insertion.

7.8.7.3 HOST POWER SUPPLY FILTERING

The host board should use the power supply filtering network shown in [Figure 152 on page 443](#), or an equivalent. Any voltage drop across a filter network on the host is counted against the host DC set point accuracy specification. Inductors with DC Resistance of less than 0.1Ω should be used in order to maintain the required voltage at the Host Edge Card Connector. Selection of the filter capacitors (shown as 10 uF in [Figure 152](#)) is left to the implementer, but it is recommended that 10 uF capacitors with approximately 0.33 Ohms of Equivalent Series Resistance be used to provide adequate filtering without high frequency ringing. The time constant of the filter circuit is the important quantity.

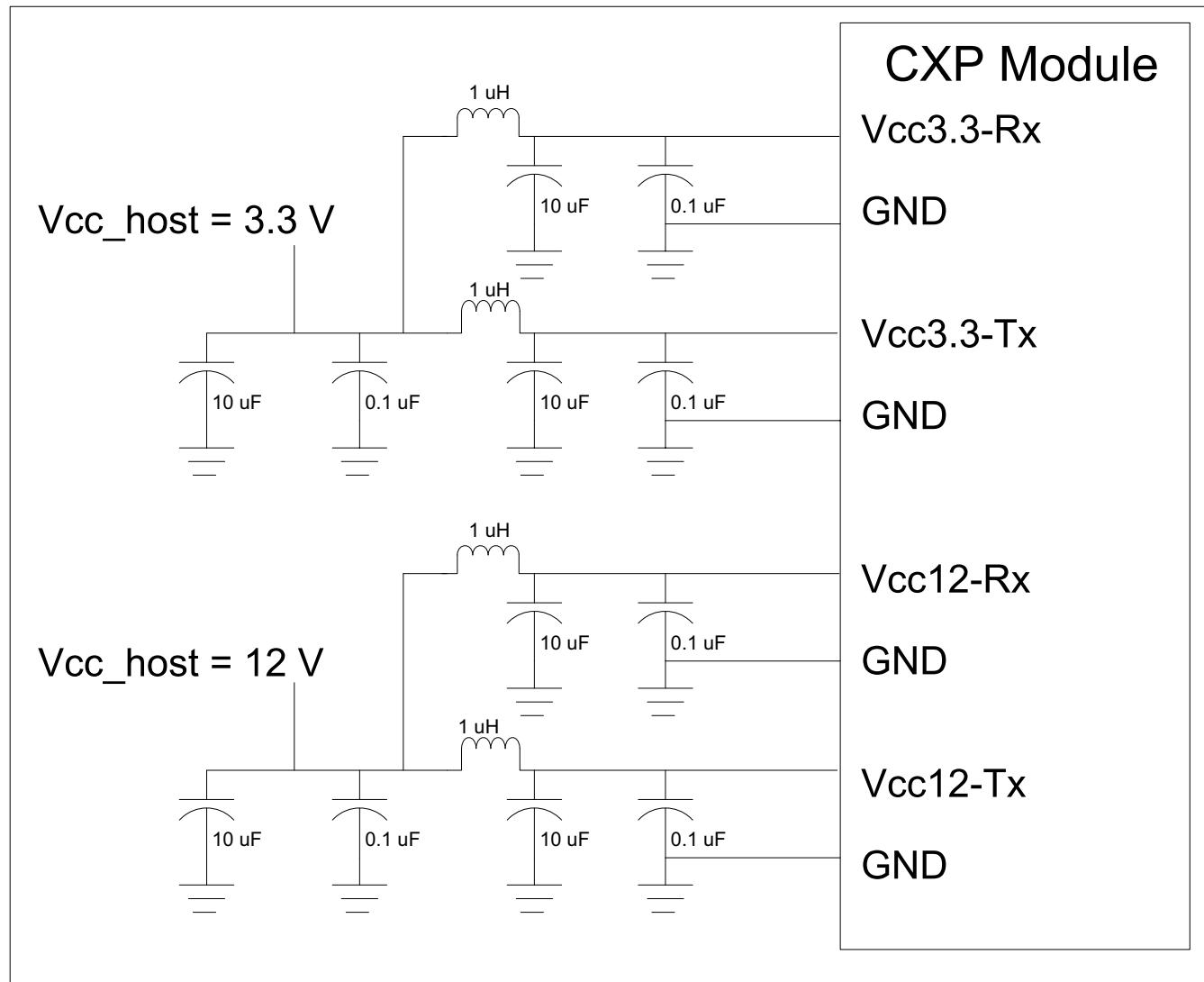


Figure 152 Recommended CXP Host Board Power Supply Filtering

7.8.7.4 HOST POWER SUPPLY SPECIFICATIONS

The specification for the power supply is shown in [Table 115](#).

Table 115 Power Supply Specification

Parameter	Min.	Nominal	Max.	Unit	Condition
Vcc12_host		12			Measured at Vcc12-Tx and Vcc12-Rx
Vcc12 set point accuracy	-5		+5	%	Measured at Vcc12-Tx and Vcc12-Rx.
Vcc12 Power supply noise including ripple			50	mVpp	1 kHz to frequency of operation
Vcc3.3_host		3.3		V	Measured at Vcc3.3-Tx and Vcc3.3-Rx
Vcc3.3 set point accuracy	-5		+5	%	Measured at Vcc3.3-Tx and Vcc3.3-Rx. Note ^a
Vcc3.3 Power supply noise including ripple			50	mVpp	1 kHz to frequency of operation
Module Maximum Current Inrush			1.25	A	On any contact.
Module Current Ramp Rate			100	mA/uS	

a. 5%-accurate power needed for VCSEL laser drivers

7.8.7.5 POWER BUDGET CLASSES

Power levels associated with classifications of modules are shown in [Table 116](#). In general, the higher power classification levels are associated with higher data rates and longer reach, for a particular technology family.

Table 116 Power Budget Classification

Power Class	Max. Power (W)	Power Class	Max. Power (W)
0	0.25 or less	1	1.0 or less
2	1.5 or less	3	2.5 or less
4	4.0 or less	5	6.0 or less
6	Higher than 6 Watts	7	Reserved

Power Class 0 supports a management-interface-only power level, for devices such as passive copper cables which require little or no signal power.

Power Classes 1 through 5 describe devices with between 0.25W and 6.0 Watts, with roughly a factor of 1.5 differentiating each power class.

Power Class 6 (higher than 6 Watts), is intended as a special high-power class, allowing higher-power devices and cables to operate with more than 6 Watts only after negotiation with a host that can support the cooling and power-delivery infrastructure to support such devices. As described in the Memory map, at Byte 42 in the Tx lower page and Byte 148 in Upper Page 00h, a device may not draw more than 6 Watts unless actively allowed by a host system. The actual amount of power used by such a high-power device or cable is described in the Read-only Byte 148 of Upper Page 00h.

The highest maximum power budget is determined by a current limit of 1.0 A for each power contact, and by the cooling capability provided by the system. Two contacts at each voltage level allow power supply of up to 6.6W of power at 3.3V, and 24W of power at 12V. Generally, cooling capability will limit the amount of power that a module may dissipate. The system designer is responsible for ensuring that the maximum temperature does not exceed the case temperature requirements.

7.8.7.6 ESD REQUIREMENTS

The module high speed signal contacts shall withstand 1000 V electrostatic discharge using the Human Body Model module and all other contacts shall withstand 2000 V electrostatic discharge using the Human Body Model, per JEDEC Standard JESD22-A114B (March 2006). The module power and ground contacts shall withstand 500 V using the charged device model, per JEDEC Standard JESD22-C101E (Dec. 2009). High speed and management interface contacts do not have charged device model ESD requirements, since they are recessed. The module shall meet ESD requirements given in EN 61000-4-2, criterion B test specification, such that when installed in a properly grounded housing and chassis the unit is subjected to 15 kV air discharges during operation and 8 kV direct contact discharges to the case. For all three tests, the module shall withstand these discharge levels without damage.

7.8.7.7 HOT INSERTION AND REMOVAL

CXP modules shall not be damaged by removal or insertion while power is applied. Removal may occur while the link is operating without damage to either the port or the link.

7.8.8 CXP MEMORY MAP

The memory map for CXP devices is described in 8.7.

7.9 ELECTRICAL INTERFACE CABLE ASSEMBLIES

This section defines the characteristics of InfiniBand cables that utilize an electrical interface to the host device. It is the intent of this specification to allow for and encourage innovation in the market. Therefore, the emphasis is on interoperability and minimum requirements needed to insure functionality without restricting the means of implementation. Specific bulk wire conductor size, insulation materials, etc. are explicitly not specified herein, allowing suppliers to determine the optimum design for a given application.

7.9.1 PHYSICAL REQUIREMENTS

This section defines the bend radius and other attributes of InfiniBand cables. The bend radius is the radius to which the cable can be bent while continuing to meet the electrical characteristics as defined in [6.8 Compliant Channels](#).

7.9.1.1 MICROGIGACN INTERFACE CABLE KEYING

It is recommended that cables using the microGigaCN interface for InfiniBand use keys as shown in [Figure 134 on page 403](#).

7.9.1.2 MICROGIGACN INTERFACE CABLE BEND RADIUS

It is recommended that cable assemblies with a microGigaCN interface to be used for InfiniBand products have a minimum bend radius of no more than 100 mm (4 inches), as shown in [Figure 153](#).

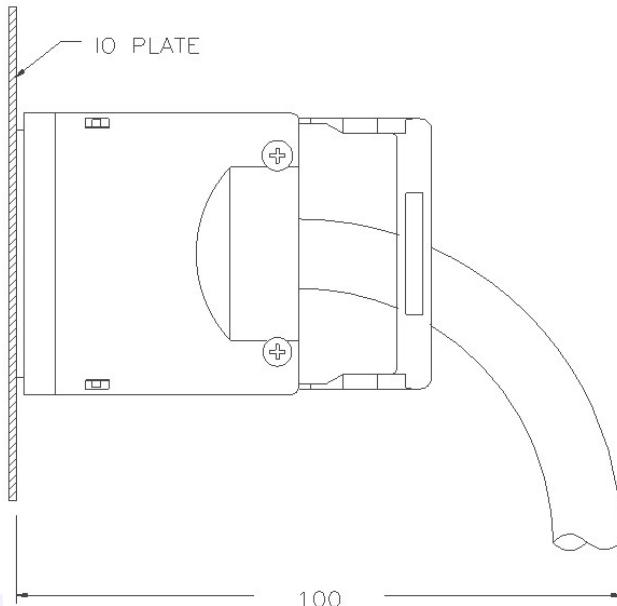


Figure 153 microGigaCN cable assembly bend radius

7.9.1.3 QSFP+ INTERFACE CABLE BEND RADIUS

It is recommended that QSFP+ cables be bent no tighter than the limits shown in [Figure 154](#) and listed in [Table 117](#).

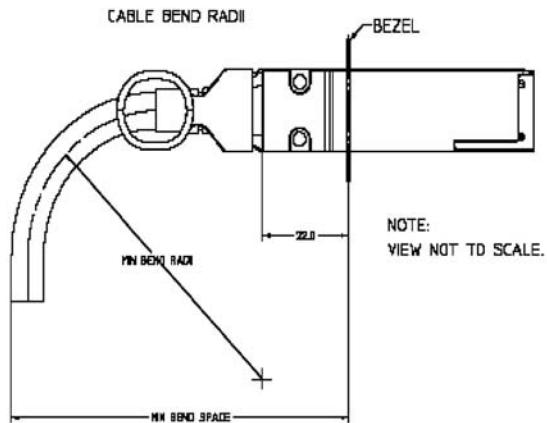


Figure 154 QSFP+ cable assembly bend radius

Table 117 QSFP+ bend radius limits

Primary signal conductor size, AWG	Outer diameter, mm	Min. bend radius, mm (from bezel)
24	9.5	100
26	8.25	95
28	7.0	87
30	6.0	80

The AWG 30 bend radius limit is suggested for fiber optic cables, with QSFP+ transceivers.

7.9.1.4 CXP INTERFACE CABLE BEND RADIUS

It is recommended that CXP cables be bent no tighter than 100 mm for copper cables, and 90 mm for optical cables as shown in [Figure 155 on page 448](#).

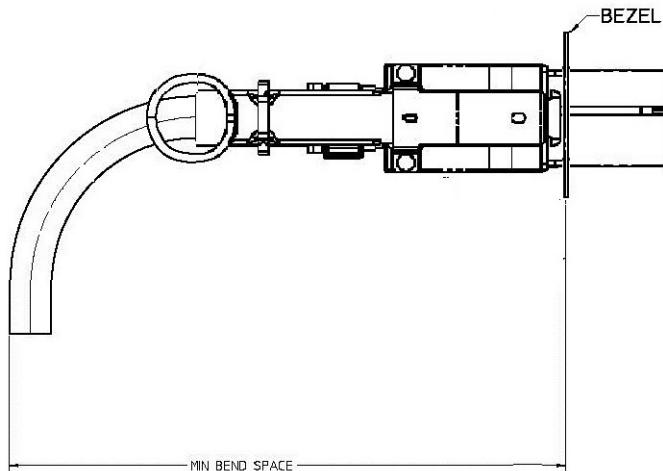


Figure 55
CXP cable assembly bend radius

7.9.2 CABLE SHIELDING

C7-17: The **IB_Sh_Ret** connections are specified on the cable connector to support the isolation of the high speed differential inputs and outputs. These shield returns shall be connected to logic ground on the module.

Implementation Note

The primary purpose of these connections is to provide for isolation of the differential signals from each other. These shields also help to insure that the desired impedance of the link is maintained.

7.9.2.1 CONNECTOR SHIELDING

A metal backshell or other means which fully shields the connector shall be bonded to the cable bulk shield through a continuous 360 degree contact to minimize EMI (Electromagnetic Interference).

7.9.2.2 INNER (SIGNAL PAIR OR QUAD) SHIELD

C7-18: The **IB_Sh_Ret** signals shall be connected to the cable inner shield(s) in the cable connector.

C7-19: A continuous ground path from the cable's inner shield(s) through the connector to the board signal ground shall be provided to insure low jitter, low crosstalk and EMI containment.

Note

The primary purpose of these connections is to provide for isolation of the differential signals from each other. These shields also help to insure that the desired impedance of the link is maintained.

A cable connector or receptacle constructed with only a bulk shield is not likely to meet the electrical requirements.

7.9.2.3 OUTER (BULK) SHIELD

C7-20: The cable bulk shield used over the outside of the collection of conductors is used for EMI control and shall be connected to chassis ground at both ends as defined in [7.9.2 on page 448](#).

This specification does not permit AC coupling of the bulk shield.

7.9.3 1X INTERFACE CABLE

This interface is obsolete. Refer to volume 2, version 1.2.1 of the InfiniBand specification for information.

7.9.4 MICROGIGACN INTERFACE CABLES

7.9.4.1 4X MICROGIGACN INTERFACE PASSIVE CABLES

C7-21: The pin assignment listed in [Table 118 on page 450](#) shall be used for passive InfiniBand 4X cables.

There are no unused pins in the cable plug.

Note

Cable signal nets shall be connected from each *transmit* signal pair on the source port to the appropriate *receive* signal pair on the desired destination port. For example, a cable from port y to port z should be wired to connect IBtyOp(0) (positive output signal from port y, bit 0), to IBtzIp(0) (positive input to port z, bit 0) and IBtyOn(0) to IBtzIn(0), as shown in [Table 118](#).

Table 118 4X passive cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
G1-G9	Signal Ground	G1-G9	Signal Ground
S1	IBtxIp(0)	S16	IBtxOp(0)
S2	IBtxIn(0)	S15	IBtxOn(0)
S3	IBtxIp(1)	S14	IBtxOp(1)
S4	IBtxIn(1)	S13	IBtxOn(1)
S5	IBtxIp(2)	S12	IBtxOp(2)
S6	IBtxIn(2)	S11	IBtxOn(2)
S7	IBtxIp(3)	S10	IBtxOp(3)
S8	IBtxIn(3)	S9	IBtxOn(3)
S9	IBtxOn(3)	S8	IBtxIn(3)
S10	IBtxOp(3)	S7	IBtxIp(3)
S11	IBtxOn(2)	S6	IBtxIn(2)
S12	IBtxOp(2)	S5	IBtxIp(2)
S13	IBtxOn(1)	S4	IBtxIn(1)
S14	IBtxOp(1)	S3	IBtxIp(1)
S15	IBtxOn(0)	S2	IBtxIn(0)
S16	IBtxOp(0)	S1	IBtxIp(0)
Housing	Chassis Ground	Housing	Chassis Ground

C7-22: Signal Ground shall not be connected to Chassis Ground in the cable or cable connector. See [Section 7.9.2](#).

7.9.4.2 4X MICROGIGACN INTERFACE ACTIVE CABLES

Note

A 4X active cable used with InfiniBand version 1 (unpowered) board connector pinout will not function.

C7-22.2.1: The pin assignment listed in [Table 119 on page 451](#) shall be used for active InfiniBand 4X cables.

C7-22.2.2: V_{cc} shall not be wired through the cable to the other end.

The 12 V or 3.3 V Sense signal is used to enable the respective voltage on the V_{cc} power supply pin used to provide power to the active components in the cable. Both signals are considered active if their voltage level is between 0.9 V and 2.4 V DC.

C7-22.1.3: Both 12 V and 3.3 V Sense signals shall not be active simultaneously.**Note**

Cable signal nets shall be connected from each *transmit* signal pair on the source port to the appropriate *receive* signal pair on the desired destination port. For example, a cable from port y to port z should be wired to connect IBtxOp(0) (positive output signal from port y, bit 0), to IBtzIp(0) (positive input to port z, bit 0) and IBtxOn(0) to IBtzIn(0), as shown in [Table 119](#).

Table 119 4X active cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
G1	Sense-12V	G9	Signal Ground
G4-G6	Signal Ground	G4-G6	Signal Ground
S1	IBtxIp(0)	S16	IBtxOp(0)
S2	IBtxIn(0)	S15	IBtxOn(0)
G2	Signal Ground	G8	V_{cc}
S3	IBtxIp(1)	S14	IBtxOp(1)
S4	IBtxIn(1)	S13	IBtxOn(1)
G3	Signal Ground	G7	Sense-3.3V
S5	IBtxIp(2)	S12	IBtxOp(2)
S6	IBtxIn(2)	S11	IBtxOn(2)
S7	IBtxIp(3)	S10	IBtxOp(3)
S8	IBtxIn(3)	S9	IBtxOn(3)
S9	IBtxOn(3)	S8	IBtxIn(3)
S10	IBtxOp(3)	S7	IBtxIp(3)
S11	IBtxOn(2)	S6	IBtxIn(2)
S12	IBtxOp(2)	S5	IBtxIp(2)

Table 119 4X active cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
G7	Sense-3.3V	G3	Signal Ground
S13	IBtxOn(1)	S4	IBtxIn(1)
S14	IBtxOp(1)	S3	IBtxIp(1)
G8	Vcc	G2	Signal Ground
S15	IBtxOn(0)	S2	IBtxIn(0)
S16	IBtxOp(0)	S1	IBtxIp(0)
G9	Signal Ground	G1	Sense-12V
Housing	Chassis Ground	Housing	Chassis Ground

7.9.4.3 8X MICROGIGACN INTERFACE PASSIVE CABLES

The 8X cables provide for simultaneous transmit and receive of eight bits of encoded differential data. They use the same cable connector as the 12X microGigaCN copper cable interface, but with the pin assignments listed in [Table 120](#). Sixteen pairs of signals are used, eight each for transmit and receive.

Table 120 8X passive cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
G1-25	Signal Ground	G1-25	Signal Ground
S1	IBtxIp(0)	S48	IBtxOp(0)
S2	IBtxIn(0)	S47	IBtxOn(0)
S3	IBtxIp(1)	S46	IBtxOp(1)
S4	IBtxIn(1)	S45	IBtxOn(1)
S5	IBtxIp(2)	S44	IBtxOp(2)
S6	IBtxIn(2)	S43	IBtxOn(2)
S7	IBtxIp(3)	S42	IBtxOp(3)
S8	IBtxIn(3)	S41	IBtxOn(3)
S9	IBtxIp(4)	S40	IBtxOp(4)
S10	IBtxIn(4)	S39	IBtxOn(4)

Table 120 8X passive cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
S11	IBtxIp(5)	S38	IBtxOp(5)
S12	IBtxIn(5)	S37	IBtxOn(5)
S13	IBtxIp(6)	S36	IBtxOp(6)
S14	IBtxIn(6)	S35	IBtxOn(6)
S15	IBtxIp(7)	S34	IBtxOp(7)
S16	IBtxIn(7)	S33	IBtxOn(7)
S17	reserved	S32	reserved
S18	reserved	S31	reserved
S19	reserved	S30	reserved
S20	reserved	S29	reserved
S21	reserved	S28	reserved
S22	reserved	S27	reserved
S23	reserved	S26	reserved
S24	reserved	S25	reserved
S25	reserved	S24	reserved
S26	reserved	S23	reserved
S27	reserved	S22	reserved
S28	reserved	S21	reserved
S29	reserved	S20	reserved
S30	reserved	S19	reserved
S31	reserved	S18	reserved
S32	reserved	S17	reserved
S33	IBtxOn(7)	S16	IBtxIn(7)
S34	IBtxOp(7)	S15	IBtxIp(7)
S35	IBtxOn(6)	S14	IBtxIn(6)
S36	IBtxOp(6)	S13	IBtxIp(6)
S37	IBtxOn(5)	S12	IBtxIn(5)
S38	IBtxOp(5)	S11	IBtxIp(5)
S39	IBtxOn(4)	S10	IBtxIn(4)

Table 120 8X passive cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
S40	IBtxOp(4)	S9	IBtxIp(4)
S41	IBtxOn(3)	S8	IBtxIn(3)
S42	IBtxOp(3)	S7	IBtxIp(3)
S43	IBtxOn(2)	S6	IBtxIn(2)
S44	IBtxOp(2)	S5	IBtxIp(2)
S45	IBtxOn(1)	S4	IBtxIn(1)
S46	IBtxOp(1)	S3	IBtxIp(1)
S47	IBtxOn(0)	S2	IBtxIn(0)
S48	IBtxOp(0)	S1	IBtxIp(0)
Housing	Chassis Ground	Housing	Chassis Ground

7.9.4.4 8X MICROGIGACN INTERFACE ACTIVE CABLES

Note

A 8X active cable used with InfiniBand version 1 (unpowered) board connector pinout will not function.

C7-22.2.4: The pin assignment listed in [Table 121 on page 455](#) shall be used for active InfiniBand 8X cables.

C7-22.2.5: V_{cc} shall not wired through the cable to the other end.

The 12 V or 3.3 V Sense signal is used to enable the respective voltage on the V_{cc} power supply pin used to provide power to the active components in the cable. Both signals are considered active if their voltage level is between 0.9 V and 2.4 V DC.

C7-22.1.6: Both 12 V and 3.3 V Sense signals shall not be active simultaneously.

Note

Cable signal nets shall be connected from each *transmit* signal pair on the source port to the appropriate *receive* signal pair on the desired destination port. For example, a cable from port y to port z should be wired to connect IBtxOp(0) (positive output signal from port y, bit 0), to IBtzIp(0) (positive input to port z, bit 0) and IBtxOn(0) to IBtzIn(0), as shown in [Table 121](#).

Table 121 8X active cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
G1	Sense-12V	G25	Signal Ground
G4-10, G12-22	Signal Ground	G4-10, G12-22	Signal Ground
S1	IBtxIp(0)	S48	IBtxOp(0)
S2	IBtxIn(0)	S47	IBtxOn(0)
G2	Signal Ground	G24	Vcc
S3	IBtxIp(1)	S46	IBtxOp(1)
S4	IBtxIn(1)	S45	IBtxOn(1)
G3	Vcc	G23	Sense-3.3V
S5	IBtxIp(2)	S44	IBtxOp(2)
S6	IBtxIn(2)	S43	IBtxOn(2)
S7	IBtxIp(3)	S42	IBtxOp(3)
S8	IBtxIn(3)	S41	IBtxOn(3)
S9	IBtxIp(4)	S40	IBtxOp(4)
S10	IBtxIn(4)	S39	IBtxOn(4)
S11	IBtxIp(5)	S38	IBtxOp(5)
S12	IBtxIn(5)	S37	IBtxOn(5)
S13	IBtxIp(6)	S36	IBtxOp(6)
S14	IBtxIn(6)	S35	IBtxOn(6)
S15	IBtxIp(7)	S34	IBtxOp(7)
S16	IBtxIn(7)	S33	IBtxOn(7)

Table 121 8X active cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
S17	reserved	S32	reserved
S18	reserved	S31	reserved
S19	reserved	S30	reserved
S20	reserved	S29	reserved
G11	Vcc	G16	Signal Ground
S21	reserved	S28	reserved
S22	reserved	S27	reserved
S23	reserved	S26	reserved
S24	reserved	S25	reserved
S25	reserved	S24	reserved
S26	reserved	S23	reserved
S27	reserved	S22	reserved
S28	reserved	S21	reserved
S29	reserved	S20	reserved
S30	reserved	S19	reserved
G16	Signal Ground	G11	Vcc
S31	reserved	S18	reserved
S32	reserved	S17	reserved
S33	IBtxOn(7)	S16	IBtxIn(7)
S34	IBtxOp(7)	S15	IBtxIp(7)
S35	IBtxOn(6)	S14	IBtxIn(6)
S36	IBtxOp(6)	S13	IBtxIp(6)
S37	IBtxOn(5)	S12	IBtxIn(5)
S38	IBtxOp(5)	S11	IBtxIp(5)
S39	IBtxOn(4)	S10	IBtxIn(4)
S40	IBtxOp(4)	S9	IBtxIp(4)
S41	IBtxOn(3)	S8	IBtxIn(3)
S42	IBtxOp(3)	S7	IBtxIp(3)
S43	IBtxOn(2)	S6	IBtxIn(2)

Table 121 8X active cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
S44	IBtxOp(2)	S5	IBtxIp(2)
G23	Sense-3.3V	G3	Vcc
S45	IBtxOn(1)	S4	IBtxIn(1)
S46	IBtxOp(1)	S3	IBtxIp(1)
G24	Vcc	G2	Signal Ground
S47	IBtxOn(0)	S2	IBtxIn(0)
S48	IBtxOp(0)	S1	IBtxIp(0)
G25	Signal Ground	G1	Sense-12V
Housing	Chassis Ground	Housing	Chassis Ground

7.9.4.5 12X MICROGIGACN INTERFACE PASSIVE CABLES

C7-23: The pin assignment listed in [Table 122](#) shall be used for passive InfiniBand 12X cables.

There are no unused pins in the cable plug.

Note

Cable signal nets shall be connected from each *transmit* signal pair on the source port to the appropriate *receive* signal pair on the desired destination port. For example, a cable from port y to port z should be wired to connect IBtyOp(0) (positive output signal from port y, bit 0), to IBtzIp(0) (positive input to port z, bit 0) and IBtyOn(0) to IBtzIn(0), as shown in [Table 122](#).

Table 122 12X passive cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
G1-G25	Signal Ground	G1-G25	Signal Ground
S1	IBtxIp(0)	S48	IBtxOp(0)
S2	IBtxIn(0)	S47	IBtxOn(0)

Table 122 12X passive cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
S3	IBtxIp(1)	S46	IBtxOp(1)
S4	IBtxIn(1)	S45	IBtxOn(1)
S5	IBtxIp(2)	S44	IBtxOp(2)
S6	IBtxIn(2)	S43	IBtxOn(2)
S7	IBtxIp(3)	S42	IBtxOp(3)
S8	IBtxIn(3)	S41	IBtxOn(3)
S9	IBtxIp(4)	S40	IBtxOp(4)
S10	IBtxIn(4)	S39	IBtxOn(4)
S11	IBtxIp(5)	S38	IBtxOp(5)
S12	IBtxIn(5)	S37	IBtxOn(5)
S13	IBtxIp(6)	S36	IBtxOp(6)
S14	IBtxIn(6)	S35	IBtxOn(6)
S15	IBtxIp(7)	S34	IBtxOp(7)
S16	IBtxIn(7)	S33	IBtxOn(7)
S17	IBtxIp(8)	S32	IBtxOp(8)
S18	IBtxIn(8)	S31	IBtxOn(8)
S19	IBtxIp(9)	S30	IBtxOp(9)
S20	IBtxIn(9)	S29	IBtxOn(9)
S21	IBtxIp(10)	S28	IBtxOp(10)
S22	IBtxIn(10)	S27	IBtxOn(10)
S23	IBtxIp(11)	S26	IBtxOp(11)
S24	IBtxIn(11)	S25	IBtxOn(11)
S25	IBtxOn(11)	S24	IBtxIn(11)
S26	IBtxOp(11)	S23	IBtxIp(11)
S27	IBtxOn(10)	S22	IBtxIn(10)
S28	IBtxOp(10)	S21	IBtxIp(10)
S29	IBtxOn(9)	S20	IBtxIn(9)
S30	IBtxOp(9)	S19	IBtxIp(9)
S31	IBtxOn(8)	S18	IBtxIn(8)

Table 122 12X passive cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
S32	IBtxOp(8)	S17	IBtxIp(8)
S33	IBtxOn(7)	S16	IBtxIn(7)
S34	IBtxOp(7)	S15	IBtxIp(7)
S35	IBtxOn(6)	S14	IBtxIn(6)
S36	IBtxOp(6)	S13	IBtxIp(6)
S37	IBtxOn(5)	S12	IBtxIn(5)
S38	IBtxOp(5)	S11	IBtxIp(5)
S39	IBtxOn(4)	S10	IBtxIn(4)
S40	IBtxOp(4)	S9	IBtxIp(4)
S41	IBtxOn(3)	S8	IBtxIn(3)
S42	IBtxOp(3)	S7	IBtxIp(3)
S43	IBtxOn(2)	S6	IBtxIn(2)
S44	IBtxOp(2)	S5	IBtxIp(2)
S45	IBtxOn(1)	S4	IBtxIn(1)
S46	IBtxOp(1)	S3	IBtxIp(1)
S47	IBtxOn(0)	S2	IBtxIn(0)
S48	IBtxOp(0)	S1	IBtxIp(0)
Housing	Chassis Ground	Housing	Chassis Ground

C7-24: Signal Ground shall not be connected to Chassis Ground in the cable or cable connector. See Section [7.9.2 on page 448](#).

7.9.4.6 12X MICROGIGACN INTERFACE ACTIVE CABLES

Note

A 12X active cable used with InfiniBand version 1 (unpowered) board connector pinout will not function.

C7-24.2.1: The pin assignment listed in [Table 123](#) shall be used for active InfiniBand 12X cables.

C7-24.2.2: V_{cc} shall not be wired through the cable to the other end.

The 12 V or 3.3 V Sense signal is used to enable the respective voltage on the V_{cc} power supply pin used to provide power to the active components in the cable. Both signals are considered active if their voltage level is between 0.9 V and 2.4 V DC.

C7-24.1.3: Both 12 V and 3.3 V Sense signals shall not be active simultaneously.**Note**

Cable signal nets shall be connected from each *transmit* signal pair on the source port to the appropriate *receive* signal pair on the desired destination port. For example, a cable from port y to port z should be wired to connect IBtxOp(0) (positive output signal from port y, bit 0), to IBtzIp(0) (positive input to port z, bit 0) and IBtxOn(0) to IBtzIn(0), as shown in [Table 123](#).

Table 123 12X active cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
G1	Sense-12V	G25	Signal Ground
G4-10, G12-22	Signal Ground	G4-10, G12-22	Signal Ground
S1	IBtxIp(0)	S48	IBtxOp(0)
S2	IBtxIn(0)	S47	IBtxOn(0)
G2	Signal Ground	G24	V_{cc}
S3	IBtxIp(1)	S46	IBtxOp(1)
S4	IBtxIn(1)	S45	IBtxOn(1)
G3	V_{cc}	G23	Sense-3.3V
S5	IBtxIp(2)	S44	IBtxOp(2)
S6	IBtxIn(2)	S43	IBtxOn(2)
S7	IBtxIp(3)	S42	IBtxOp(3)
S8	IBtxIn(3)	S41	IBtxOn(3)
S9	IBtxIp(4)	S40	IBtxOp(4)
S10	IBtxIn(4)	S39	IBtxOn(4)
S11	IBtxIp(5)	S38	IBtxOp(5)
S12	IBtxIn(5)	S37	IBtxOn(5)

Table 123 12X active cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
S13	IBtxIp(6)	S36	IBtxOp(6)
S14	IBtxIn(6)	S35	IBtxOn(6)
S15	IBtxIp(7)	S34	IBtxOp(7)
S16	IBtxIn(7)	S33	IBtxOn(7)
S17	IBtxIp(8)	S32	IBtxOp(8)
S18	IBtxIn(8)	S31	IBtxOn(8)
S19	IBtxIp(9)	S30	IBtxOp(9)
S20	IBtxIn(9)	S29	IBtxOn(9)
G11	Vcc	G16	Signal Ground
S21	IBtxIp(10)	S28	IBtxOp(10)
S22	IBtxIn(10)	S27	IBtxOn(10)
S23	IBtxIp(11)	S26	IBtxOp(11)
S24	IBtxIn(11)	S25	IBtxOn(11)
S25	IBtxOn(11)	S24	IBtxIn(11)
S26	IBtxOp(11)	S23	IBtxIp(11)
S27	IBtxOn(10)	S22	IBtxIn(10)
S28	IBtxOp(10)	S21	IBtxIp(10)
S29	IBtxOn(9)	S20	IBtxIn(9)
S30	IBtxOp(9)	S19	IBtxIp(9)
G16	Signal Ground	G11	Vcc
S31	IBtxOn(8)	S18	IBtxIn(8)
S32	IBtxOp(8)	S17	IBtxIp(8)
S33	IBtxOn(7)	S16	IBtxIn(7)
S34	IBtxOp(7)	S15	IBtxIp(7)
S35	IBtxOn(6)	S14	IBtxIn(6)
S36	IBtxOp(6)	S13	IBtxIp(6)
S37	IBtxOn(5)	S12	IBtxIn(5)
S38	IBtxOp(5)	S11	IBtxIp(5)
S39	IBtxOn(4)	S10	IBtxIn(4)

Table 123 12X active cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
S40	IBtxOp(4)	S9	IBtxIp(4)
S41	IBtxOn(3)	S8	IBtxIn(3)
S42	IBtxOp(3)	S7	IBtxIp(3)
S43	IBtxOn(2)	S6	IBtxIn(2)
S44	IBtxOp(2)	S5	IBtxIp(2)
G23	Sense-3.3V	G3	Vcc
S45	IBtxOn(1)	S4	IBtxIn(1)
S46	IBtxOp(1)	S3	IBtxIp(1)
G24	Vcc	G2	Signal Ground
S47	IBtxOn(0)	S2	IBtxIn(0)
S48	IBtxOp(0)	S1	IBtxIp(0)
G25	Signal Ground	G1	Sense-12V
Housing	Chassis Ground	Housing	Chassis Ground

7.9.4.7 12X TO 3-4X MICROGIGACN PASSIVE CABLES

The 12x to 3-4x cables are used for connecting to devices which may be configured to operate with a single 12x port, or with three separate 4x ports, using the same pins. The 12x to 3-4x cable provides an interface to a 12x interface board CXP connector, operating as either a single 12x port or as three 4x ports. The opposite end of the cable provides three separate 4x cable connectors.

Notes

1. When using a 12X to 3-4X cable, ports 2 and 3 will not width negotiate since there is no lane 0 on those ports.
2. Port 3 (plug 4) of a 12X to 3-4X active cable plugged into an 8X board receptacle will not function.

Twenty-four pairs of signals are used, twelve each for transmit and receive. Eight pairs are used for each 4X port.

C7-24.2.4: The pin assignment listed in [Table 124](#) shall be used for 12X to 3-4X cables using the microGigaCN interface.

Note

Cable signal nets shall be connected from each *transmit* signal pair on the source port to the appropriate *receive* signal pair on the desired destination port. For example, a cable from port y to port z should be wired to connect IBtxOp(0) (positive output signal from port y, bit 0), to IBtzIp(0) (positive input to port z, bit 0) and IBtxOn(0) to IBtzIn(0), as shown in [Table 124](#).

Table 124 12X to 3-4X cable connector signal assignment

Plug 1		Plug number		
Pin number	Signal		Pin number	Signal
G1-G25	Signal Ground	2	G1-G9	Signal Ground
		3	G1-G9	Signal Ground
		4	G1-G9	Signal Ground
S1	IBtx.1Ip(0)	2	S16	IBtx.1Op(0)
S2	IBtx.1In(0)	2	S15	IBtx.1On(0)
S3	IBtx.1Ip(1)	2	S14	IBtx.1Op(1)
S4	IBtx.1In(1)	2	S13	IBtx.1On(1)
S5	IBtx.1Ip(2)	2	S12	IBtx.1Op(2)
S6	IBtx.1In(2)	2	S11	IBtx.1On(2)
S7	IBtx.1Ip(3)	2	S10	IBtx.1Op(3)
S8	IBtx.1In(3)	2	S9	IBtx.1On(3)
S9	IBtx.2Ip(0)	3	S16	IBtx.2Op(0)
S10	IBtx.2In(0)	3	S15	IBtx.2On(0)
S11	IBtx.2Ip(1)	3	S14	IBtx.2Op(1)
S12	IBtx.2In(1)	3	S13	IBtx.2On(1)
S13	IBtx.2Ip(2)	3	S12	IBtx.2Op(2)
S14	IBtx.2In(2)	3	S11	IBtx.2On(2)
S15	IBtx.2Ip(3)	3	S10	IBtx.2Op(3)
S16	IBtx.2In(3)	3	S9	IBtx.2On(3)
S17	IBtx.3Ip(0)	4	S16	IBtx.3Op(0)

Table 124 12X to 3-4X cable connector signal assignment

Plug 1		Plug number		
Pin number	Signal		Pin number	Signal
S18	IBtx.3In(0)	4	S15	IBtx.3On(0)
S19	IBtx.3Ip(1)	4	S14	IBtx.3Op(1)
S20	IBtx.3In(1)	4	S13	IBtx.3On(1)
S21	IBtx.3Ip(2)	4	S12	IBtx.3Op(2)
S22	IBtx.3In(2)	4	S11	IBtx.3On(2)
S23	IBtx.3Ip(3)	4	S10	IBtx.3Op(3)
S24	IBtx.3In(3)	4	S9	IBtx.3On(3)
S25	IBtx.3On(3)	4	S8	IBtx.3In(3)
S26	IBtx.3Op(3)	4	S7	IBtx.3Ip(3)
S27	IBtx.3On(2)	4	S6	IBtx.3In(2)
S28	IBtx.3Op(2)	4	S5	IBtx.3Ip(2)
S29	IBtx.3On(1)	4	S4	IBtx.3In(1)
S30	IBtx.3Op(1)	4	S3	IBtx.3Ip(1)
S31	IBtx.3On(0)	4	S2	IBtx.3In(0)
S32	IBtx.3Op(0)	4	S1	IBtx.3Ip(0)
S33	IBtx.2On(3)	3	S8	IBtx.2In(3)
S34	IBtx.2Op(3)	3	S7	IBtx.2Ip(3)
S35	IBtx.2On(2)	3	S6	IBtx.2In(2)
S36	IBtx.2Op(2)	3	S5	IBtx.2Ip(2)
S37	IBtx.2On(1)	3	S4	IBtx.2In(1)
S38	IBtx.2Op(1)	3	S3	IBtx.2Ip(1)
S39	IBtx.2On(0)	3	S2	IBtx.2In(0)
S40	IBtx.2Op(0)	3	S1	IBtx.2Ip(0)
S41	IBtx.1On(3)	2	S8	IBtx.1In(3)
S42	IBtx.1Op(3)	2	S7	IBtx.1Ip(3)
S43	IBtx.1On(2)	2	S6	IBtx.1In(2)
S44	IBtx.1Op(2)	2	S5	IBtx.1Ip(2)
S45	IBtx.1On(1)	2	S4	IBtx.1In(1)
S46	IBtx.1Op(1)	2	S3	IBtx.1Ip(1)

Table 124 12X to 3-4X cable connector signal assignment

Plug 1		Plug number		
Pin number	Signal		Pin number	Signal
S47	IBtx.1On(0)	2	S2	IBtx.1In(0)
S48	IBtx.1Op(0)	2	S1	IBtx.1Ip(0)
Housing	Chassis Ground	2, 3, 4	Housing	Chassis Ground

7.9.5 QSFP+ INTERFACE CABLES

These cables use QSFP+ connectors described in [Section 7.5, “4X QSFP+ Interface connectors,” on page 373](#). In some special cases a second connector of a different type is also used as described in the relevant section below.

7.9.5.1 QSFP+ TO QSFP+ CABLES

4X cables utilizing the QSFP and QSFP+ interface shall be wired according to [Table 125](#).

Note that the signal naming is different than that used for a 4X cable with the microGigaCN interface. Tx1n, for instance, is the lowest-order input bit on the cable and which receives its input from the host transmitter. This is equivalent to the microGigaCN interface host output signal IBtxOn(0).

Table 125 QSFP cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
1	Signal Ground	20	Signal Ground
2	Tx2n	21	Rx2n
3	Tx2p	22	Rx2p
4	Signal Ground	23	Signal Ground
5	Tx4n	24	Rx4n
6	Tx4p	25	Rx4p
7	Signal Ground	26	Signal Ground
8	ModSelL (See note 1)	27	ModPrsL (See note 1)
9	LPMode_Reset (See note 1)	28	IntL (See note 1)
10	VccRx (See note 1)	29	VccTx (See note 1)
11	SCL (See note 1)	30	Vcc1 (See note 1)
12	SDA (See note 1)	31	Reserved

Table 125 QSFP cable connector signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
13	Signal Ground	32	Signal Ground
14	Rx3p	33	Tx3p
15	Rx3n	34	Tx3n
16	Signal Ground	35	Signal Ground
17	Rx1p	36	Tx1n
18	Rx1n	37	Tx1p
19	Signal Ground	38	Signal Ground
20	Signal Ground	1	Signal Ground
21	Rx2n	2	Tx2n
22	Rx2p	3	Tx2p
23	Signal Ground	4	Signal Ground
24	Rx4n	5	Tx4n
25	Rx4p	6	Tx4p
26	Signal Ground	7	Signal Ground
27	ModPrsL (See note 1)	8	ModSelL (See note 1)
28	IntL (See note 1)	9	LPMode_Reset (See note 1)
29	VccTx (See note 1)	10	VccRx (See note 1)
30	Vcc1 (See note 1)	11	SCL (See note 1)
31	Reserved	12	SDA (See note 1)
32	Signal Ground	13	Signal Ground
33	Tx3p	14	Rx3p
34	Tx3n	15	Rx3n
35	Signal Ground	16	Signal Ground
36	Tx1p	17	Rx1p
37	Tx1n	18	Rx1n
38	Signal Ground	19	Signal Ground
Housing	Chassis Ground	Housing	Chassis Ground

Note 1: These signals shall not be wired through the cable from Plug 1 to Plug 2.

7.9.5.2 QSFP TO 4X MICROGIGACN PASSIVE CABLES

C7-25: The pin assignment shown in [Table 126 on page 467](#) shall be used for links connecting the 4x QSFP+ interface with passive 4X microGigaCN board connectors.

Table 126 Signal connections between 4x-QSFP+ and 4x microGigaCN connectors

4x micriGigaCN			QSFP	
Pin Number	Signal		QSFP Signal Name	QSFP Pin Number
G1-G9	Signal Ground	<->	Signal Ground	1,4,7,13,16,19,20, 23,26,32,35,38
S1	IBtxIp(0)	<- p	Tx1p	36
S2	IBtxIn(0)	<- n	Tx1n	37
S3	IBtxIp(1)	<- p	Tx2p	3
S4	IBtxIn(1)	<- n	Tx2n	2
S5	IBtxIp(2)	<- p	Tx3p	33
S6	IBtxIn(2)	<- n	Tx3n	34
S7	IBtxIp(3)	<- p	Tx4p	6
S8	IBtxIn(3)	<- n	Tx4n	5
S9	IBtxOn(3)	--> n	Rx4n	24
S10	IBtxOp(3)	--> p	Rx4p	25
S11	IBtxOn(2)	--> n	Rx3n	15
S12	IBtxOp(2)	--> p	Rx3p	14
S13	IBtxOn(1)	--> n	Rx2n	21
S14	IBtxOp(1)	--> p	Rx2p	22
S15	IBtxOn(0)	--> n	Rx1n	18
S16	IBtxOp(0)	--> p	Rx1p	17
Housing	Chassis Ground	a	Housing	Chassis ground (case common)

a. Chassis ground is isolated from Signal Ground (GND) in the cable and connector, on both 4x board connector and QSFP ends, to allow equipment designer flexibility regarding connections between EMI shields and circuit ground.

7.9.5.3 QSFP TO 4X MICROGIGACN ACTIVE CABLES

C7-26: The pin assignment shown in [Table 126 on page 467](#) shall be used for links connecting the 4x QSFP+ interface with previously-specified 4X active board connectors.

Table 127 Signal connections between 4x-QSFP+ and 4x microGigaCN connectors

4x microGigaCN			QSFP	
Pin Number	Signal		QSFP Signal Name	QSFP Contact Number
G1	Sense 12 V	<-->	Signal Ground	1,4,7,13,16,19,20, 23,26,32,35,38
S1	IBtxIp(0)	<-- p	Tx1p	36
G2-G6	Signal Ground		Signal Ground	1,4,7,13,16,19,20, 23,26,32,35,38
S2	IBtxIn(0)	<-- n	Tx1n	37
S3	IBtxIp(1)	<-- p	Tx2p	3
S4	IBtxIn(1)	<-- n	Tx2n	2
S5	IBtxIp(2)	<-- p	Tx3p	33
S6	IBtxIn(2)	<-- n	Tx3n	34
S7	IBtxIp(3)	<-- p	Tx4p	6
S8	IBtxIn(3)	<-- n	Tx4n	5
S9	IBtxOn(3)	--> n	Rx4n	24
S10	IBtxOp(3)	--> p	Rx4p	25
S11	IBtxOn(2)	--> n	Rx3n	15
S12	IBtxOp(2)	--> p	Rx3p	14
S13	IBtxOn(1)	--> n	Rx2n	21
S14	IBtxOp(1)	--> p	Rx2p	22
G7	Sense 3.3 V			
S15	IBtxOn(0)	--> n	Rx1n	18
G8	Vcc			
S16	IBtxOp(0)	--> p	Rx1p	17
G9	Signal Ground		Signal Ground	1,4,7,13,16,19,20, 23,26,32,35,38
Housing	Chassis Ground	a	Housing	Chassis ground (case common)

a. Chassis ground is isolated from Signal Ground (GND) in the cable and connector, on both 4x board connector and QSFP ends, to allow equipment designer flexibility regarding connections between EMI shields and circuit ground.

7.9.5.4 Two 4x QSFP+ CABLES FOR 8X CONFIGURATION

Devices may have the capability to configure two separate 4x ports to operate as an 8X port as described in [Section 7.7.3.4 on page 407](#).

Similarly, it is possible for two separate 4x pluggable devices to be configured to operate as the equivalent of an 8X device.

As shown in [Figure 156](#), two 4X QSFP+ devices operating as a single 8X link shall be oriented such that lanes 0-3 on each end are interconnected by one pair of pluggable devices, and lanes 4-7 on each end are interconnected by a second pair of pluggable devices.

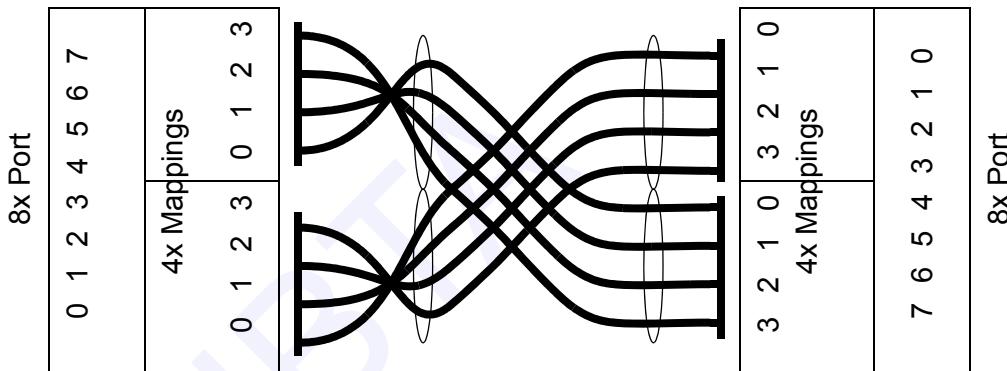


Figure 156 8x port using two 4X Pluggable devices

7.9.5.4.1 SKEW BUDGET

Combinations of 4x QSFP+ cable or module pairs operating as 8X links shall have matched transmission times such that the aggregate bit-to-bit skew requirements, as defined in [Volume 2, Release 1.2.1, Section 6.9 Bit to Bit Skew 4x Optical Pluggable Devices](#), are met. Retiming repeaters may be incorporated into the pluggable devices, but no retiming repeaters shall be inserted between the transmitter or receiver and the pluggable interfaces.

Implementation Note

The full amount of bit-to-bit skew allowed at the receiver is 60 UI (24, 12, or 6 ns at SDR, DDR, and QDR, respectively). Of this, 1.5 ns of skew must be reserved to allow for 500 ps of bit-to-bit skew in the Driving Transmitter (S_{DBtB}) and 500 ps of Adapter board skew (S_{ABtB}) at each end, so the maximum bit-to-bit skew between sets of pluggable devices is limited to 22.5 ns, 10.5 ns, or 4.5 ns at SDR, DDR, and QDR rates, respectively. This budget must be allocated, in a vendor-dependent manner, between skew in the pluggable devices, potentially including embedded retiming repeaters, and skew in signal propagation through the cable itself.

7.9.5.4.2 LINK TRAINING

As described in [Volume 2, Release 1.2.1, Section Link Initialization and Training](#), a link may operate with only a subset of the lanes connected. For example, if some lanes of an 8X port are disconnected, the port may come up as a 4X link with the other lanes disabled. This feature has implications for the case of multiple 4X QSFP+ cables configured to construct an 8X link.

If the ports on both ends of an 8X link are active (i.e., their Link Training State Machines are in the Polling state) while the 4X QSFP+ cables are being plugged in, the ports will generally traverse the full initialization and training sequence after the first 4X QSFP+ cable is connected, coming up as a 4X link. Plugging in the second or third cable may not force retraining, and the link may continue to operate with only the first pair active as a 4X link.

In order to ensure that all lanes of an 8X link constructed using 4X QSFP+ cables or modules, the link must be trained or retrained after all devices are connected and operational. Since aggregation of multiple 4X links into a wider link is unspecified, the methods for accomplishing this retraining are implemented in a vendor-dependent manner.

When multiple 4x cables or modules are used to construct an 8X link, it is important that the devices be plugged with lanes correctly connected, both within each 4X link and across 4X links, as shown in [Figure 156 on page 469](#). There are several incorrect ways of connecting multiple 4x devices that result in the reversal of subsets of lanes. One example of an incorrect configuration for an 8x link is shown in [Figure 157 on page 470](#). The lane reversal function described in [Volume 2, Release 1.2.1, Section 5.6.2.2 Lane Reversal Correction \(Optional\)](#) can not correct for these situations since it operates across the full width of the link.

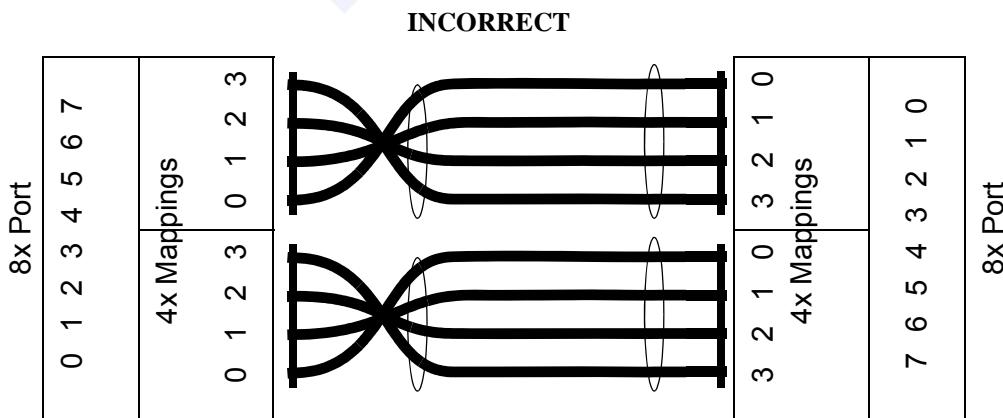


Figure 157 Example of incorrect cabling of an 8X port using two 4X Pluggable devices

7.9.5.5 4x QSFP+ CABLES FOR 12X CONFIGURATION

Devices may have the capability to configure three separate 4x ports to operate as a 12X port as described in [Section 7.7.3.4 on page 407](#).

Similarly, it is possible for three separate 4x QSFP+ cables or modules to be configured to operate as the equivalent of a 12X device.

As shown in [Figure 158 on page 471](#) three 4X QSFP+ cables or modules operating as a single 12X link shall be oriented such that lanes 0-3 on each end are interconnected by one QSFP+ cable lanes 4-7 on each end are interconnected by a second QSFP+ cable, and lanes 8-11 on each end are interconnected by a third QSFP+ cable.

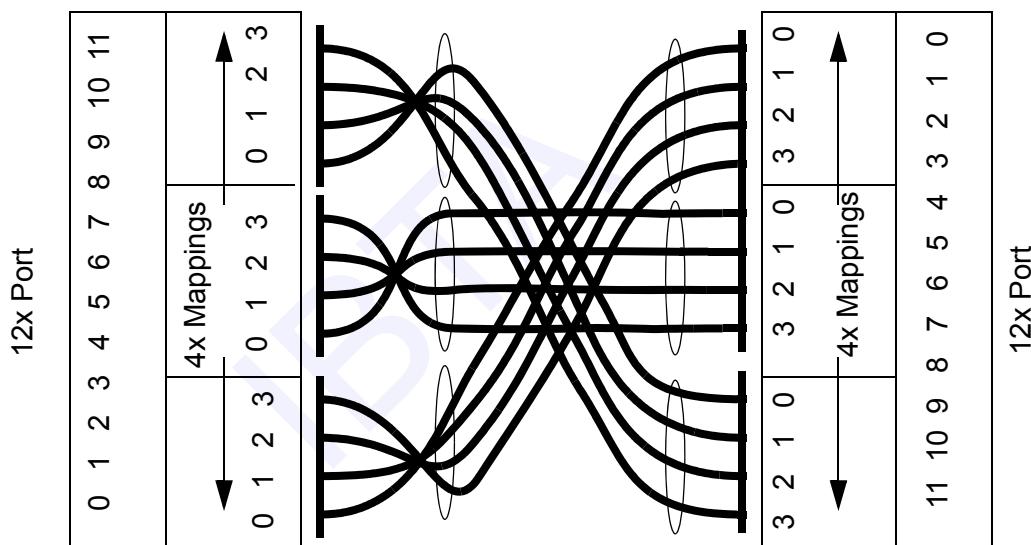


Figure 158 12x port using three 4X Pluggable devices

7.9.5.5.1 SKew BUDGET

Combinations of 4x QSFP+ cables operating as 12X links shall have matched transmission times such that the aggregate bit-to-bit skew requirements, as defined in [Volume 2, Release 1.2.1, Section 6.9 Bit to Bit Skew 4x Optical Pluggable Devices](#), are met. Retiming repeaters may be incorporated into the pluggable devices, but no retiming repeaters shall be inserted between the transmitter or receiver and the pluggable interfaces.

Implementation Note

The full amount of bit-to-bit skew allowed at the receiver is 60 UI (24, 12, or 6 ns at SDR, DDR, and QDR, respectively). Of this, 1.5 ns of skew must be reserved to allow for 500 ps of bit-to-bit skew in the Driving Transmitter (S_{DBtB}) and 500 ps of Adapter board skew (S_{ABtB}) at each end, so the maximum bit-to-bit skew between sets of pluggable devices is limited to 22.5 ns, 10.5 ns, or 4.5 ns at SDR, DDR, and QDR rates, respectively. This budget must be allocated, in a vendor-dependent manner, between skew in the pluggable devices, potentially including embedded retiming repeaters, and skew in signal propagation through the cable itself.

7.9.5.5.2 LINK TRAINING

As described in [Volume 2, Release 1.2.1, Section Link Initialization and Training](#), a link may operate with only a subset of the lanes connected. For example, if some lanes of an 12X port are disconnected, the port may come up as a 4X link with the other lanes disabled. This feature has implications for the case of multiple 4X Pluggable devices configured to construct an 12X link.

If the ports on both ends of an 12X link are active (i.e., their Link Training State Machines are in the Polling state) while the 4X pluggable devices are being plugged in, the ports will generally traverse the full initialization and training sequence after the first 4X pluggable device pair is connected, coming up as a 4X link. Plugging in the second or third device pairs may not force retraining, and the link may continue to operate with only the first pair active as a 4X link.

In order to ensure that all lanes of an 12X link constructed using 4X QSFP+ links are used, the link must be trained or retrained after all devices are connected and operational. Since aggregation of multiple 4X links into a wider link is unspecified, the methods for accomplishing this retraining are implemented in a vendor-dependent manner.

When multiple 4x devices are used to construct a 12X link, it is important that the devices be plugged with lanes correctly connected, both within each 4X link and across 4X links, as shown in [Figure 158 on page 471](#). There are several incorrect ways of connecting multiple 4x devices that result in the reversal of subsets of lanes. One example of an incorrect configuration for an 8x link is shown in [Figure 157 on page 470](#). Particular care must be taken with 12X links constructed from three 4X device pairs because of the potential for five improper configurations of pairs. The lane reversal function described in [Volume 2, Release 1.2.1, Section 5.6.2.2 Lane Reversal Correction \(Optional\)](#) can not correct for these situations since it operates across the full width of the link.

7.9.6 CXP INTERFACE CABLES

A variety of modules, devices, and cables are envisioned for this interface. This list is not normative or restrictive, but will clarify potential design options. Some examples of these would include passive or active optical cables as shown in [Figure 159](#), or passive or active copper cables as shown in [Figure 160 on page 473](#).

For a description of the CXP module see [Section 7.8 on page 413](#).

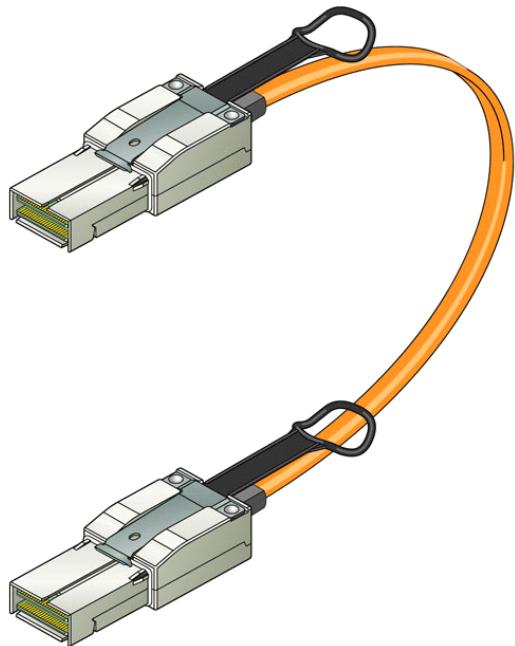


Figure 159 Active Optical Cable

[Figure 160](#) shows a copper cable, which may be passively or actively equalized.

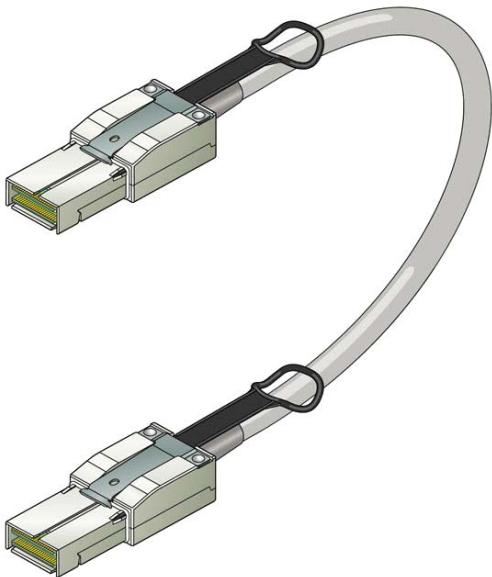


Figure 160 Passive or Active Copper Cable

The contact assignment for CXP cables is listed in [Table 128 on page 475](#).

Note that the signal naming is different than that used for a 12X cable with the microGigaCN interface. Tx0n, for instance, is the lowest-order input bit on the cable and which receives its input from the host transmitter. This is equivalent to the microGigaCN interface host output signal IBtxOn(0).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

7.9.6.1 12x CXP to CXP CABLES

The contact assignment listed in [Table 128](#) shall be used for 12X CXP to CXP cables.

Table 128 12X CXP to CXP cable signal assignment

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
A2	Tx1p	C2	Rx1p
A3	Tx1n	C3	Rx1n
A5	Tx3p	C5	Rx3p
A6	Tx3n	C6	Rx3n
A8	Tx5p	C8	Rx5p
A9	Tx5n	C9	Rx5n
A11	Tx7p	C11	Rx7p
A12	Tx7n	C12	Rx7n
A14	Tx9p	C14	Rx9p
A15	Tx9n	C15	Rx9n
A17	Tx11p	C17	Rx11p
A18	Tx11n	C18	Rx11n
B2	Tx0p	D2	Rx0p
B3	Tx0n	D3	Rx0n
B5	Tx2p	D5	Rx2p
B6	Tx2n	D6	Rx2n
B8	Tx4p	D8	Rx4p
B9	Tx4n	D9	Rx4n
B11	Tx6p	D11	Rx6p
B12	Tx6n	D12	Rx6n
B14	Tx8p	D14	Rx8p
B15	Tx8n	D15	Rx8n
B17	Tx10p	D17	Rx10p
B18	Tx10n	D18	Rx10n

Plug 1		Plug 2	
Pin number	Signal	Pin number	Signal
C2	Rx1p	A2	Tx1p
C3	Rx1n	A3	Tx1n
C5	Rx3p	A5	Tx3p
C6	Rx3n	A6	Tx3n
C8	Rx5p	A8	Tx5p
C9	Rx5n	A9	Tx5n
C11	Rx7p	A11	Tx7p
C12	Rx7n	A12	Tx7n
C14	Rx9p	A14	Tx9p
C15	Rx9n	A15	Tx9n
C17	Rx11p	A17	Tx11p
C18	Rx11n	A18	Tx11n
D2	Rx0p	B2	Tx0p
D3	Rx0n	B3	Tx0n
D5	Rx2p	B5	Tx2p
D6	Rx2n	B6	Tx2n
D8	Rx4p	B8	Tx4p
D9	Rx4n	B9	Tx4n
D11	Rx6p	B11	Tx6p
D12	Rx6n	B12	Tx6n
D14	Rx8p	B14	Tx8p
D15	Rx8n	B15	Tx8n
D17	Rx10p	B17	Tx10p
D18	Rx10n	B18	Tx10n

A1, A4, A7, A10, A13, A16, A19, B1, B4, B7, B10, B13, B16, B19, C1, C4, C7, C10, C13, C16, C19, D1, D4, D7, D10, D13, D16, D19 on each plug are connected to local Signal Ground. They are not connected through cable.
 A20, A21, C20, and C21 are connected to local management interface. They are not connected through cable.
 B20, B21, D20, and D21 provide local power. They are not connected through cable.

7.9.6.2 12x CXP TO 3-4X CABLES

[Figure 161](#) shows exemplary configurations for 12X to 3-4X copper and active optical cables.

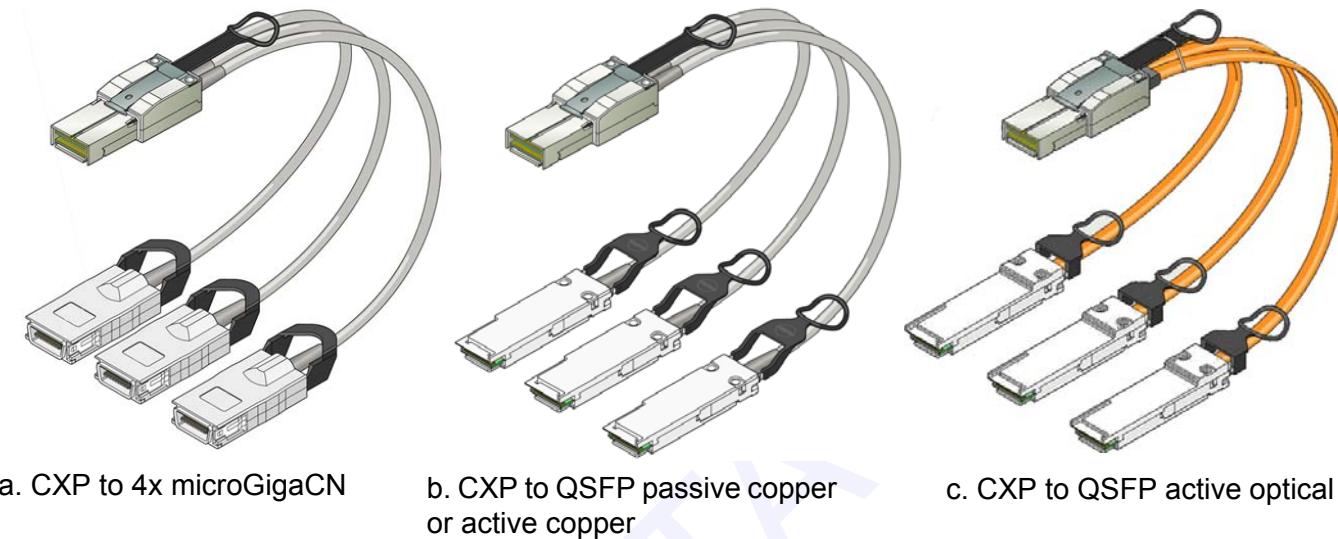


Figure 161 12x to 3-4x cables

The contact assignment listed in [Table 129](#) shall be used for 12X CXP to 3-4X microGigaCN passive cables.

Table 129 12X CXP to 3-4X microGigaCN passive cable signal assignment

CXP Plug 1		microGiga CN Plug number	microGiga CN pin number	Signal	Plug 1		microGiga CN Plug number	microGiga CN pin number	Signal
CXP pin Number	Signal	CXP pin Number	Signal	CXP pin Number	Signal	CXP pin Number	Signal	CXP pin Number	Signal
A2	Tx1.1p	2	S3	IBtx.1Ip(1)	C2	Rx1.1p	2	S14	IBtx.1Op(1)
A3	Tx1.1n	2	S4	IBtx.1In(1)	C3	Rx1.1n	2	S13	IBtx.1On(1)
A5	Tx1.3p	2	S7	IBtx.1Ip(3)	C5	Rx1.3p	2	S10	IBtx.1Op(3)
A6	Tx1.3n	2	S8	IBtx.1In(3)	C6	Rx1.3n	2	S9	IBtx.1On(3)
A8	Tx2.1p	3	S3	IBtx.2Ip(1)	C8	Rx2.1p	3	S14	IBtx.2Op(1)
A9	Tx2.1n	3	S4	IBtx.2In(1)	C9	Rx2.1n	3	S13	IBtx.2On(1)
A11	Tx2.3p	3	S7	IBtx.2Ip(3)	C11	Rx2.3p	3	S10	IBtx.2Op(3)
A12	Tx2.3n	3	S8	IBtx.2In(3)	C12	Rx2.3n	3	S9	IBtx.2On(3)
A14	Tx3.1p	4	S3	IBtx.3Ip(1)	C14	Rx3.1p	4	S14	IBtx.3Op(1)
A15	Tx3.1n	4	S4	IBtx.3In(1)	C15	Rx3.1n	4	S13	IBtx.3On(1)
A17	Tx3.3p	4	S7	IBtx.3Ip(3)	C17	Rx3.3p	4	S10	IBtx.3Op(3)
A18	Tx3.3n	4	S8	IBtx.3In(3)	C18	Rx3.3n	4	S9	IBtx.3On(3)
B2	Tx1.0p	2	S1	IBtx.1Ip(0)	D2	Rx1.0p	2	S16	IBtx.1Op(0)
B3	Tx1.0n	2	S2	IBtx.1In(0)	D3	Rx1.0n	2	S15	IBtx.1On(0)
B5	Tx1.2p	2	S5	IBtx.1Ip(2)	D5	Rx1.2p	2	S12	IBtx.1Op(2)
B6	Tx1.2n	2	S6	IBtx.1In(2)	D6	Rx1.2n	2	S11	IBtx.1On(2)
B8	Tx2.0p	3	S1	IBtx.2Ip(0)	D8	Rx2.0p	3	S16	IBtx.2Op(0)
B9	Tx2.0n	3	S2	IBtx.2In(0)	D9	Rx2.0n	3	S15	IBtx.2On(0)
B11	Tx2.2p	3	S5	IBtx.2Ip(2)	D11	Rx2.2p	3	S12	IBtx.2Op(2)
B12	Tx2.2n	3	S6	IBtx.2Inp2	D12	Rx2.2n	3	S11	IBtx.2On(2)
B14	Tx3.0p	4	S1	IBtx.3Ip(0)	D14	Rx3.0p	4	S16	IBtx.3Op(0)
B15	Tx3.0n	4	S2	IBtx.3In(0)	D15	Rx3.0n	4	S15	IBtx.3On(0)
B17	Tx3.2p	4	S5	IBtx.3Ip(2)	D17	Rx3.2p	4	S12	IBtx.3Op(2)
B18	Tx3.2n	4	S6	IBtx.3In(2)	D18	Rx3.2n	4	S11	IBtx.3On(2)

On the CXP connector, A1, A4, A7, A10, A13, A16, A19, B1, B4, B7, B10, B13, B16, B19, C1, C4, C7, C10, C13, C16, C19, D1, D4, D7, D10, D13, D16, D19 are connected to local Signal Ground. A20, A21, C20, and C21 are connected to the local management interface. B20, B21, D20, and D21 provide local power. None of these pins is connected through the cable.
 On the microGigaCN connectors, pins G1-G9 are connected to local Signal Ground.

The pin assignment listed in [Table 130](#) shall be used for 12X CXP to 3-4X microGigaCN active cables.

Table 130 12X CXP to 3-4X microGigaCN active cable signal assignment

CXP Plug 1		microGiga CN Plug number	microGiga CN pin number	Signal
CXP pin Number	Signal			
A2	Tx1.1p	2	S3	IBtx.1Ip(1)
A3	Tx1.1n	2	S4	IBtx.1In(1)
A5	Tx1.3p	2	S7	IBtx.1Ip(3)
A6	Tx1.3n	2	S8	IBtx.1In(3)
A8	Tx2.1p	3	S3	IBtx.2Ip(1)
A9	Tx2.1n	3	S4	IBtx.2In(1)
A11	Tx2.3p	3	S7	IBtx.2Ip(3)
A12	Tx2.3n	3	S8	IBtx.2In(3)
A14	Tx3.1p	4	S3	IBtx.3Ip(1)
A15	Tx3.1n	4	S4	IBtx.3In(1)
A17	Tx3.3p	4	S7	IBtx.3Ip(3)
A18	Tx3.3n	4	S8	IBtx.3In(3)
B2	Tx1.0p	2	S1	IBtx.1Ip(0)
B3	Tx1.0n	2	S2	IBtx.1In(0)
B5	Tx1.2p	2	S5	IBtx.1Ip(2)
B6	Tx1.2n	2	S6	IBtx.1In(2)
B8	Tx2.0p	3	S1	IBtx.2Ip(0)
B9	Tx2.0n	3	S2	IBtx.2In(0)
B11	Tx2.2p	3	S5	IBtx.2Ip(2)
B12	Tx2.2n	3	S6	IBtx.2In(2)
B14	Tx3.0p	4	S1	IBtx.3Ip(0)
B15	Tx3.0n	4	S2	IBtx.3In(0)
B17	Tx3.2p	4	S5	IBtx.3Ip(2)
B18	Tx3.2n	4	S6	IBtx.3In(2)

On the CXP connector, A1, A4, A7, A10, A13, A16, A19, B1, B4, B7, B10, B13, B16, B19, C1, C4, C7, C10, C13, C16, C19, D1, D4, D7, D10, D13, D16, D19 are connected to local Signal Ground. A20, A21, C20, and C21 are connected to the local management interface. B20, B21, D20, and D21 provide local power. None of these pins is connected through the cable.

On the microGigaCN connectors, pins G2, G3, G4-6, and G9 are connected to local Signal Ground.

Plug 1		microGiga CN Plug number	microGiga CN pin number	Signal
CXP pin Number	Signal			
C2	Rx1.1p	2	S14	IBtx.1Op(1)
C3	Rx1.1n	2	S13	IBtx.1On(1)
C5	Rx1.3p	2	S10	IBtx.1Op(3)
C6	Rx1.3n	2	S9	IBtx.1On(3)
C8	Rx2.1p	3	S14	IBtx.2Op(1)
C9	Rx2.1n	3	S13	IBtx.2On(1)
C11	Rx2.3p	3	S10	IBtx.2Op(3)
C12	Rx2.3n	3	S9	IBtx.2On(3)
C14	Rx3.1p	4	S14	IBtx.3Op(1)
C15	Rx3.1n	4	S13	IBtx.3On(1)
C17	Rx3.3p	4	S10	IBtx.3Op(3)
C18	Rx3.3n	4	S9	IBtx.3On(3)
D2	Rx1.0p	2	S16	IBtx.1Op(0)
D3	Rx1.0n	2	S15	IBtx.1On(0)
D5	Rx1.2p	2	S12	IBtx.1Op(2)
D6	Rx1.2n	2	S11	IBtx.1On(2)
D8	Rx2.0p	3	S16	IBtx.2Op(0)
D9	Rx2.0n	3	S15	IBtx.2On(0)
D11	Rx2.2p	3	S12	IBtx.2Op(2)
D12	Rx2.2n	3	S11	IBtx.2On(2)
D14	Rx3.0p	4	S16	IBtx.3Op(0)
D15	Rx3.0n	4	S15	IBtx.3On(0)
D17	Rx3.2p	4	S12	IBtx.3Op(2)
D18	Rx3.2n	4	S11	IBtx.3On(2)

The pin assignment listed in [Table 131](#) shall be used for 12X CXP to QSFP cables.

Table 131 12X CXP to 3-4X QSFP cable signal assignment

Plug 1		QSFP plug number	QSFP pin number	Signal	Plug 1		QSFP plug number	QSFP pin number	Signal
CXP pin Number	Signal				CXP pin Number	Signal			
A2	Tx1.1p	2	22	Rx1.2p	C2	Rx1.1p	2	3	Tx1.2p
A3	Tx1.1n	2	21	Rx1.2n	C3	Rx1.1n	2	2	Tx1.2n
A5	Tx1.3p	2	25	Rx1.4p	C5	Rx1.3p	2	6	Tx1.4p
A6	Tx1.3n	2	24	Rx1.4n	C6	Rx1.3n	2	5	Tx1.4n
A8	Tx2.1p	3	22	Rx2.2p	C8	Rx2.1p	3	3	Tx2.2p
A9	Tx2.1n	3	21	Rx2.2n	C9	Rx2.1n	3	2	Tx2.2n
A11	Tx2.3p	3	25	Rx2.4p	C11	Rx2.3p	3	6	Tx2.4p
A12	Tx2.3n	3	24	Rx2.4n	C12	Rx2.3n	3	5	Tx2.4n
A14	Tx3.1p	4	22	Rx3.2p	C14	Rx3.1p	4	3	Tx3.2p
A15	Tx3.1n	4	21	Rx3.2n	C15	Rx3.1n	4	2	Tx3.2n
A17	Tx3.3p	4	25	Rx3.4p	C17	Rx3.3p	4	6	Tx3.4p
A18	Tx3.3n	4	24	Rx3.4n	C18	Rx3.3n	4	5	Tx3.4n
B2	Tx1.0p	2	17	Rx1.1p	D2	Rx1.0p	2	37	Tx1.1p
B3	Tx1.0n	2	18	Rx1.1n	D3	Rx1.0n	2	36	Tx1.1n
B5	Tx1.2p	2	14	Rx1.3p	D5	Rx1.2p	2	33	Tx1.3p
B6	Tx1.2n	2	15	Rx1.3n	D6	Rx1.2n	2	34	Tx1.3n
B8	Tx2.0p	3	17	Rx2.1p	D8	Rx2.0p	3	37	Tx2.1p
B9	Tx2.0n	3	18	Rx2.1n	D9	Rx2.0n	3	36	Tx2.1n
B11	Tx2.2p	3	14	Rx2.3p	D11	Rx2.2p	3	33	Tx2.3p
B12	Tx2.2n	3	15	Rx2.3n	D12	Rx2.2n	3	34	Tx2.3n
B14	Tx3.0p	4	17	Rx3.1p	D14	Rx3.0p	4	37	Tx3.1p
B15	Tx3.0n	4	18	Rx3.1n	D15	Rx3.0n	4	36	Tx3.1n
B17	Tx3.2p	4	14	Rx3.3p	D17	Rx3.2p	4	33	Tx3.3p
B18	Tx3.2n	4	15	Rx3.3n	D18	Rx3.2n	4	34	Tx3.3n
On the CXP connector, A1, A4, A7, A10, A13, A16, A19, B1, B4, B7, B10, B13, B16, B19, C1, C4, C7, C10, C13, C16, C19, D1, D4, D7, D10, D13, D16, D19 are connected to local Signal Ground. A20, A21, C20, and C21 are connected to the local management interface. B20, B21, D20, and D21 provide local power. None of these pins is connected through the cable. On the QSFP connector, pins 1, 4, 7, 13, 16, 20, 23, 26, 32, and 35 are connected to local Signal Ground. Pins 8, 9, 11, 12, 27, and 28 are connected to the local management interface. Pins 10, 29, and 30 provide local power. Pin 31 is reserved.									

CHAPTER 8: MANAGEMENT INTERFACE

8.1 INTRODUCTION

A management interface, as already commonly used in other form factors like GBIC, SFP, XFP and QSFP is specified for accessing the EEPROM present in the QSFP and CXP modules in order to enable flexible use of the transceiver by the host system. The specification has been modeled on the definition of the SFF QSFP (Quad Small Form-factor Pluggable) multi-lane receiver, with extensions as needed to support InfiniBand signaling at QDR, FDR, and EDR speeds, and to support both (4+4)-lane and (12+12)-lane operation.

The ‘Extended Module Code’ bits inside the QSFP and CXP EEPROMs are optionally used by link endpoints to determine speeds to attempt during link bring-up. The subnet management entity is not expected nor desired to have any involvement in determining these speeds through fabric-level queries of the cable EEPROM. Instead, the information inside the cable EEPROM should in general be viewed as informative at the fabric level and used for link bring-up decisions locally by link endpoints only.

Some timing requirements are critical, especially for a multi-lane device, so the interface speed has been increased relative to single-lane devices such as GBIC, SFP, and XFP.

8.2 VOLTAGE AND TIMING SPECIFICATION

8.2.1 MANAGEMENT INTERFACE VOLTAGE SPECIFICATION

Management signaling logic levels are based on Low Voltage CMOS operating at 3.3V Vcc. Host shall use a pull-up to Vcc3.3 for the two wire interface SCL (clock), SDA (address & data), and Int_L/Reset_L signals.

The electrical specifications are given in [Table 132](#). This specification ensures compatibility between host bus masters and the two wire interface

Table 132 Low Speed Control and Sense Signal Specifications

Parameter	Symbol	Min	Max	Unit	Condition
Module Input Voltage Low	Vil	-0.3	0.4	V	Pull-up to 3.3 V
Module Input Voltage High	Vih	2.3	3.6	V	Min Vih = 0.7*3.3 V
Module Output Voltage Low	Vol	-0.3	0.3	V	Condition IOL=3.0 mA. Pull-up to 3.3 V
Module Output Voltage High	Voh	2.8	3.6	V	Min Voh = 3.3 V - 0.5 V
Module Output Current High	Ioh	-10	10	µA	-0.3V < Voutput < 3.6 V
Capacitance of module on SCL & SDA contacts	C _{i,SCLSDA}		36	pF	Allocate 28 pF for IC(s), 8 pF for module PCB(s)
Capacitance of module on Int_L/Reset_L I/O contact(s)	C _{i,INT_L}		36	pF	Allocate 28 pF for IC(s), 8 pF for module PCB(s)
Total bus capacitive load, SCL, SDA, and Int_L/Reset_L I/O pin	C _b		100	pF	3.0 kOhm Pullup resistor, max.
			200	pF	1.6 kOhm Pullup resistor, max.

8.2.2 MANAGEMENT INTERFACE TIMING SPECIFICATION

In order to support a multi-lane device a 400 kHz clock rate for the serial interface is expected. The timing requirements are shown in [Figure 162 on page 482](#) and specified in [Table 133 on page 482](#). All values are referred to VIH(min) and VIL(max) levels shown in [Table 132](#).

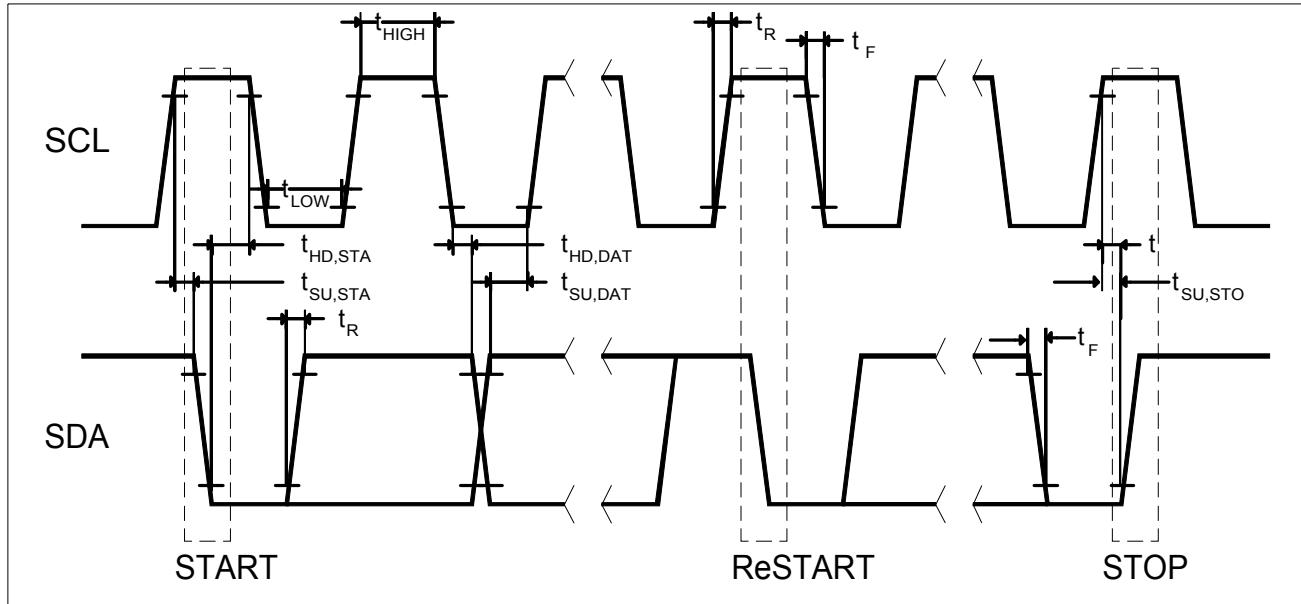


Figure 162 Management interface two wire serial interface timing diagram

Table 133 Management interface two wire serial interface timing specifications

Parameter	Symbol	Min	Max	Unit	Condition
Clock Frequency	f_{SCL}	0	400	kHz	
Clock Pulse Width Low	t_{LOW}	1.3		μs	
Clock Pulse Width High	t_{HIGH}	0.6		μs	
Time bus free before new transmission can start	t_{BUF}	20		μs	Note ^a
START Set-up Time	$t_{SU,STA}$	0.6		μs	
START Hold Time	$t_{Hd,STA}$	0.6		μs	
Data Set-up Time	$t_{SU,DAT}$	0.1		μs	Note ^b
Data Hold Time	$t_{HD,DAT}$	0		μs	Note ^c
SDA and SCL rise time	$t_{R,400}$		0.3	μs	Note ^d
SDA and SCL fall time	$t_{F,400}$		0.3	μs	Note ^e
STOP Set-up Time	$t_{SU,STO}$	0.6		μs	
ModSell Set-up Time (QSFP)	$t_{host_select_setup}$	2		ms	Note ^f
ModSell Hold Time (QSFP)	$t_{host_select_hold}$	10		μs	Note ^g
ModSell Aborted sequence - bus release (QSFP)	$t_{deselect_abort}$	2		ms	Note ^h

a. Between STOP & START and between ACK & ReSTART.

b. Data In Set Up Time is measured from $V_{il(max)}$ SDA or $V_{ih(min)}$ SDA to $V_{il(max)}$ SCL.

c. Data In Hold Time is measured from $V_{il(max)}$ SCL to $V_{il(max)}$ SDA or $V_{ih(min)}$ SDA.

d. Rise Time is measured from $V_{oh(min)}$ SDA to $V_{oh(min)}$ SDA.

e. Fall Time is measured from $V_{oh(min)}$ SDA to $V_{oh(min)}$ SDA.

- f. Setup time on the select lines before start of a host-initiated serial bus sequence.
 g. Delay from completion of a serial bus sequence to changes of the module select status.
 h. Delay from a host de-asserting ModSelL (at any point in a bus sequence) to the QSFP+ module releasing SCL and SDA.

8.3 MEMORY INTERACTION SPECIFICATIONS

Non-volatile memory may be accessed in either single-byte or multiple-byte memory blocks. The largest multiple-byte contiguous write operation that a module shall handle is 4 bytes. The minimum size write block is 1 byte.

8.3.1 TIMING FOR MEMORY TRANSACTIONS

The management interface memory transaction timings are given in [Table 134](#).

Table 134 Management interface memory transaction timing specification

Parameter	Symbol	Min	Max	Unit	Condition
Serial Interface Clock Holdoff - "Clock Stretching"	T_{clock_hold}		500	μs	Note ^a
Complete Single or Sequential Write	t_{WR}		40	ms	Note ^b
Endurance (Write cycles)		50,000		cycles	^c

- a. Maximum time the QSFP or CXP module may hold the SCL line low before continuing with a read or write operation.
 b. Complete up to 4 Byte write. Timing should start from Stop bit at the end of the sequential write operation and continue until the module responds to another operation.
 c. 50K write cycles at 70 C.

8.3.2 TIMING FOR CONTROL AND STATUS FUNCTIONS

Timing for QSFP+ and CXP control and status functions are described in [Table 135](#).

Table 135 I/O Timing for Control and Status Functions

Parameter	Symbol	Min	Max	Unit	Condition, and Notes
Initialization Time	t_{init}		2000	ms	Note ^{a, b, c}
Reset Pulse Width - Min.	$t_{reset_L,PW-min}$	25		ms	Note ^d
Monitor Data Ready Time	t_{data}		2000	ms	Note ^e
Reset Assert Time	$t_{RSTL,OFF}$		2000	ms	Note ^f
Int_L Assert Time	$t_{Int_L,ON}$		200	ms	Note ^g
Interrupt Pulse Width - Min (CXP)	$t_{Int_L,PW-min}$	5		μs	Note ^h
Interrupt Pulse Width - Max (CXP)	$t_{Int_L,PW-max}$		50	μs	Note ⁱ
Int_L Deassert Time	$t_{Int_L,OFF}$		100	ms	Note ^j
Rx LOS Assert Time	$t_{LOS,ON}$		100	ms	Note ^k
Tx Fault Assert Time	$t_{Txfault,ON}$		200	ms	Note ^l
Flag Assert Time	$t_{flag,ON}$		200	ms	Note ^m
Mask Assert Time	$t_{mask,OFF}$		100	ms	Note ⁿ
Mask Deassert Time	$t_{mask,ON}$		100	ms	Note ^o
Select Change Time	$t_{ratesel}$		100	ms	Note ^p

Table 135 I/O Timing for Control and Status Functions (Continued)

Parameter	Symbol	Min	Max	Unit	Condition, and Notes
Power_over-ride or Power-set Assert Time (QSFP)	$t_{P\text{down},ON}$		100	ms	Note ^q
High Power de-Assert Time (CXP)			100		
Power_over-ride or Power-set Deassert Time (QSFP)	$t_{P\text{down},OFF}$		300	ms	Note ^r
High Power Assert Time (CXP)					
Page Select Wait Time, Upper Page 00 or 01	$t_{page_00/01_select}$		100	ms	Note ^s
Page Select Wait Time, Upper Page 02	$t_{page_02_select}$		600	ms	Note ^t

- a. Time from power on, hot plug or rising edge of Reset until the module is fully functional. This time does not apply to non-Power Level 0 modules in the Low Power State.
- b. QSFP: Power on is defined as the instant when supply voltages reach and remain at or above the minimum level specified in [Table 100](#).
- CXP: Power on is defined as the instant when supply voltages reach and remain at or above the minimum level specified in
- c. Fully functional is defined as Int_L asserted due to Data Not Ready (Byte 2, bit 0) deasserted. The module should also meet optical and electrical specifications.
- d. This is the minimum Reset_L pulse width required to reset a module. Assertion of Reset_L activates a complete module reset, i.e., module returns to factory default control settings. While Reset_L is Low, Tx and Rx outputs are disabled and the module does not response to the two-wire serial interface.
- e. Time from power on to Data Not Ready (Byte 2, bit 0) deasserted and Int_L asserted.
- f. Time from rising edge on the Reset_L contact until the module is fully functional. During the Reset Time module will not respond to a “low” on the Int_L/Reset_L signal.
- g. Time from occurrence of condition triggering Int_L until Vout:Int_L = Vol.
- h. CXP: Int_L operates in pulse mode. Static mode (Int_L stays low until reset by host) is not supported for Int_L.
- i. CXP: Int_L pulse width must not exceed $t_{Int_L,PW-\text{max}}$, to distinguish Int_L from a Reset for other devices on bus.
- j. Time from clear on read operation of associated flag until Int_L Status (Lower page, byte 2, bit 1) is cleared. This includes deassert times for Rx LOS, Tx Fault and other flag bits. Measured from falling clock edge after stop bit of read transaction.
- k. Time from Rx LOS state to Rx LOS bit set (value = 1b) and Int_L asserted.
- l. Time from Tx Fault state to Tx Fault bit set (value = 1b) and Int_L asserted.
- m. Time from occurrence of condition triggering flag to associated flag bit set (value = 1b) and Int_L asserted.
- n. Time from mask bit set (value = 1b) until associated Int_L assertion is inhibited.
- o. Time from mask bit cleared (value = 0b) until associated Int_L operation resumes.
- p. Time from change of state of Application or Rate Select bit until transmitter or receiver bandwidth is in conformance with appropriate specification.
- q. QSFP: Time from P_Down bit set (value = 1b) until module power consumption enters Power Class 0.
- CXP: Time from High-Power Mode bit cleared (value = 0b) until module uses less than 6.0 Watts of power.
- r. QSFP: Time from P_Down bit cleared (value = 0b) until the module is fully functional.
- CXP: Time from High-Power Mode bit set (value = 1b) until module is fully functional with power usage indicated in Byte 148 of Upper Page 00h.
- s. Time from setting the Upper Page Select Byte (Lower Page Byte 127) to 01h from 00h, or to 00h from 01h, until the associated upper page is accessible.
- t. Time from setting the Upper Page Select Byte (Lower Page Byte 127) to 02h or from 02 to either 00h or 01h, until the associated upper page is accessible. This longer period for Upper Page 02, vs. other Upper Pages, allows more complex memory management for this infrequently-accessed upper page.

8.3.3 TIMING FOR SQUELCH AND DISABLE FUNCTIONS

Squelch and disable times are described in [Table 136](#).

Table 136 I/O Timing for Squelch and Disable

Parameter	Symbol	Min	Max	Unit	Condition and Notes
Rx Squelch Assert Time	$t_{RxSq,ON}$		0.080	ms	Note ^a
Rx Squelch Deassert Time	$t_{RxSq,OFF}$		0.080	ms	Note ^b
Tx Squelch Assert Time	$t_{TxSq,ON}$		10	ms	Note ^c
Tx Squelch Deassert Time	$t_{TxSq,OFF}$		0.500	ms	Note ^d
Tx Channel Disable & Output Disable Assert Time	$t_{Txdis,ON}$		100	ms	Note ^e
Tx Channel Disable & Output Disable Deassert Time	$t_{Txdis,OFF}$		400	ms	Note ^f
Rx Output Disable Assert Time	$t_{Rxdis,ON}$		100	ms	Note ^g
Rx Output Disable Deassert Time	$t_{Rxdis,OFF}$		100	ms	Note ^h
Squelch Disable Assert Time	$t_{Sqdis,ON}$		100	ms	Note ⁱ
Squelch Disable Deassert Time	$t_{Sqdis,OFF}$		100	ms	Note ^j

a. Time from loss of Rx input signal until the squelched output condition is reached. See [Section 7.5.3.4](#).

b. Time from resumption of Rx input signals until normal Rx output condition is reached. See [Section 7.5.3.4](#).

c. Time from loss of Tx input signal until the squelched output condition is reached. See [Section 7.5.3.4](#).

d. Time from resumption of Tx input signals until normal Tx output condition is reached. See [Section 7.5.3.4](#). Note that this specification is much tighter than typical SFF specifications, in order to assure correct InfiniBand beaconing operation, as described in [Section 6.7.2](#).

e. Time from Tx Disable or Tx Output Disable bit set (value = 1b) until optical output falls below 10% of nominal.

f. Time from Tx Disable or Tx Output Disable bit cleared (value = 0b) until optical output rises above 90% of nominal.

Measured from Stop bit low-to-high SDA transition.

g. Time from Rx Output Disable bit set (value = 1b) until Rx output falls below 10% of nominal.

h. Time from Rx Output Disable bit cleared (value = 0b) until Rx output rises above 90% of nominal.

i. This applies to Rx and Tx Squelch and is the time from bit set (value = 1b) until squelch functionality is disabled.

j. This applies to Rx and Tx Squelch and is the time from bit cleared (value = 0b) until squelch functionality is enabled.

8.4 DEVICE ADDRESSING AND OPERATION

Serial Clock (SCL): The host supplied SCL input to CXP transceivers is used to positive-edge clock data into each QSFP or CXP module and negative-edge clock data out of each module. The SCL line may be pulled low by the module during clock stretching.

Serial Data (SDA): The SDA signal is bidirectional for serial data transfer. This signal is open-drain or open-collector driven and may be wire-ORed with multiple open-drain or open collector devices, limited by aggregate capacitance vs. clock speed.

Master/Slave: QSFP and CXP transceivers operate only as slave devices. The host must provide a bus master for SCL and initiate all read/write communication.

Device Address: All QSFP and CXP modules use the same base addresses, 1010 000x and 1010 100x, where x indicates read (1) or write(0). Each module supports an internal memory map, with one or more 128B lower page and one or more 128B upper pages, depending on module capabilities. See [Section 8.7 on page 520](#) for memory map struc-

ture within each module.

Single QSFP or CXP device per SCL/SDA: Since all transceivers or modules use the same base addresses, each QSFP port requires its own ModSel signal, and each CXP port requires its own SCL/SDA bus. Support of multiple ports in a host requires multiple ModSel signals, or SCL/SDA buses, or multiplexing circuitry such as a multiplexer chip or a switch chip. See [Section 8.2.1. "Management interface Voltage Specification." on page 480](#) and [Table 132](#) for more information.

Clock and Data Transitions: The SDA signal is normally pulled high in the host. Data on the SDA signal may change only during SCL low time periods. Data changes during SCL high periods indicate a START or STOP condition. All addresses and data words are serially transmitted to and from the module in 8-bit words. Every byte on the SDA line must be 8-bits long. Data is transferred with the most significant bit (MSB) first.

START Condition: A high-to-low transition of SDA with SCL high is a START condition, which must precede any other command.

STOP Condition: A low-to-high transition of SDA with SCL high is a STOP condition.

Acknowledge: After sending each 8-bit word, the transmitter releases the SDA line for one bit time, during which the receiver is allowed to pull SDA low (zero) to acknowledge (ACK) that it has received each word. Device address bytes and write data bytes initiated by the host shall be acknowledged by QSFP or CXP modules. Read data bytes transmitted by the module shall be acknowledged by the host for all but the final byte read, for which the host shall respond with a STOP instead of an ACK.

Memory (Management Interface) Reset: After an interruption in protocol, power loss or system reset the management interface can be reset. Memory reset is intended only to reset the module management interface (to correct a hung bus). No other transceiver functionality is implied.

- 1) Clock up to 9 cycles.
- 2) Look for SDA high in each cycle while SCL is high.
- 3) Create a START condition as SDA is high

Device Addressing: QSFP or CXP modules require an 8-bit device address word following a start condition to enable a read or write operation. The device address word con-

sists of a mandatory sequence for the first seven most significant bits, as shown in [Figure 137](#). This is common to all QSFP and CXP devices.

Transmitter Functions (A0h)	1	0	1	0	0	0	0	R(1) / W(0)
Receiver Functions (A8h)	1	0	1	0	1	0	0	R(1) / W(0)
	Most Significant Bit							Least Significant Bit
Standard Two-wire Serial Device Address	1	0	1	0	A2	A1	A0	R/W

Table 137 CXP Device Addresses

The eighth bit of the device address is the read/write operating select bit. A read operation is initiated if this bit is set high and a write operation is initiated if this bit is set low. Upon compare of the device address the module shall output a zero (ACK) on the SDA line to acknowledge the address.

Nomenclature for all registers more than 1 bit long is MSB-LSB.

8.5 QSFP+ MEMORY MAP

The memory map for QSFP+ is summarized in [Figure 163](#).

Lower Page (1010 000x) - Required			
Addr	Size	Type	Functions
0-2	3	RO	ID and status
2-21	19	RO	Interrupt Flags
22-23	12	RO	Module Monitor - Temperature
26-27	12	RO	Module Monitor - Voltage
30-33	4	RO	Vendor Specific
34-41	8	RO	Channel Monitors - Rx Input Optical Power
42-49	8	RO	Ch. Monitors - Tx Bias Current
50-57	8	RO	Ch Monitors - Tx Optical Power
66-81	16	RO	Vendor Specific
86-97	12	R/W	Control
100-106	7	R/W	Module and Channel Mask
108-111	4	R/W	Free Side Device Properties
119-122	4	R/W	Password Change Entry Area
123-126	4	R/W	Password Entry Area
127	1	R/W	Upper Page Select Byte (00h, 01h, 02h, or 03h)

Notes:

RO = Read-Only,
R/W = Read/Write

Ranges not specified are Reserved

The term “Module” in this specification corresponds to the term “Free Side Device” (or “Module”) in the SFF-8636 documentation.

Upper Page 00 - Required			
Addr	Size	Type	Functions
128-191	64	RO	Base ID Fields
192-223	32		Extended ID Fields
224-255	32		Vendor Specific ID

Upper Page 01 - Optional			
Addr	Size	Type	Functions
128	1	RO	CC_APPS: Check code for the Application Select Table
129	1		AST Table Length (TL) (TL Range: 0-62)
130-131	2		Application Code Entry 0
132-133	2		Application Code Entry 1
other Application Code Entries			
130+2*TL-131+2*TL	2		Application Code Entry TL
Remainder			Reserved

Upper Page 02 - Optional			
Addr	Size	Type	Functions
128-255	128	R/W	User EEPROM Data If Upper Page 00 byte 129 bit 4 is set, bytes 128-137 store the CLEI code for the module.

Upper Page 03 - Optional for Cable Assemblies			
Addr	Size	Type	Functions
128-175	48	RO	Module Device Thresholds
176-223	48	RO	Channel Threshold
224	1	RO	Tx EQ & Rx Emphasis Magnitude
226-233	12	RW	Vendor Specific Channel Controls
234-241	4	RW	Channel Controls: Tx equalization, Rx emphasis, Rx amplitude, Tx/Rx squelch, Rx output disable
242-253	12	RW	Channel Monitor Masks
254-255	2	RW	Reserved

Figure 163 QSFP+ Memory Map

The memory map for QSFP+ devices is defined in [Table 138 on page 489](#), [Table 139 on page 499](#), [Table 140 on page 509](#) and [Table 141 on page 510](#) based on SFF-8636, Rev. 1.3, which supersedes the relevant sections in SFF-8436-2011-02-03 rev. 4.2 for modules capable of operation at speeds exceeding 25 Gb/s.

In these three tables the following indications are used: R = Required, O = Optional, C = Conditional (i.e., required, if the related functionality is implemented). In this section, the term “ASCII character” refers to printable ASCII characters, including characters in the range 20h to 7Eh.

Note: Indicated fields are required for InfiniBand cable assemblies. Fields that are not indicated are not required, but may be populated at the vendor’s discretion. If fields that are not listed here are unused, they should be initialized according to the SFF-8636 default value recommendations.

Note: A 1 value in a masking bit prevents the assertion of the hardware interrupt pin by the corresponding latched flag bit. Masking bits are volatile and startup with all unmasked (masking bits 0). The mask bits may be used to prevent continued interruption from ongoing conditions, which would otherwise continually reassert the hardware interrupt pin. A mask bit is allocated for each flag bit. The status bits (e.g., QSFP Bytes 3-14 and CXP Bytes 6-20) should not be affected by status of mask bits.

8.5.1 QSFP+ MEMORY MAP - LOWER PAGE

Table 138 QSFP+ lower page memory map (Sheet 1 of 10)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
0	7:0	Identifier	QSFP+ or QSFP28, as appropriate Shall be set to the same value as Upper Page 00, Byte 128 - see Table 139 - QSFP+ (0Dh): expected to be supported by SDR-EDR ports - QSFP28 (11h): expected to be supported by EDR ports and may be supported by SDR-FDR ports. To ensure compatibility with Rel. 1.3 legacy ports, modules should be programmed with the 0Dh module ID value.	R	0Dh or 11h	R	0Dh or 11h

Table 138 QSFP+ lower page memory map (Sheet 2 of 10)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
1	7:0	Revision compliance	<p>Describes version of SFF-8436 or SFF-8636 documents supported by the cable / module.</p> <p>00h = Revision not specified</p> <p>01h = SFF-8436 Rev. 4.7 or earlier</p> <p>02h = includes functionality described in revision 4.8 of SFF-8436 except that this byte and bytes 186-189 are as defined in the SFF-8636 document</p> <p>03h = SFF-8636 Rev. 1.3 or earlier</p> <p>04h = SFF-8636 Rev 1.4</p> <p>05h = SFF-8636 Rev 1.5</p> <p>06h = SFF-8636 Rev 2.0</p> <p>07h = SFF-8636 Rev 2.5, 2.6 and 2.7</p> <p>08 - FFh = un-allocated</p> <p>In case of differences between this document and relevant SFF specifications, this document shall take precedence.</p>	R		R	
2	7:3	Rsvd					
	2	Flat_mem	<p>Upper memory flat or paged</p> <p>1 = Flat:: Upper Page 0 only</p> <p>0 = Paged: 2 or more upper pages</p>	R		R	
	1	IntL	Digital state of the IntL Interrupt output pin	O		R	
	0	Data_Not_Ready	<p>Indicates module has not yet achieved power up and memory data is not ready. Bit remains high until data is ready to be read, at which time the device sets the bit low.</p> <p>The Data_Not_Ready bit shall be asserted high during free-side device reset, power up reset and prior to a valid suite of monitor readings. Once all monitor readings are valid, the bit is set low until the device is powered down or reset. Upon completion of power up reset, the free-side device shall assert IntL (if supported) low while de-asserting the Data_Not_Ready bit low. The IntL bit will remain asserted until a read is performed of the Data_Not_Ready bit.</p>	R		R	

Table 138 QSFP+ lower page memory map (Sheet 3 of 10)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
3	7	L-Tx4 LOS	Tx loss of signal Latched; value=1 indicates loss of signal at Tx input, i=4:1; else=0	O	0b	O	0b
	6	L-Tx3 LOS		O	0b	O	0b
	5	L-Tx2 LOS		O	0b	O	0b
	4	L-Tx1 LOS		O	0b	O	0b
	3	L-Rx4 LOS		O	0b	O	0b
	2	L-Rx3 LOS		O	0b	O	0b
	1	L-Rx2 LOS		O	0b	O	0b
	0	L-Rx1 LOS		O	0b	O	0b
4	7	L-Tx4 Adapt EQ Fault	Tx Adaptive Equalization Fault Latched Tx fault indicator Latched; value=1 indicates Tx fault on channel i, i=4:1; else=0	O	0b	O	0b
	6	L-Tx3 Adapt EQ Fault		O	0b	O	0b
	5	L-Tx2 Adapt EQ Fault		O	0b	O	0b
	4	L-Tx1 Adapt EQ Fault		O	0b	O	0b
	3	L-Tx4 Fault		O	0b	R	0b
	2	L-Tx3 Fault		O	0b	R	0b
	1	L-Tx2 Fault		O	0b	R	0b
	0	L-Tx1 Fault		O	0b	R	0b
5	7	L-Tx4 LOL	Tx CDR Loss of Lock Latched; value=1 indicate loss of lock on the TX CDR at the Tx input, i=4:1; else=0 Note: Masks for these alarms are in Byte 102	O	0b	O	0b
	6	L-Tx3 LOL		O	0b	O	0b
	5	L-Tx2 LOL		O	0b	O	0b
	4	L-Tx1 LOL		O	0b	O	0b
	3	L-Rx4 LOL	Rx CDR Loss of Lock Latched; value=1 indicate loss of lock on the Rx CDR at the Rx input, i=4:1; else=0 Masking bits for these alarms are in Byte 102	O	0b	O	0b
	2	L-Rx3 LOL		O	0b	O	0b
	1	L-Rx2 LOL		O	0b	O	0b
	0	L-Rx1 LOL		O	0b	O	0b

Table 138 QSFP+ lower page memory map (Sheet 4 of 10)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
6	7	L-Temp High Alarm	Hi/Low Alarms and Warnings for Temperature Thresholds for these alarms and warnings are at Upper Page 03, Bytes 128-135 Masking bits for these alarms and warnings are in Byte 103	O	0b	R	0b
	6	L-Temp Low Alarm		O	0b	O	0b
	5	L-Temp High Warning		O	0b	O	0b
	4	L-Temp Low Warning		O	0b	O	0b
	3:1	Rsvd					
	0	Initialization Complete flags		O	as defined	O	as defined
7	7	L-Vcc High Alarm	Hi/Low Alarms and Warnings for Voltage Thresholds for these alarms and warnings are at Upper Page 03, Bytes 144-151 Masking bits for these alarms and warnings are in Byte 104	O	0b	O	0b
	6	L-Vcc Low Alarm		O	0b	O	0b
	5	L-Vcc High Warning		O	0b	O	0b
	4	L-Vcc Low Warning		O	0b	O	0b
	3:0	Rsvd					
8	All	Vendor Specific					
9	7-4	L-Rx1 Power alarms and warnings	Rx Optical Power Alarms and Warnings Latched; value=1 indicates alarm or warning for high or low optical power at the receiver. Thresholds for these alarms and warnings are at Upper Page 03, Bytes 176-183.	O	0b	O	0b
	3-0	L-Rx2 Power alarms and warnings		O	0b	O	0b
10	7-4	L-Rx3 Power alarms and warnings	Masking bits for these alarms and warnings are at Upper Page 03, Bytes 242-243 7,3: Optical Power High Alarm 6,2: Optical Power Low Alarm 5,1: Optical Power High Warning 4,0: Optical Power Low Warning	O	0b	O	0b
	3-0	L-Rx4 Power alarms and warnings		O	0b	O	0b
11	7-4	L-Tx1 Bias Current alarms and warnings	Tx Laser Bias Current Alarms and Warnings Latched; value=1 indicates alarm or warning for high or low bias current at the laser transmitter Thresholds for these alarms and warnings are at Upper Page 03, Bytes 184-191	O	0b	O	0b
	3-0	L-Tx2 Bias Current alarms and warnings		O	0b	O	0b
12	7-4	L-Tx3 Bias Current alarms and warnings	Masking bits for these alarms and warnings are at Upper Page 03, Bytes 244-245 7,3: Tx Bias High Alarm 6,2: Tx Bias Low Alarm 5,1: Tx Bias High Warning 4,0: Tx Bias Low Warning	O	0b	O	0b
	3-0	L-Tx4 Bias Current alarms and warnings		O	0b	O	0b

Table 138 QSFP+ lower page memory map (Sheet 5 of 10)

Address (Byte number)	Bit(s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
13	7-4	L-Tx1 Power alarms and warnings	Tx Optical Power Alarms and Warnings Latched; value=1 indicates alarm or warning for high or low optical power at the receiver. Thresholds for these alarms and warnings are at Upper Page 03, Bytes 192-199	O	0b	O	0b
	3-0	L-Tx2 Power alarms and warnings		O	0b	O	0b
14	7-4	L-Tx3 Power alarms and warnings	Masking bits for these alarms and warnings are at Upper Page 03, Bytes 242-243	O	0b	O	0b
	3-0	L-Tx4 Power alarms and warnings	7,3: Optical Power High Alarm 6,2: Optical Power Low Alarm 5,1: Optical Power High Warning 4,0: Optical Power Low Warning	O	0b	O	0b
15-16	All	Reserved Channel monitor flags, set 4, Thresholds are at Upper Page 03, Bytes 200-207, Masking Bits are at Upper Page 03, Bytes 248-249					
17-18	All	Reserved Channel monitor flags, set 5					
19-21	All	Vendor Specific					
22	7:0	Temperature MSB	Internally measured module temperature	O	00h	R	00h
23	7:0	Temperature LSB	16 bit signed twos complement value in increments of 1/256 degrees Celsius, yielding a total range of -128 to +128 that is considered valid between -40 and +125 C. Temperature accuracy is Vendor Specific but must be better than +/-3C over specified operating temperature and voltage. Please see Vendor Specification for details on location of temperature sensor.	O	00h	R	00h
24-25	All	Reserved					
26	7:0	Supply voltage MSB	Internally measured module supply voltage	O	00h	O	00h
27	7:0	Supply voltage LSB	16 bit unsigned integer with the voltage defined as the full 16 bit value (0-65535), with LSB equal to 100 µV, yielding a total measurement range of 0 to +6.55 V. Accuracy is Vendor Specific but must be better than ±3% of the manufacturer's nominal value over specified operating temperature and voltage.	O	00h	O	00h
28-29	All	Reserved					
30-33	All	Vendor Specific					

Table 138 QSFP+ lower page memory map (Sheet 6 of 10)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
34-35	All	Rx1, MSB-LSB	Internally measured Rx input optical power. Represented as either an average received optical power or OMA, depending upon how Upper Page 00, byte 220, bit 3 is set. (1=Avg., 0=OMA).	O		O	
36-37	All	Rx2, MSB-LSB					
38-39	All	Rx3, MSB-LSB					
40-41	All	Rx4, MSB-LSB	The parameter is encoded as a 16 bit unsigned integer with the power defined as the full 16 bit value (0 to 65535) with LSB equal to 0.1 uW, yielding a total measurement range of 0 to 6.5535 mW (~-40 to +8.2 dBm). total measurement range of 0 to 6.5535 mW (~-40 to +8.2 dBm). Absolute accuracy is dependent upon the exact optical wavelength. For the vendor specified wavelength, accuracy shall be better than ±3 dB over specified temperature and voltage. This accuracy shall be maintained for input power levels up to the lesser of maximum transmitted or maximum received optical power per the appropriate standard. It shall be maintained down to the minimum transmitted power minus cable plant loss (insertion loss or passive loss) per the appropriate standard. Absolute accuracy beyond this minimum required received input optical power range is Vendor Specific.	O		O	
42-43	All	Tx1 Bias MSB-LSB	Internally measured Tx laser bias current	O		O	
44-45	All	Tx2 Bias MSB-LSB	Measured TX bias current is represented in mA as a 16-bit unsigned integer with the current defined as the full 16 bit value (0 to 65535) with LSB equal to 2 uA, yielding a total measurement range of 0 to 131 mA. Accuracy is Vendor Specific but must be better than ±10% of the manufacturer's nominal value over specified operating temperature and voltage.	O		O	
46-47	All	Tx3 Bias MSB-LSB					
48-49	All	Tx4 Bias MSB-LSB					
50-51	All	Tx1 Power MSB-LSB	Internally measured Tx laser output power	O		O	
52-53	All	Tx2 Power MSB-LSB	Represented in mW as an average power. The parameter is encoded as a 16 bit unsigned integer with the power defined as the full 16 bit value (0 to 65535) with LSB equal to 0.1 uW, yielding a total measurement range of 0 to 6.5535 mW (~-40 to +8.2 dBm). For the vendor specified wavelength, accuracy shall be better than ±3 dB over specified temperature and voltage.	O		O	
54-55	All	Tx3 Power MSB-LSB					
56-57	All	Tx4 Power MSB-LSB					
58-65	All	Reserved channel monitors, set 4					
66-81	All	Vendor Specific					

Table 138 QSFP+ lower page memory map (Sheet 7 of 10)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
82-85	All	Reserved					
86	7:4	Rsvd					
	3:0	M-Tx[4:1] Disable	Read/Write bit that allows software disable of transmitters. 1: laser of the channel is disabled, 0: laser of the channel is enabled	C	0h	R	0h
87	7-6	Rx4_Rate_select, [MSB:LSB]	Software rate select for Rx channels If Upper Page 0, Byte 221[2-3] = 01 : If (Byte141[0] = 1 : 00, 01, 10, 11 = <2.2 Gbps, 2.2-6.6, ≥6.6, Rsvd If Byte 141[1] = 1 : 00, 01, 10, 11 = <12, 12-24, 24-26, ≥26 Gbps	C	0b	C	0b
	5-4	Rx3_Rate_select, [MSB:LSB]	Not used for InfiniBand.	C	0b	C	0b
	3-2	Rx2_Rate_select, [MSB:LSB]		C	0b	C	0b
	1-0	Rx1_Rate_select, [MSB:LSB]		C	0b	C	0b
88	7-6	Tx4_Rate_select, [MSB:LSB]	Software rate select for Tx channels For format, see Byte 87	C	0b	C	0b
	5-4	Tx3_Rate_select, [MSB:LSB]	Not used for InfiniBand	C	0b	C	0b
	3-2	Tx2_Rate_select, [MSB:LSB]		C	0b	C	0b
	1-0	Tx1_Rate_select, [MSB:LSB]		C	0b	C	0b
89	All	Rx4_Application_Select	Software Application Select, per SFF-8079	O		O	
90	All	Rx3_Application_Select	Not used for InfiniBand.				
91	All	Rx2_Application_Select					
92	All	Rx1_Application_Select					

Table 138 QSFP+ lower page memory map (Sheet 8 of 10)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
93	7:3	Rsvd					
	2	High Power Class Enable (Classes 5-7)	1: enables modules to exceed Power Class 4 (3.5 W max.) - See Byte 129 for definition of Power Classes 5 to 7. 0: modules (including modules with Power Classes 5, 6 & 7) must dissipate less than 3.5W (but are not required to be fully functional).	O	0b	O	0b
	1	Power_set	Power set to Low Power Mode. Default 0	O	0b	O	0b
	0	Power_override	Override of the LPmode signal setting the power mode with software	R	0b	R	0b
94	All	Tx4 Application_Select	Software Application Select, per SFF-8079	O	O		
95	All	Tx3 Application_Select					
96	All	Tx2 Application_Select					
97	All	Tx1 Application_Select					
98	7	TX4_CDR_enable	When a bit is set to 1, the CDR for the corresponding lane is enabled. When it is set to 0, it is bypassed (i.e., disabled). CDR presence is indicated in Upper Page 0, byte 129, and CDR bypass implementation status is given in Upper Page 0, byte 194. For InfiniBand active cables and modules, the default is that Tx and Rx CDRs are disabled, for interoperability with FDR, QDR, DDR, and SDR hosts.	C	0b	C	0b
	6	TX3_CDR_enable		C	0b	C	0b
	5	TX2_CDR_enable		C	0b	C	0b
	4	TX1_CDR_enable		C	0b	C	0b
	3	RX4_CDR_enable		C	0b	C	0b
	2	RX3_CDR_enable		C	0b	C	0b
	1	RX2_CDR_enable		C	0b	C	0b
	0	RX1_CDR_enable		C	0b	C	0b
99	All	Reserved					
100	7:4	M-Tx LOS mask[4:1]	Masking bits for Tx LOS (Loss of Signal) alarms in Byte 3	C	0h	R	0h
	3:0	M-Rx LOS mask[4:1]	Masking bits for Rx LOS (Loss of Signal) alarms in Byte 3	C	0h	R	0h
101	7:4	Rsvd					
	3:0	M-Tx Fault[4:1] mask	Masking bits for Tx fault alarms in Byte 4	C	0h	R	0h

Table 138 QSFP+ lower page memory map (Sheet 9 of 10)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
102	7:4	M-Tx[4:1] CDR LOL	Masking bits for Tx CDR Loss of Lock alarms in Byte 5	C	0b	C	0b
	3:0	M-Rx[4:1] CDR LOL	Masking bit for Rx CDR Loss of Lock alarms in Byte 5	C	0b	C	0b
103	7	M-Temp High Alarm	Masking bits for Hi/Low Alarms and Warnings for Temperature in Byte 6	C	0b	C	0b
	6	M-Temp Low Alarm		C	0b	C	0b
	5	M-Temp High Warning		C	0b	C	0b
	4	M-Temp Low Warning		C	0b	C	0b
	3:0	Rsvd					
104	7	M-Vcc High Alarm	Masking bits for Hi/Low Alarms and Warnings for Voltage in Byte 7	C	0b	C	0b
	6	M-Vcc Low Alarm		C	0b	C	0b
	5	M-Vcc High Warning		C	0b	C	0b
	4	M-Vcc Low Warning		C	0b	C	0b
	3:0	Rsvd					
105-106	All	Vendor-Specific					
107	All	Reserved Read-Write					
108	7:0	Prop. Delay MSB	Propagation Delay of cable ("non-separable free-side device")	R		O	00h
109	7:0	Prop. Delay LSB	16 bit unsigned integer indicating the propagation delay, in units of 10 nsec, with fractional values rounded up to the next unit. Total measurement range: 0 to 655 µsec.	R		O	00h

Table 138 QSFP+ lower page memory map (Sheet 10 of 10)

Address (Byte number)	Bit(s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
110	7:4	Advanced Low Power Mode	Module power consumption levels ≤ 1.5 W. 0000: a power consumption limit below 1.5 W is not available. 0001: module shall consume ≤ 1 W 0010: module shall consume ≤ 0.75 W 0011: module shall consume ≤ 0.5 W	C	0000	C	0000
	3	Far Side Managed	1: the far end of the cable assembly is managed and complies with SFF-8636. 0: all other cases, including use of other management interfaces specifications and separable applications where the modules and transmission media can be physically separated from each other. For Passive or Active Cables, with QSFP+ connectors on both ends, this bit shall be set to 1.	R		O	
	2:0	Min Operating Voltage	000 - operates from nominal 3.3 V supply 001 - can operate properly from 2.5 V supply 010 - can operate properly from 1.8 V supply	C	000b	C	000b
111-112	All	Reserved - Assigned for use by PCI Express - See relevant PCI-SIG Documents					
113-118	All	Reserved Read-Write					
119-122	7:0	Password Change Entry Area		O	0	O	0
123-126	7:0	Password Entry Area	Password to control read/write access to vendor specific page 02h, or to Serial ID or other read-only information.	O	0	O	0
127	7:0	Page Select Byte	Points to currently loaded upper page of paged memory mapped to locations 128h to 255h. Valid values: 00h, 01h, 02h, or 03h.	R	00h	R	00h

8.5.2 QSFP+ MEMORY MAP - UPPER PAGE 00

Table 139 QSFP+ upper page 00 memory map (Sheet 1 of 11)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
128	7:0	Identifier	Identifier type of serial module: QSFP+: 0Dh, QSFP28: 11h (Shall be set to the same value as byte 0; see Table 138)	R	0Dh or 11h	R	0Dh or 11h
129	7:6	Power Class (1 of 2 - see bits 1:0)	Extended Identifier of module Use 00b for EEPROM-only passive copper assemblies. See bits 1:0 to for definitions of Power Classes 5-7, and Byte 93, bit 2 to enable Power Classes 5-7. 00: Power Class 1 (1.5 W max.) 01: Power Class 2 (2.0 W max.) 10: Power Class 3 (2.5 W max.) 11: Power Class 4 (3.5 W max.)	R		R	
5	Reserved						
4	CLEI Code Presence	1: CLEI code present in upper page 02h 0: No CLEI code present		R		R	
3	Tx CDR	1: CDR present in Tx 0: No CDR present in Tx		R		R	
2	Rx CDR	1: CDR present in Rx 0: No CDR present in Rx		R		R	
1:0	Power Class (2 of 2 - see bits 7:6)	Power Classes 5, 6, & 7 have been added to enable modules requiring more than 3.5W of dissipation. However, legacy systems have generally been designed to a maximum of 3.5W. To ensure legacy systems are not harmed by power classes 5, 6 or 7, a lockout feature is added in Address 93, bit 2 to enable them. A legacy system will not know about Address 129 bits 1-0 or about Address 93, bit 2. A new system will know about both and can configure Power Classes 5-7 support accordingly. The power class identifiers specify maximum power dissipation over operating conditions and lifetime with all supported settings set to worst case values. See Byte 93, bit 2 to enable 00: unused (legacy setting) 01: Power Class 5 (4.0 W max.) 10: Power Class 6 (4.5 W max.) 11: Power Class 7 (5.0 W max.)	R		R		

Table 139 QSFP+ upper page 00 memory map (Sheet 2 of 11)

Address (Byte number)	Bit(s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
130	7:0	Connector	Code for connector type For cable assemblies with no separable interface, value = 23h; if modules with separable connectors are implemented they should specify the connector type see SFF-8636	R	23h	R	
131-138	7:0	Specification Compliance	Code for electronic compatibility or optical compatibility. Set to 00h for InfiniBand. Electronic compatibility is described in Byte 164.	R	00h	R	00h
139	7:0	Encoding	Code for serial encoding algorithm. Set to 00h (Unspecified) for InfiniBand FDR & faster, since must support both 64b/66b & 8b10b.	R		R	
140	7:0	BR, nominal	Nominal bit rate Nominal bit rate, in units of 100 Mb/s, rounded to the nearest 100 Mb/s. For >25.5 Gb/s, set to FFh and use byte 164 to indicate supported IB data rates, and byte 222 for nominal bit rate in units of 250 Mb/s. Expected Values: (Allow 5% tolerance, for compatibility with other interface specifications). EDR: 255 (FFh) (Use Byte 222, in 250 Mb/s units) FDR: 140 (8Ch) QDR: 100 (64h) DDR: 50 (32h) SDR: 25 (19h)	R		R	
141	7:2	Reserved					
	1	V2: 12-26 Gbps	QSFP+ Extended Rate Select (if Upper Page 0, Byte 221[2-3] = 01) Not used for InfiniBand				
	0	V1: 2.2-6.6 Gbps					
142	7:0	Length (SMF)	Link length supported for single mode fiber (km) Value specifies the link length that is supported by the QSFP+ module using single mode fiber. Zero means that the module does not support single mode fiber or that length information must be determined from the module technology. For all cable assemblies, including active optical cables, the value shall be zero (0).	R	00h	R	

Table 139 QSFP+ upper page 00 memory map (Sheet 3 of 11)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
143	7:0	Length (OM3 50 µm optical fiber)	Link length supported for EBW 50/125 µm OM1 mm fiber, units of 2 m Value specifies the link length that is supported by the QSFP+ module using 2000 MHz*km (850 nm) extended bandwidth 50 µm core multimode fiber. Value is in units of two meters; 0 means that the module does not support OM3 fiber or that length information must be determined from the module technology. For all cable assemblies, including active optical cables, the value shall be zero (0).	R	00h	R	
144	7:0	Length (OM2 50 µm optical fiber)	Link length supported for 50/125 µm OM2 mm fiber, units of 1 m Value specifies the link length that is supported by the QSFP+ module using 500 MHz*km (850 and 1310 nm) extended bandwidth 50 µm core multimode fiber. Value is in units of one meter; 0 means that the module does not support OM2 fiber or that length information must be determined from the module technology. For all cable assemblies, including active optical cables, the value shall be zero (0).	R	00h	R	
145	7:0	Length (OM1 62.5 µm fiber)	Link length supported for 62.5/125 µm OM1 mm fiber, units of 1 m. Value specifies the link length supported by QSFP+ module using 200 MHz*km (850 nm) & 500 MHz*km (1310 nm) extended bandwidth 62.5 µm core multimode fiber. 0: module does not support OM1 fiber or that length information must be determined from the module technology. For all cable assemblies, including active optical cables, the value shall be zero (00h).	R	00h	R	
146	7:0	Length (copper cable) or Length (OM4 50 µm optical fiber)	For copper or fiber cable assemblies, equal to length in units of 1 m, rounded to nearest whole number. Populate field with 1 if cable is <1 m length, 255 if length is >254 m. For optical modules with 850 nm VCSEL device technology (Byte 147) supported length over 50/125 µm (OM4) fiber, in 2 m units..	R		R	

Table 139 QSFP+ upper page 00 memory map (Sheet 4 of 11)

Address (Byte number)	Bit(s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
147	7:4	Device technology	Informational only for InfiniBand. 0000b = 850 nm VCSEL 0001b = 1300 nm VCSEL 0010b = 1550 nm VCSEL 0011b = 1310 nm FP 0100b = 1310 nm DFB 0101b = 1550 nm DFB 0110b = 1310 nm EML 0111b = 1550 nm EML 1000b = other 1001b = 1490 nm DFB 1010b = copper cable, unequalized 1011b = copper, passive equalized 1100b = copper, near & far end limiting active 1101b = copper cable, far end limiting active 1110b = copper cable, near end limiting active 1111b = copper cable, linear active equalizers	O		O	
			1: Active wavelength control 0: No wavelength control	R		R	
			1: Cooled transmitter 0: Uncooled transmitter	O		O	
			1: APD receiver photodetector 0: PIN receiver photodetector	O		O	
			1: Transmitter tunable 0: Transmitter not tunable	O		O	
148-163	All	Vendor name	QSFP+ vendor name (ASCII) 16 character field containing ASCII characters, left-aligned and padded on the right with ASCII spaces (20h).	R		R	
164	4:0	Extended module codes	Extended module codes for InfiniBand 7-5 Reserved 4 = EDR 3 = FDR 2 = QDR 1 = DDR 0 = SDR Populate applicable bit rates with a 1b. FDR and EDR modules that perform the SDR portions of link initialization using FDR/EDR electrical high-speed signal specifications may set the SDR bit to either 0 or 1.	R		R	

Table 139 QSFP+ upper page 00 memory map (Sheet 5 of 11)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
165-167	All	Vendor OUI	Vendor company ID 3 byte field that contains the IEEE company ID for the vendor. InfiniBand requires that this field not be set to 00h.	R		R	
168-183	All	Vendor PN	Part # provided by vendor 16 byte field that contains ASCII characters, left 22 aligned and padded on the right with ASCII spaces (20h)	R		R	
184-185	All	Vendor revision	Revision level for part number provided by vendor 2 byte field that contains ASCII characters, left aligned and padded on the right with ASCII spaces (20h). The leftmost byte must be a ASCII character other than a space (20h).	R		R	

Table 139 QSFP+ upper page 00 memory map (Sheet 6 of 11)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
186	7:0	Optical Wavelength MSB or Copper cable attenuation at 2.5 GHz	Nominal laser wavelength ($\lambda = \text{value}/20$ in nm) or copper cable attenuation in dB at 2.5 GHz Optical Transmitter: 16 bit hex value with byte 186 as high order byte and byte 187 as low order byte. The laser wavelength is equal to the 16 bit integer value divided by 20 in nm (units of 0.05 nm). Limiting Active Cable (Optical or Full-Active Copper): Vendor-dependent - value may be zero or non-zero. Passive Copper and Linear Active Copper Cable: 8 bit hex value indicating the copper cable attenuation at 2.5 GHz in units of dB. Note that a tolerance of ± 1.5 dB between the programmed and actual measured values is used for PlugFest compliance testing. Half-active limiting copper cables (Near- or far-end limiting): If extra Tx equalization is required, this field should be used to request host to provide extra Tx equalization. Method for specification is not defined in this specification release.				
187	7:0	Optical wavelength LSB or Copper cable attenuation at 5.0 GHz	Nominal laser wavelength ($\lambda = \text{value}/20$ in nm) or copper cable attenuation in dB at 5 GHz Optical Transmitter: see byte 186 Limiting Active Cable (Optical or Full-Active Copper): Vendor-dependent - value may be zero or non-zero. Passive Copper and Linear Active Copper Cable: 8 bit hex value indicating the copper cable attenuation at 5.0 GHz in units of dB. Note that a tolerance of ± 1.5 dB between the programmed and actual measured values is used for PlugFest compliance testing. Half-active limiting copper cables (Near- or far-end limiting): If extra Tx equalization is required, this field should be used to request host to provide extra Tx equalization. Method for specification is not defined in this specification release.				

Table 139 QSFP+ upper page 00 memory map (Sheet 7 of 11)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
188	7:0	Optical wavelength tolerance, MSB or Copper cable attenuation at 7.0 GHz	<p>Wavelength: guaranteed range of laser wavelength (\pm value) from nominal (λ tolerance = value/200 in nm) or Copper cable: attenuation in dB at 7 GHz not including test boards</p> <p>Optical Transmitter: guaranteed \pm range of transmitter output wavelength under all normal operating conditions. 16 bit hex value with bytes 188 and 189 as high and low order bytes. The laser wavelength is equal to the 16 bit value divided by 200 in nm (units of 0.005 nm).</p> <p>Limiting Active Cable (Optical or Full-Active Copper): Vendor-dependent - value may be zero or non-zero.</p> <p>Passive Copper and Linear Active Copper Cable: 8 bit hex value indicating the cable attenuation at 7.0 GHz in units of dB. Note that a tolerance of ± 1.5 dB between the programmed and actual measured values is used for PlugFest compliance testing.</p> <p>Half-active limiting copper cables (Near- or far-end limiting): This field should be used to request host to provide extra Tx equalization. Method for specification is not defined in this specification release.</p>	R		R	
189	7:0	Optical wavelength tolerance, LSB or Copper cable attenuation at 12.9 GHz	<p>Wavelength: guaranteed range of laser wavelength (\pm value) from nominal (λ tolerance = value/200 in nm) or Copper cable: attenuation in dB at 12.9 GHz not including test boards</p> <p>Optical transmitter: see Byte 188.</p> <p>Limiting Active Cable (Optical or Full-Active Copper): Vendor-dependent - value may be zero or non-zero.</p> <p>Passive Copper and Linear Active Copper Cable: 8 bit hex value indicating the cable attenuation at 12.9 GHz in units of 1 dB. Note that a tolerance of ± 1.5 dB between the programmed and actual measured values is used for PlugFest compliance testing.</p> <p>Half-active limiting copper cables (Near- or far-end limiting): This field should be used to request host to provide extra Tx equalization. Method for specification is not defined in this specification release.</p>	R		R	

Table 139 QSFP+ upper page 00 memory map (Sheet 8 of 11)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
190	7:0	Max. case temp.	Maximum case temperature in degrees C. Eight bit value indicating maximum allowed operating case temperature in degrees C. A value of 00h indicates the standard rating of 70 C.	R		R	
191	7:0	CC_BASE	Check code for base ID fields (addresses 128-190) Low order eight bits of the sum of the contents of all the bytes from 128 to 190, inclusive.	R		R	
192	7:0	Extended Ethernet Compliance Codes	Ethernet compliance codes for 100G (AOC, -SR4, -LR4,.....) & 40GBASE-ER4, see SFF-8024 Set to 00h:Unspecified for InfiniBand products.	O		O	
193	7-4	Reserved					
	3	Tx Input Equalization Auto Adaptive Capable	Indicates if Tx Input Equalization Auto Adaptive is implemented. 1 if implemented, else 0				
	2	Tx Input Equalization Fixed Programmable	Indicates if Tx input equalization Fixed Programmable Settings is implemented. 1 if implemented, else 0	R		R	
	1	Rx Output Emphasis Fixed Programmable	Indicates if Rx Output Emphasis Fixed Programmable Settings is implemented. 1 if implemented, else 0	R		R	
	0	Rx Output Amplitude Fixed Programmable	Indicates if Rx Output Amplitude Fixed Programmable Settings is implemented. 1 if implemented, else 0	R		R	

Table 139 QSFP+ upper page 00 memory map (Sheet 9 of 11)

Address (Byte number)	Bit(s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
194	7	Tx CDR On/Off Control implemented	Indicates if Tx CDR On/Off Control (Bypass) is implemented. 1 if implemented, else 0	R	1b	R	1b
	6	Rx CDR On/Off Control implemented	Indicates if Rx CDR On/Off Control (Bypass) is implemented. 1 if implemented, else 0	R	1b	R	1b
	5	Tx CDR Loss of Lock (LOL) Flag implemented	Indicates if Tx CDR Loss of Lock (LOL) Flag is implemented. 1 if implemented, else 0	R	0b	R	
	4	Rx CDR Loss of Lock (LOL) Flag implemented	Indicates if Rx CDR Loss of Lock (LOL) Flag is implemented. 1 if implemented, else 0	R	0b	R	
	3	Rx Squelch Disable implemented	Indicates if Rx Squelch Disable is implemented. 1 if implemented, else 0	R	0b	R	
	2	Rx Output Disable capable	Indicates if Rx Output Disable is implemented. 1 if implemented, else 0	R	0b	R	
	1	Tx Squelch Disable implemented	Indicates if Tx Squelch Disable is implemented. 1 if implemented, else 0	R	0b	R	
	0	Tx Squelch implemented	Indicates if Tx Squelch is implemented. 1 if implemented, else 0	R	0b	R	
195	7	Upper page 02 provided	1 if Memory page 02 is provided, else 0	R	0b	R	
	6	Upper page 01 provided	1 if Memory page 01 is provided, else 0	R	0b	R	
	5	RATE_SELECT is implemented	1 if active control of select bits in upper page is required to change rates.. 0 if no control of select bits is required.	C	C	C	C
	4	Tx Disable implemented (also disables serial output)	1 if Tx_Disable is implemented to disable the serial output, else 0	R		R	
	3	Tx fault reporting implemented	1 if Tx_Fault signal implemented, else 0	R		R	
	2	Tx Squelch implementation	1 if TX Squelch reduces OMA, 0 if Tx Squelch reduce Pave (average power)				
	1	LOS and reporting implemented	1 if Tx Loss of Signal implemented, else 0	R		R	
	0	Reserved					

Table 139 QSFP+ upper page 00 memory map (Sheet 10 of 11)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
196-211		Vendor SN	Serial number provided by vendor 16 character field containing ASCII characters, 2 left aligned and padded on the right with ASCII spaces (20h). Serial numbers on the two ends of a cable are allowed to be different.	R		R	
212-217		Date code	Date code 6 byte field containing the vendor's YYMMDD date code, ASCII characters	R		R	
218-219		Lot Code	Vendor specific lot code, ASCII, may be blank	O		O	
220	7-4	Reserved					
	3:2	Diagnostic Monitoring Type	Bit 3: Received Optical Power Measurement: 1: Average optical power is monitored. 0: OMA (Optical Modulation Amplitude) is monitored Bit 2: Transmitter Average Optical Power Channels are: 1=monitored, 0=not monitored.	O		O	
	1-0	Reserved					
221	7-4	Reserved					
	3	Rate Selection Declaration	Rate Selection and Application Selection using bytes For InfiniBand, set to 00: no support for the rate selection mechanism used in Bytes 221(bits 3-2), Byte 141, and Bytes 87-92.	O	0b	O	0b
	2	Application select table declaration	Bit 3: 0: the module or cable does not support rate selection; 1: rate selection is implemented using extended rate selection, SFF-8636, Section 6.2.7.2 Bit 2: 0: the module or cable does not support application select and Upper Page 01 does not exist; 1: the module or cable supports rate selection using application select table mechanism. Values: 00: No Rate selection is supported 01: Rate Selection using 2-bit "Extended Rate Selection" values by Byte 141 and Bytes 87-88 10: Rate selection using Application Select method defined in Upper Page 01h. 11: Reserved.	O	0b	O	0b
	1-0	Reserved					

Table 139 QSFP+ upper page 00 memory map (Sheet 11 of 11)

Address (Byte number)	Bit (s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
222	7:0	BR, nominal - extended	Nominal bit rate Nominal bit rate, in units of 250 Mb/s, rounded off to the nearest 250 Mb/s. Extends range available in Byte 140. Supports range, in 250 Mb/s per step, up to 63.5 Gb/s. Expected values: EDR: 103 (67h) ($25.78125 = 0.250 * 103.125$) FDR, QDR, DDR, SDR: 0 (00h) - Byte 140 indicates nominal bit rate	R		R	
223	7:0	CC_EXT	Check code for the Extended ID fields (addresses 192-222) The check code shall be the low order 8 bits of the sum of the contents of all the bytes from byte 192 to 222 inclusive	R		R	
224-255		Vendor specific ID or other info	Vendor specific EEPROM Read-only	O		O	

8.5.3 QSFP+ MEMORY MAP - UPPER PAGE 02**Table 140 QSFP+ upper page 02 memory map**

Address (Byte number)	Bit(s)	Name	Description	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
128-255	All	Vendor specific	User Writable and Vendor Specific Memory The host system may read or write this memory for any purpose. If bit 4 of Page 00 byte 129 is set, however, then bytes 128-137 will be used to store the CLEI code for the module.	O		O	

8.5.4 QSFP+ MEMORY MAP - UPPER PAGE 03

Table 141 QSFP+ upper page 03 memory map (Sheet 1 of 5)

Address (Byte number)	Bit (s)	Name	Comment	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
128-129	All	Temp Hi Alarm Thrsh	Read-only; Threshold levels for Module Temperature High and Low Alarms and Warnings..	C		C	
130-131	All	Temp Lo Alarm Thrsh					
132-133		Temp Hi Warn Thrsh	Alarms and Warnings appear at Byte 6				
134-135		Temp Lo Warn Thrsh	For format, see Bytes 22-23.				
136-143	All	Reserved					
144-145	All	Vcc Hi Alarm Thrsh	Read-only; Threshold levels for Module Supply Voltage High and Low Alarms and Warnings.. Alarms and Warnings appear at Byte 7 For format, see Bytes 26-27	C		C	
146-147	All	Vcc Lo Alarm Thrsh					
148-149		Vcc Hi Warn Thrsh					
150-151		Vcc Lo Warn Thrsh					
152-159	All	Reserved					
160-175	All	Vendor Specific					
176-177	All	Rx Power High Alarm Threshold	Read-only; Threshold levels for Rx Receiver Optical Power High and Low Alarms and Warnings. Alarms and Warnings appear in Bytes 9-10 For format, see Bytes 34-35	C		C	
178-179	All	Rx Power Low Alarm Threshold					
180-181		Rx Power High Warning Threshold					
182-183		Rx Power Low Warning Threshold					
184-185	All	Tx Bias High Alarm Threshold	Read-only; Threshold levels for Tx Transmitter Bias Current High and Low Alarms and Warnings. Alarms and Warnings appear in Bytes 11-12 For format, see Bytes 42-49	C		C	
186-187	All	Tx Bias Low Alarm Threshold					
188-189		Tx Bias High Warning Threshold					
190-191		Tx Bias Low Warning Threshold					

Table 141 QSFP+ upper page 03 memory map (Sheet 2 of 5)

Address (Byte number)	Bit (s)	Name	Comment	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
192-193	All	Tx Power High Alarm Threshold	Read-only; Threshold levels for Tx Transmitter Optical Power High and Low Alarms and Warnings. Alarms and Warnings appear in Bytes 13-14 For format, see Bytes 50-57	C		C	
194-195	All	Tx Power Low Alarm Threshold					
196-197		Tx Power High Warning Threshold					
198-199		Tx Power Low Warning Threshold					
200-207	All	Reserved for channel parameter set 4					
208-223	All	Vendor Specific					
224	7:4	Tx Input Equalization Magnitude	Max Tx Input EQ magnitude supported by the module. Controls are in Bytes 234-235	O		O	
	3:0	Rx Output Emphasis magnitude Identifier	Max Rx Output Emphasis magnitude supported by the module. Controls are in Bytes 236-237				
225	All	Reserved					
226-233	All	Vendor Specific					
234	7:4	Tx1 input equalization	Transmitter Input equalization levels 1111-1011: Reserved 1010, 1001, 1000: 10, 9, 8 dB - not required for InfiniBand modules 0111: 7 dB 0110: 6 dB 0101: 5 dB 0100: 4 dB 0011: 3 dB 0010: 2 dB 0001: 1 dB 0000: 0 dB - No EQ Where the specification compliance given in bytes 131 -138 or byte 192 defines input equalization ranges (for example, CEI-28G-VSR), that specification takes precedence.	O	0000b	O	0000b
	3:0	Tx2 input equalization			0000b		0000b
235	7:4	Tx3 input equalization	0111: 7 dB 0110: 6 dB 0101: 5 dB 0100: 4 dB 0011: 3 dB 0010: 2 dB 0001: 1 dB 0000: 0 dB - No EQ Where the specification compliance given in bytes 131 -138 or byte 192 defines input equalization ranges (for example, CEI-28G-VSR), that specification takes precedence.	O	0000b	O	0000b
	3:0	Tx4 input equalization			0000b		0000b

Table 141 QSFP+ upper page 03 memory map (Sheet 3 of 5)

Address (Byte number)	Bit(s)	Name	Comment	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
236	7:4	Rx1 output emphasis	Receiver Output Emphasis levels 1xxx: Vendor Specific 0111, 0110, 0101: 7, 6, 5 dB - not required for InfiniBand modules 0100: 4 dB 0011: 3 dB 0010: 2 dB 0001: 1 dB 0000: 0 dB - No Emphasis Where the specification compliance given in bytes 131 -138 or byte 192 defines input equalization ranges (for example, CEI-28G-VSR), that specification takes precedence.	O	0000b	O	0000b
	3:0	Rx2 output emphasis		O	0000b	O	0000b
237	7:4	Rx3 output emphasis	0100: 4 dB 0011: 3 dB 0010: 2 dB 0001: 1 dB 0000: 0 dB - No Emphasis Where the specification compliance given in bytes 131 -138 or byte 192 defines input equalization ranges (for example, CEI-28G-VSR), that specification takes precedence.	O	0000b	O	0000b
	3:0	Rx4 output emphasis		O	0000b	O	0000b
238	7:4	Rx1 output amplitude	Output amplitude level with no equalization enabled. See Table 85 on page 342 0000b = range 0 0001b = range 1 0010b = range 2 0011-0111b = reserved 1xxxb = Rsvd	O	0000b	O	0000b
	3:0	Rx2 output amplitude		O	0000b	O	0000b
239	7:4	Rx3 output amplitude	0000b = range 0 0001b = range 1 0010b = range 2 0011-0111b = reserved 1xxxb = Rsvd	O	0000b	O	0000b
	3:0	Rx4 output amplitude		O	0000b	O	0000b
240	7:4	Rx[4:1] SQ Disable	Value = 1 to disable receiver output squelch, else 0	O	0b	O	0b
	3:0	Tx[4:1] SQ Disable	Value = 1 to disable transmitter output squelch, else 0	O	0b	O	0b
241	7:4	Rx[4:1] Output Disable	Value = 1 to disable receiver output, else 0	O	0b	O	0b
	3:0	Tx[4:1] Adaptive Equalization Control	Tx Input Adaptive Equalization Control. Upper Page 0, Byte 193, bit 3 identifies support 1: Enable Adaptive Equalization 0: Disable	O	0b	O	0b

Table 141 QSFP+ upper page 03 memory map (Sheet 4 of 5)

Address (Byte number)	Bit (s)	Name	Comment	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
242	7	M-Rx1 PwrHigh Alrm	Masking bits for Rx Channel High and Low Power Alarms and Warnings for Channels 1 & 2	C	0b	C	0b
	6	M-Rx1 PwrLow Alrm		C	0b	C	0b
	5	M-Rx1 PwrHigh Wrng		C	0b	C	0b
	4	M-Rx1 PwrLow Wrng		C	0b	C	0b
	3	M-Rx2 PwrHigh Alrm		C	0b	C	0b
	2	M-Rx2 PwrLow Alrm		C	0b	C	0b
	1	M-Rx2 PwrHigh Wrng		C	0b	C	0b
	0	M-Rx2 PwrLow Wrng		C	0b	C	0b
243	7	M-Rx3 PwrHigh Alrm	Masking bits for Rx Channel High and Low Power Alarms and Warnings for Channels 3 & 4	C	0b	C	0b
	6	M-Rx3 PwrLow Alrm		C	0b	C	0b
	5	M-Rx3 PwrHigh Wrng		C	0b	C	0b
	4	M-Rx3 PwrLow Wrng		C	0b	C	0b
	3	M-Rx4 PwrHigh Alrm		C	0b	C	0b
	2	M-Rx4 PwrLow Alrm		C	0b	C	0b
	1	M-Rx4 PwrHigh Wrng		C	0b	C	0b
	0	M-Rx4 PwrLow Wrng		C	0b	C	0b
244	7	M-Tx1 BiasHigh Alrm	Masking bits for Tx Channel High and Low Bias Alarms and Warnings for Channels 1 & 2	C	0b	C	0b
	6	M-Tx1 BiasLow Alrm		C	0b	C	0b
	5	M-Tx1 BiasHigh Wrng		C	0b	C	0b
	4	M-Tx1 BiasLow Wrng		C	0b	C	0b
	3	M-Tx2 BiasHigh Alrm		C	0b	C	0b
	2	M-Tx2 BiasLow Alrm		C	0b	C	0b
	1	M-Tx2 BiasHigh Wrng		C	0b	C	0b
	0	M-Tx2 BiasLow Wrng		C	0b	C	0b

Table 141 QSFP+ upper page 03 memory map (Sheet 5 of 5)

Address (Byte number)	Bit(s)	Name	Comment	Passive copper, Active Copper, Active Optical		Optical Module	
				R/O/C	Default	R/O/C	Default
245	7	M-Tx3 BiasHigh Alrm	Masking bits for Tx Channel High and Low Bias Alarms and Warnings for Channels 3 & 4	C	0b	C	0b
	6	M-Tx3 BiasLow Alrm		C	0b	C	0b
	5	M-Tx3 BiasHigh Wrng		C	0b	C	0b
	4	M-Tx3 BiasLow Wrng		C	0b	C	0b
	3	M-Tx4 BiasHigh Alrm		C	0b	C	0b
	2	M-Tx4 BiasLow Alrm		C	0b	C	0b
	1	M-Tx4 BiasHigh Wrng		C	0b	C	0b
	0	M-Tx4 BiasLow Wrng		C	0b	C	0b
246	7	M-Tx1 PwrHigh Alrm	Masking bits for Tx Channel High and Low Power Alarms and Warnings for Channels 1 & 2	C	0b	C	0b
	6	M-Tx1 PwrLow Alrm		C	0b	C	0b
	5	M-Tx1 PwrHigh Wrng		C	0b	C	0b
	4	M-Tx1 PwrLow Wrng		C	0b	C	0b
	3	M-Tx2 PwrHigh Alrm		C	0b	C	0b
	2	M-Tx2 PwrLow Alrm		C	0b	C	0b
	1	M-Tx2 PwrHigh Wrng		C	0b	C	0b
	0	M-Tx2 PwrLow Wrng		C	0b	C	0b
247	7	M-Tx3 PwrHigh Alrm	Masking bits for Tx Channel High and Low Power Alarms and Warnings for Channels 3 & 4	C	0b	C	0b
	6	M-Tx3 PwrLow Alrm		C	0b	C	0b
	5	M-Tx3 PwrHigh Wrng		C	0b	C	0b
	4	M-Tx3 PwrLow Wrng		C	0b	C	0b
	3	M-Tx4 PwrHigh Alrm		C	0b	C	0b
	2	M-Tx4 PwrLow Alrm		C	0b	C	0b
	1	M-Tx4 PwrHigh Wrng		C	0b	C	0b
	0	M-Tx4 PwrLow Wrng		C	0b	C	0b
248-249	All	Reserved channel monitor masks set 4					
250-253	All	Reserved channel monitor masks					
254-255	All	Reserved					

8.6 READ/WRITE FUNCTIONALITY FOR CXP

This section describes the functionality for read/write operations for CXP modules.
Read/write functionality for QSFP devices is defined in the QSFP documentation.

8.6.1 CXP MEMORY ADDRESS COUNTER (READ AND WRITE OPERATIONS)

CXP devices maintain an internal data word address counter containing the last address accessed during the latest read or write operation, incremented by one. The address counter is incremented whenever a data word is received or sent by the transceiver. This address stays valid between operations as long as CXP power is maintained. The address “roll over” during read and writes operations is from the last byte of the 128-byte memory page to the first byte of the same page.

8.6.2 READ OPERATIONS (CURRENT ADDRESS READ)

A current address read operation requires only the device address read word (10100001-Tx base address or 10101001-Rx base address) be sent, see [Figure 164 on page 515](#). Once acknowledged by the CXP, the current address data word is serially clocked out. The host does not respond with an acknowledge, but does generate a STOP condition once the data word is read.

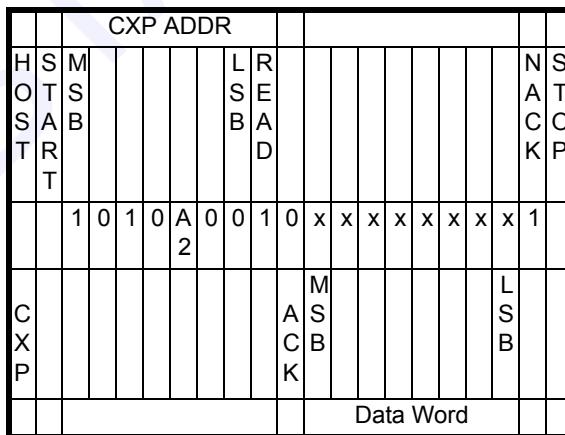


Figure 164 Read Operation on Current Address

8.6.3 READ OPERATIONS (RANDOM READ)

A random read operation requires a “dummy” write operation to load in the target byte address as shown in [Figure 165 on page 516](#). This is accomplished by the following sequence: The target 8-bit data word address is sent following the device address write word (1010 0000 or 1010 1000) and acknowledged by the CXP module. The host then generates another START condition (aborting the dummy write without incrementing the counter) and a current address read by sending a device read base address (1010 0001 for Tx or 1010 1001 for Rx). The CXP acknowledges the device address and serially

clocks out the requested data word. The host does not respond with an acknowledge, but does generate a STOP condition once the data word is read.

	CXP ADDR								MEMORY ADDR								CXP ADDR															
H O S T T	S T R T	M A B			L S B	W R I T E	M S B							L S B	S T A R T	M S B					R E A D							N S A C O K P				
		1 0 1 0 A 0 0 0 0 x x x x x x x 0			1 0 1 0 A 0 0 0 0 x x x x x x x 0									1 0 1 0 A 0 0 0 0 x x x x x x x 1																		
C X P							A C K							A C K								A C K	M S B					L S B				
																												Data Word 1				

Figure 165 Random Read

8.6.4 READ OPERATIONS (SEQUENTIAL READ)

Sequential reads are initiated by either a current address read as depicted in [Figure 167](#) or a random address read as depicted in [Figure 167 on page 517](#). To specify a sequential read, the host responds with an acknowledge (instead of a STOP) after each data word. As long as the CXP device receives an acknowledge, it shall serially clock out sequential data words. The sequence is terminated when the host responds with a NACK and a STOP instead of an acknowledge.

Figure 166 Sequential Address Read Starting at Current Address

Figure 167 Sequential Address Read Starting with Random CXP Read

8.6.5 WRITE OPERATIONS (BYTE WRITE)

A write operation requires an 8-bit data word address following the device address write word (0101 efg0) and acknowledgment, see [Figure 168 on page 518](#). Upon receipt of this address, the CXP shall again respond with a zero (ACK) to acknowledge and then clock in the first 8-bit data word. Following the receipt of the 8-bit data word, the CXP shall output a zero (ACK) and the host master must terminate the write sequence with a STOP condition for the write cycle to begin. If a START condition is sent in place of a STOP condition (i.e. a repeated START per the two wire interface specification) the write is aborted and the data received during that operation is discarded. Upon receipt of the proper STOP condition, the CXP enters an internally timed write cycle, t_{WR} , to internal memory. The CXP disables its management interface input during this write cycle and shall not respond or acknowledge subsequent commands until the write is complete.

Note that two wire interface “Combined Format” using repeated START conditions is not supported on CXP write commands.

	CXP ADDR								MEMORY ADDR								Data Word									
H	S	M						L	W	M					L	M				L	S	T				
O	T	S	A	B				S	R	S					S	S				S	B	O				
T	R	T						B	I	B					B	B				B						
			1	0	1	0	A	0	0	0	x	x	x	x	x	x	x	0	x	x	x	x	x	x	0	
C	X	P								A									A						A	C
										C									C						K	K

Figure 168 Write Byte Operation

8.6.6 WRITE OPERATIONS (SEQUENTIAL WRITE)

A CXP device shall support up to a 4 sequential byte write without repeatedly sending CXP address and memory address information as shown in [Figure 169 on page 519](#). A “sequential” write is initiated the same way as a single byte write, but the host master does not send a stop condition after the first word is clocked in. Instead, after the CXP acknowledges receipt of the first data word, the host can transmit up to three more data words. The CXP shall send an acknowledge after each data word received. The host must terminate the sequential write sequence with a STOP condition or the write opera-

tion shall be aborted and data discarded. Note that two wire interface “combined format” using repeated START conditions is not supported on CXP write.

	CXP ADDR				MEM ADDR				Data Word 1				Data Word 2				Data Word 3				Data Word 4					
H	S	M	O	T	S	R	M	L	S	M	L	S	M	L	S	M	L	S	M	L	S	M	L	T	O	P
A	B	I	B	E																						
1	0	1	0	A	0	0	0	x	x	x	x	x	x	x	0	x	x	x	x	x	0	x	x	x	x	
0	2																									0
C	X	P						A			A			A			A			A			A		A	C
								C			C			C			C			C			C		C	K

Figure 169 Sequential Write Operation

8.6.7 WRITE OPERATIONS (ACKNOWLEDGE POLLING)

Once the CXP internally timed write cycle has begun (and inputs are being ignored on the bus) acknowledge polling can be used to determine when the write operation is complete. This involves sending a START condition followed by the device address word. Only if the internal write cycle is complete shall the CXP respond with an acknowledge to subsequent commands, indicating read or write operations can continue.

8.7 CXP MEMORY MAP

This section defines the Memory Map for the CXP transceiver used for serial ID, digital monitoring and certain control functions. The interface is mandatory for all CXP devices. The interface has been designed largely after the XFP MSA as defined in INF-8077i Rev.4.0. The memory map has been modified to accommodate 12 lanes per direction and to limit the required memory space. Paging on upper pages is used to allow slower access to less time-critical information.

The memory map has also been configured to support a range of device types, from simple passive cables with only EEPROM chips with two-wire serial interfaces for identification, to, for example, optical transceivers with tunable equalization and per-lane optical power monitoring. All devices conforming to this interface are required to implement a basic memory map, including various fields such as fields to identify the device type and manufacturer.

The structure of the memory map is shown in [Figure 170 on page 521](#). It includes two ranges of serial addresses, at A0h (for Tx and basic required functions), and A8h (for Rx and optional extensions). Each address (A0h and A8h) contains one lower page and at least one upper page (00h), with one optional other upper page (01h) per address range. Each page contains 128 bytes of address space. The lower page or pages contain Read-Only information, and may contain Read-Write fields as well, for more sophisticated devices. The upper pages 00h and 01h contain only Read-Only (RO) information. Only the first upper page (00h) is required. Other(s) are optional, to allow construction of modules or cables with non-pageable memory. In this section, the term “ASCII character” refers to printable ASCII characters, including characters in the range 20h to 7Eh.

This structure permits timely access to fields in the lower page, which contain time-critical information, such interrupt flags, alarms, critical monitors (temperature, voltage,...) and per-lane control. Read-only device information such as serial ID information, vendor information, is available in Upper Page 00, which is identical for both A0h and A8h device addresses. Less time-critical Read-Only information on more complex devices, such as threshold settings or per-channel monitors, are available with the optional Upper Page Select function.

In Rel. 1.3, an upper page 02H, in the Tx Address range, has been added, for compatibility with the capabilities described in SFF-8472, Diagnostic Monitoring for Optical Transceivers. This upper page is optional, but if it is implemented, it must be implemented as described. The structure also allows for address expansion by adding additional upper pages as needed. This expansion is vendor-specific, and is not described in this document.

Note: A 1 value in a masking bit prevents the assertion of the hardware interrupt pin by the corresponding latched flag bit. Masking bits are volatile and startup with all unmasked (masking bits 0). The mask bits may be used to prevent continued interruption from ongoing conditions, which would otherwise continually reassert the hardware interrupt pin. A mask bit is allocated for each flag bit. The status bits (e.g., QSFP Bytes 3-14 and CXP Bytes 6-20) should not be affected by status of mask bits.

Tx Lower Page (1010 000x / A0x) - Required		
Byte	Type	Functions
0-6	RO	Tx Status: Page 02 presence, A8h presence, Flat/Paging memory presence, Interrupt, Data not Ready, Version control, Loss of Signal, Fault, Summary of Alarms
7-21	RO	Latched Tx Alarms: Loss of Signal, Fault, Per-channel Alarms (Power or Bias Current high/low), Module alarms (Temp, Vcc3.3 or Vcc12), per-channel Loss of Lock for Tx CDR circuits
22-29	RO	Module Monitors: Temps, Voltages
38-39	RO	Module Monitor: Elapsed Operating Time
40-41	RW	Module Control: Rate / Application Select
42	RW	High-Power Mode control
43	RW	Tx CDR Bypass/Enable Control
51	RW	Module Control: Tx Reset
52-67	RW	Tx Channel Control: Channel Disable, Output Disable, Squelch, Polarity Flip, Margin, Equalization
95-109	RW	Masks for Alarms: Channel (LOS, Fault), Channel Internal (Power or Current high/low), Module (Temp, Voltage), Loss of Lock for CDR circuit
110-118	RW	Vendor-Specific Read/Write Registers for Tx
119-126	RW	Password
127	RW	Upper Page Select Byte (00h or 01h or 02h)

Tx Upper Page 01h (Optional)		
Byte	Type	Functions
128-167	RO	Module Alarm Threshold Settings
168-179	RO	Channel Alarm Threshold Settings
180-181	RO	Checksum
182-229	RO	Per-Channel Monitors: Tx Bias current and light output
230-255		Reserved - Vendor-Specific Tx Functions

Rx Lower Page (1010 100x / A8x) - Optional		
Byte	Type	Functions
0-6	RO	Rx Status: Flat/Paging memory, Interrupt, Data not Ready, Loss of Signal, Fault, Summary of Alarms
7-21	RO	Latched Rx Alarms: Loss of Signal, Fault, Per-channel Alarms (Power or Current high/low), Device alarms (temp, Vcc3.3 or Vcc12), per-channel Loss of Lock for Tx CDR circuits
22-29	RO	Module Monitors: Temps, Voltages
38-39	RO	Module Monitor: Elapsed Operating Time
40-41	RW	Module Control: Rate / Application Select
43	RW	Tx CDR Bypass/Enable Control
51	RW	Module Control: Rx Reset
52-73	RW	Rx Channel Control: Channel Disable, Output Disable, Squelch, Polarity Flip, Margin, Amplitude, De-emphasis
95-106	RW	Masks for Alarms: Channel (LOS, Fault), Channel Internal (Power high/low), Module (Temp, Voltage), Loss of Lock for CDR circuit
109-118	RW	Vendor-Specific Read/Write Registers for Rx
119-126	RW	Password
127	RW	Upper Page Select Byte (00h, 01h, or 02h)

Upper Page 00h (Identical for Tx & Rx) Required		
Bytes	Type	Functions
128	RO	SFF-style Type Identifier
129	RO	Power Class, Tx/Rx CDR Presence
130-146	RO	Device Description: Cable & Connector, Power supplies, Max Case Temp, Min/Max Signal Rate, Laser wavelength or copper attenuation, and supported functions
147	RO	Description: Device Technology
148	RO	Max Power Utilization
149	RO	Data rates supported, cable fanout
150-151	RO	Cable length, 0.5 m units
152-222	RO	Vendor Information: Name & OUI, PN & PN rev, Serial number, Date code, Lot code/customer-specific information
223	RO	Checksum on 128-222
224-255	RO	Vendor-specific Read-only Registers

Tx and/or Rx Upper Page 02h (Optional)		
Byte	Type	Functions
128-247	RW	User Writable EEPROM (120 B)
248-255		Reserved - Vendor Specific (8B)

Rx Upper Page 01h (Optional)		
Byte	Type	Functions
128-167	RO	Module Alarm Threshold Settings
168-179	RO	Channel Alarm Threshold Settings
180-181	RO	Checksum
206-229	RO	Per-Channel Monitors: Rx Input power
230-255		Reserved - Vendor-Specific Rx Functions

Figure 170 Memory map two wire serial addresses A0x (Tx) & A8x (Rx)

8.7.1 CXP MEMORY MAP - TX LOWER PAGE

[Table 142](#) describes the memory map for the Tx lower page.

Table 142 CXP Tx Lower Page Memory Map (Sheet 1 of 7)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional-/ (Not Applicable)		
					Passive	Active Electrc	Active Optical
0 00h	All	Reserved - 1B	Coded 00h (unspecified)	RO			
1 01h	All	Reserved: Extended Status	00h	RO			
2 02h	7-6	Reserved	0000b	RO			
	5-4	Tx and/or Rx Upper Page 02 Presence	00b = no optional Upper Page 02 supported 10b = Upper Page 02 supported, Tx (A0h) address 01b = Upper Page 02 supported, Rx (A8h) address 11b = Upper Page 02 supported, accessible through either Tx (A0h) address or Rx (A8h) address		R	R	R
	3	Rx A8h Device Address Presence	0 = Rx Device Address fields (A8h) are present. 1 = Rx Device Address fields (A8h) are not present		R	R	R
	2	Flat/Paging Memory Presence	0 = Paging is present. 1 = Upper Page 00h only, no other Tx Upper pages		R	R	R
	1	Int_L Status	Coded 1 for asserted Int_L. Clears to 0 when all flags including LOS and Fault are cleared.		R	R	R
	0	Data_Not_Ready	Indicates transceiver has not yet achieved power up and monitor data is not ready. Bit remains high until data is ready to be read at which time the device sets the bit low. The Data_Not_Ready bit shall be asserted high during free-side device reset, power up reset and prior to a valid suite of monitor readings. Once all monitor readings are valid, the bit is set low until the device is powered down or reset. Upon completion of power up reset, the free-side device shall assert IntL (if supported) low while de-asserting the Data_Not_Ready bit low. The IntL bit will remain asserted until a read is performed of the Data_Not_Ready bit.		R	R	R
3 03h	All	Version control - used to identify to which version of the CXP specification the cable or optical transceiver is compliant	00h = Undefined 01h = compliant to CXP rev v1.1.2011-03-09 or earlier 02h = compliant to CXP Specification as defined in InfiniBand Volume 2 Rev 1.3 03h = compliant to CXP Specification as defined in InfiniBand Volume 2 Rev 1.3.1 04h-FFh = reserved	RO	R	R	R
4-5	All	Reserved - 3B	Reserved for Status info	RO			

Table 142 CXP Tx Lower Page Memory Map (Sheet 2 of 7)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/- (Not Applicable)		
					Passive	Active Electrl	Active Optical
6 06h	7	LOS Tx Status Summary	Coded 1 when a LOS Tx flag (bytes 7-8) is asserted for any channel, else 0. Clears when LOS flags are cleared.	RO	-	O	O
	6	Reserved	Coded 0b. Reserved for Rx LOS Status Summary in Rx Lower Page				
	5	Fault Tx Status Summary	Coded 1 when a Fault Tx flag (bytes 9-10) is asserted for any channel, else 0. Clears when all Fault flags are cleared.		-	O	O
	4	Bias Tx Status Summary	Coded 1 when a Tx Bias Hi-Lo Alarm (bytes 11-13) is asserted, else 0. Clears when alarm is cleared.		-	-	O
	3	Power Tx Status Summary	Coded 1 when a Tx Optical Power Hi-Lo Alarm (bytes 14-16) is asserted, else 0. Clears when alarm is cleared.		-	-	O
	2	Reserved	Coded 0b. Reserved for Rx Optical Power Hi-Lo Alarm in Rx Lower Page				
	1	Module Tx Status Summary	Coded 1 when any Tx Temperature or Voltage alarm (bytes 17-18) or reserved module Tx monitor alarm (reserved in byte 19) or Loss of Lock Tx (Bytes 20-21) are asserted, else 0. Clears when all these alarms are cleared.		-	O	O
	0	Reserved	Reserved for other Module Monitor alarm				
7 07h	7-4	Reserved	Loss of Signal Tx Channel: Coded 1 when asserted, Latched, Clears on Read.	RO	-	O	O
	3-0	L-LOS Tx11 - Tx08					
8 08h	7-0	L-LOS Tx07 - Tx00	Byte 7, bit 3 encodes for channel Tx11 Byte 7, bit 2 encodes for channel Tx10 Byte 7, bit 1 encodes for channel Tx09 Byte 7, bit 0 encodes for channel Tx08 Byte 8, bits 7-0 encode for channels Tx07-Tx00 respectively. The following registers follow the same pattern Masking bits for these alarms are at Lower Page, Bytes 95-96.				
	9 09h	7-4	Reserved				
	3-0	L-Fault Tx11 - Tx08		-	O	O	
	10 0Ah	7-0					L-Fault Tx07 - Tx00

Table 142 CXP Tx Lower Page Memory Map (Sheet 3 of 7)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/- (Not Applicable)		
					Passive	Active Electrl	Active Optical
11 0Bh	7-0	L-Bias Hi-Lo Alarm Tx11 - Tx08	Tx Bias Current Hi-Lo Alarm Latched: 2 bits / channel Coded 10b when High Bias current alarm is asserted Coded 01b when Low Bias current alarm is asserted Coded 00b for no alarm.	RO	-	-	O
12 0Ch	7-0	L-Bias Hi-Lo Alarm Tx07 - Tx04	Latched, Clears on Read. Thresholds for these alarms are at Upper Page 01, Bytes 168-171.				
13 0Dh	7-0	L-Bias Hi-Lo Alarm Tx03 - Tx00	Masking bits for these alarms are at Lower Page, Bytes 99-101.				
14 0Eh	7-0	L-Power Hi-Lo Alarm Tx11 - Tx08	Tx Optical Power Hi-Lo Alarm Latched, 2 bits per channel Coded 10b when High Optical output power alarm is asserted		-	-	O
15 0Fh	7-0	L-Power Hi-Lo Alarm Tx07 - Tx04	Coded 01b when Low Optical output power alarm is asserted				
16 10h	7-0	L-Power Hi-Lo Alarm Tx03 - Tx00	Coded 00b for no alarm. Latched, Clears on Read. Thresholds for these alarms are at Upper Page 01, Bytes 172-175 Masking bits for these alarms are at Lower Page, Bytes 102-104.				
17 11h	7	L-Temp High Alarm - Tx	High Internal Temperature Alarm Latched: Coded 1 when asserted, Latched, Clears on Read. Thresholds for all temperature alarms are at Upper Page 01, Bytes 128-135. Masking bit for all Temp alarms are at Lower Page, Byte 105.	RO	O	O	R
	6	L-Temp Low Alarm - Tx	Low Internal Temperature Alarm Latched: Coded 1 when asserted, Latched, Clears on Read.				
	5-0	Reserved					
18 12h	7	L-Vcc3.3 High Alarm - Tx	High Internal Vcc3.3 Alarm Latched: Coded 1 when asserted, Latched, Clears on Read. Thresholds for all voltage alarms are at Upper Page 01, Bytes 144-151. Masking bits for all voltage alarms are at Lower Page, Byte 106.	RO	-	O	O
	6	L-Vcc3.3 Low Alarm - Tx	Low Internal Vcc3.3 Alarm Latched: Coded 1 when asserted, Latched, Clears on Read.				
	5-4	Reserved					
	3	L-Vcc12 High Alarm - Tx	High Internal Vcc12 Alarm Latched: Coded 1 when asserted, Latched, Clears on Read.		-	O	O
	2	L-Vcc12 Low Alarm - Tx	Low Internal Vcc12 Alarm Latched: Coded 1 when asserted, Latched, Clears on Read.				
	1-0	Reserved					

Table 142 CXP Tx Lower Page Memory Map (Sheet 4 of 7)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/- (Not Applicable)		
					Passive	Active Electrl	Active Optical
19	All	Reserved	Reserved - Tx module alarm	RO			
20 14h	7-4	Reserved	Loss of Lock, Tx CDR: Coded 1 when asserted (i.e., when CDR is enabled and not locked to data stream), 0 when (CDR is enabled AND locked) OR (CDR is bypassed). Latched.	RO	-	O	O
	3-0	LOL Tx11 - Tx08					
21	7-0	LOL Tx07 - Tx00					
22 16h	All	1st Tx Temp Monitor MSB	1st Internal Temperature Monitor for Tx MSB: Integer part coded in signed 2's complement. Tolerance is $\pm 3^{\circ}\text{C}$. If implemented, sensor location is vendor-defined - intended to be most temperature-sensitive Tx location.	RO	O	O	R
23 17h	All	1st Tx Temp Monitor LSB	1st Internal Temperature Monitor for Tx LSB: Fractional part in units of $1/256$ coded in binary.				
24-25 18-19h	All	2nd Tx Temp Monitor	2nd Internal Temperature Monitor for Tx. Same 2 Byte format as 1st. If implemented, sensor location is vendor-defined.	RO	O	O	O
26-27 1A-1Bh	All	Tx Vcc3.3 Monitor MSB Tx Vcc3.3 Monitor LSB	Internal Vcc3.3 Monitor for Tx: Voltage in $100 \mu\text{V}$ units coded as 16 bit unsigned integer, Low byte is MSB. Tolerance is $\pm 0.10\%$.	RO	O	O	O
28-29 1C-1Dh	All	Tx Vcc12 Monitor MSB Tx Vcc12 Monitor LSB	Internal Vcc12 Monitor for Tx: Voltage in $250 \mu\text{V}$ units coded as 16 bit unsigned integer, Low byte is MSB. Tolerance is $\pm 0.1\%$.	RO	O	O	O
30-37	All	Reserved - 8B	Reserved - Module Monitors	RO			
38-39 26-27h	All	Elapsed Operating Time	Elapsed (Power-on) Operating Time: Elapsed time in 2 hour units coded as 16 bit unsigned integer, Low byte is MSB, Tolerance is $\pm 10\%$	RO	O	O	O
40 28h	All	Tx Module application select	Format to be determined as other applications besides InfiniBand arise	RW	-	O	O
41 29h	7-5	Reserved	Reserved - Rate Select	RW			
	4-0	Tx Rate Select	Tx Rate Select / optimization bit-map Bit 4: EDR Bit 3:FDR Bit 2: QDR Bit 1: DDR Bit 0:SDR Examples: 00111: Configured for QDR / DDR / SDR operation 01000: Configured for FDR operation 10000: Configured for EDR operation 00000: no info provided	RW	-	O	O
42 2Ah	7-1	Reserved	0: Device or cable may not draw more than 6 Watts of power. 1: Device or cable may draw more than 6.0 W, up to limit denoted in Upper Page 00, Byte 148 (94h)	RW	O	O	O
	0	High-Power Mode					

Table 142 CXP Tx Lower Page Memory Map (Sheet 5 of 7)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/- (Not Applicable)		
					Passive	Active Electrl	Active Optical
43 2Bh	7-1	Reserved	0: TX CDR is Bypassed/Enabled. 1: TX CDR is Enabled.	RW	-	O	O
	0	TX CDR Bypassed/Enabled					
44-50	All	Reserved - 8B	Reserved - Module Control	RW	-	-	-
51 33h	7-1	Reserved	Reset: Writing 1 return all registers on Tx pages (non-volatile RW, if present in vendor-specific area) to factory default values. Reads 0 after operation.	RW	R	R	R
	0	Reset					
52 34h	7-4	Reserved	Tx Channel Disable: Writing 1 disables the whole channel. For optical module output power reporting, see Tx Upper Page 01h, Bytes 206-229. Default is 0.	RW	-	O	O
	3-0	Channel Disable Tx11 - Tx08					
53	7-0	Channel Disable Tx07 - Tx00					
54 36h	7-4	Reserved	Tx Output Disable: Writing 1 disables (i.e., squelches) the output for the channel. Some internal circuitry (e.g., CDR) may be kept active. If implemented, functionality is vendor-specific. Default is 0.	RW	-	O	O
	3-0	Output Disable Tx11 - Tx08					
55	7-0	Output Disable Tx07 - Tx00					
56 38h	7-4	Reserved	Tx Squelch Disable: Writing 1 disables squelch for the channel. Default is 0 (Squelch enabled).	RW	-	O	O
	3-0	Squelch Disable Tx11-Tx08					
57	7-0	Squelch Disable Tx07 - Tx00					
58 3Ah	7-4	Reserved	Tx Channel input polarity flip: Writing 1 inverts the polarity of outputs relative to the inputs. Default is 0 (No polarity flip)	RW	-	O	O
	3-0	Polarity flip Tx11 - Tx08					
59	7-0	Polarity flip Tx07 - Tx00					
60 3Ch	7-4	Reserved	Tx Channel Margin Activation: Writing 1 places Tx in "Margin Mode", reducing signal amplitude (electrical) or OMA (optical) by a vendor-specific amount (nominally equivalent to roughly 1 dB). Intended use is for testing of link signal integrity margin. Specific implementation is vendor-specific, and may be different for modules vs. active cables (AOCs or AECs). Default is 0.	RW	-	O	O
	3-0	Margin Select Tx11 - Tx08					
61	7-0	Margin Select Tx07 - Tx00					

Table 142 CXP Tx Lower Page Memory Map (Sheet 6 of 7)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/- (Not Applicable)		
					Passive	Active Electrl	Active Optical
62 3Eh	7-4 3-0	Input Equalization Tx11 Input Equalization Tx10	Tx Input Equalization Control: Four bit code blocks (bits 7-4 or 3-0) are assigned to each channel.	RW	-	O	O
63 3Fh	7-4 3-0	Input Equalization Tx09 Input Equalization Tx08	Codes 1xxxb are reserved. Writing 0111b calls for full-scale equalization.				
64 40h	7-4 3-0	Input Equalization Tx07 Input Equalization Tx06	Writing 0000b calls for no equalization. Intermediate code values call for intermediate levels of equalization.				
65 41h	7-4 3-0	Input Equalization Tx05 Input Equalization Tx04	Exact equalization parameters (e.g., crossover frequency, equalization levels, slopes vs. frequency, etc.) are vendor-specific, and shall be appropriate to the intended use of the module or device.				
66 42h	7-4 3-0	Input Equalization Tx03 Input Equalization Tx02					
67 43h	7-4 3-0	Input Equalization Tx01 Input Equalization Tx00					
68-94	All	Reserved - 27B	Reserved - Per-Channel Control	RW			
95 5Fh	7-4 3-0	Reserved Mask LOS Flag Tx11 - Tx08	Mask Tx LOS Flag: Writing 1 prevents Int_L on Tx LOS. Default = 0; mask is required if corresponding optional alarm is implemented.	RW	-	C	C
96	7-0	Mask LOS Flag Tx07 - Tx00					
97 61h	7-4 3-0	Reserved Mask Tx Fault Flag Tx11 - Tx08	Mask Tx Fault Flag: Writing 1 prevents Int_L on Tx Fault. Default = 0; mask is required if corresponding optional alarm is implemented.		-	C	C
98	7-0	Mask Tx Fault Flag Tx07 - Tx00					
99 63h	7-0	Mask Bias Hi-Lo Alarm Tx11 - Tx08	Mask Tx Bias Current Hi-Lo Alarm: Writing 10b prevents Int_L on Tx High Bias Current	RW	-	-	C
100 64h	7-0	Mask Bias Hi-Lo Alarm Tx07 - Tx04	Writing 01b prevents Int_L on Tx Low Bias Current				
101 65h	7-0	Mask Bias Hi-Lo Alarm Tx03 - Tx00	Writing 11b prevents Int_L on both High and Low Bias Current Alarms. Default = 00b; mask is required if corresponding optional alarm is implemented.				
102 66h	7-0	Mask Pwr Hi-Lo Alarm Tx11 - Tx08	Mask Tx Optical Power Hi-Lo Alarm: Writing 10b prevents Int_L on Tx High Optical Power	RW	-	-	C
103 67h	7-0	Mask Pwr Hi-Lo Alarm Tx07 - Tx04	Writing 01b prevents Int_L on Tx Low Optical Power				
104 68h	7-0	Mask Pwr Hi-Lo Alarm Tx03 - Tx00	Writing 11b prevents Int_L on both High and Low Optical Power alarms. Default = 00b; mask is required if corresponding optional alarm is implemented.				

Table 142 CXP Tx Lower Page Memory Map (Sheet 7 of 7)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/- (Not Applicable)		
					Passive	Active Electrl	Active Optical
105 69h	7	Mask Temp High Alarm - Tx	Mask High Internal Temperature Alarm: Writing 1 prevents Int_L on High Tx Internal temperature. Default = 0; mask is required if corresponding optional alarm is implemented.	RW	C	C	R
	6	Mask-Temp Low Alarm - Tx	Mask Low Internal Temperature Alarm: Writing 1 prevents Int_L on Low Tx internal temperature. Default = 0; mask is required if corresponding optional alarm is implemented.				
	5-0	Reserved					
106 6Ah	7	Mask Vcc3.3-Tx High Alarm	Mask High Internal 3.3 Vcc Alarm: Writing 1 prevents Int_L on High Vcc3.3-Tx Voltage alarm. Default = 0; mask is required if corresponding optional alarm is implemented.	RW	C	C	C
	6	Mask Vcc3.3-Tx Low Alarm	Mask Low Internal 3.3 Vcc Alarm: Writing 1 prevents Int_L on Low Vcc3.3-Tx Voltage alarm. Default = 0; mask is required if corresponding optional alarm is implemented.		C	C	C
	5-4	Reserved					
	3	Mask Vcc12-Tx High Alarm	Mask High Internal Vcc12 Alarm: Writing 1 prevents Int_L on High Vcc12-Tx Voltage alarm. Default = 0; mask is required if corresponding optional alarm is implemented.		C	C	C
	2	Mask Vcc12-Tx Low Alarm	Mask Low Internal Vcc12 Alarm: Writing 1 prevents Int_L on Low Vcc12-Tx Voltage alarm. Default = 0; mask is required if corresponding optional alarm is implemented.		C	C	C
	1-0	Reserved					
107	All	Vendor Specific	Vendor specific mask	RW			
108 6Ch	7-4	Reserved	Mask Tx LOL Flag: Writing 1 prevents Int_L on Tx Loss of Lock on Tx CDR. Default = 0, mask is required if corresponding optional alarm is implemented.	RW	-	C	C
	3-0	Mask LOL Flag Tx11 - Tx08					
109	7-0	Mask LOL Flag Tx07 - Tx00					
110-118 6Eh-76h	All	Vendor Specific - 9B	Vendor Specific Read-Write Registers for Tx	RW			
119-122 77h-7Ah	All	Password Change Entry Area	Password Change Entry Area for Tx register space. See SFF-8636 for definition and usage.	RW	O	O	O
123-126 7Bh-7Eh	All	Password Entry Area	Password Entry Area for Tx register space. See SFF-8636 for definition and usage.	RW	O	O	O
127 7Fh	All	Page Select Byte	Selects Upper Page - Required if paging is used on upper page(s). Not required if paging is not used. Writing 00h selects Tx & Rx Upper Page 00h Writing 01h selects Tx Upper Page 01h, etc.	RW	See Description		

8.7.2 CXP MEMORY MAP - Rx LOWER PAGE

[Table 143](#) describes the memory map for the Rx lower page. This page is optional, and may not be implemented on simple modules or devices such as passive cables.

Table 143 CXP Rx Lower Page Memory Map (Optional) (Sheet 1 of 7)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional - (Not Applicable)		
					Passive	Active copper	Active Optical
0 00h	All	Reserved - 1B	Coded 00h (unspecified)	RO			
1 01h	All	Reserved: Extended Status	00h	RO			
2 02h	7-4	Reserved	0000b	RO			
	3	Reserved	0 - used in Tx Lower Page to indicate presence of Rx				
	2	Flat/Paging Memory Presence	0 = Paging is present. 1 = Upper Page 00h only, no other Rx Upper pages		R	R	R
	1	Int_L Status	Coded 1 for asserted Int_L. Clears to 0 when all flags are cleared.		O	R	R
	0	Data_Not_Ready	Indicates transceiver has not yet achieved power up and monitor data is not ready. Bit remains high until data is ready to be read at which time the device sets the bit low. The Data_Not_Ready bit shall be asserted high during free-side device reset, power up reset and prior to a valid suite of monitor readings. Once all monitor readings are valid, the bit is set low until the device is powered down or reset. Upon completion of power up reset, the free-side device shall assert IntL (if supported) low while de-asserting the Data_Not_Ready bit low. The IntL bit will remain asserted until a read is performed of the Data_Not_Ready bit.		O	R	R
3-5	All	Reserved - 3B	Reserved for Status info	RO			

Table 143 CXP Rx Lower Page Memory Map (Optional) (Sheet 2 of 7)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
06h	6	Reserved - 1b	Coded 0b. Reserved for Tx status info	RO			
	6	LOS Rx Status Summary	Coded 1 when a LOS Rx flag (bytes 7-8) is asserted for any channel, else 0. Clears when Fault flags are cleared.		-	O	O
	5	Fault Rx Status Summary	Coded 1 when a Fault Rx flag (bytes 9-10) is asserted for any channel, else 0. Clears when all Fault flags are cleared.				
	4-3	Reserved	Coded 00b. Reserved for Tx status info				
	2	Power Rx Status Summary	Coded 1 when a Rx Optical Power Hi-Lo Alarm (bytes 14-16) is asserted, else 0. Clears when alarm is cleared.		-	-	O
	1	Module Rx Status Summary	Coded 1 when any Rx Temperature or Voltage alarm (bytes 17-18) or reserved module Rx monitor alarm (reserved in byte 19) or Loss of Lock Rx (Bytes 20-21) are asserted, else 0. Clears when all these alarms are cleared.		-	O	O
	0	Reserved	Reserved for other Module Monitor alarm				
07h	7-4	Reserved	Loss of Signal Rx Channel: Coded 1 when asserted, Latched, Clears on Read.	RO	-	O	O
	3-0	L-LOS Rx11 - Rx08					
8	7-0	L-LOS Rx07 - Rx00	Masking bits for these alarms are at Lower Page, Bytes 95-96.				
09h	7-4	Reserved	Fault Rx Channel: Coded 1 when asserted, Latched, Clears on Read.	RO	-	O	O
	3-0	L-Fault Rx11 - Rx08					
10	7-0	L-Fault Rx07 - Rx00	Masking bits for these alarms are at Lower Page, Bytes 97-98.				
11-13	All	Reserved - 3B	Reserved - Module Alarms - used in Tx Lower Page for Optical Bias Current Hi-Lo alarms	RO			
14 0Eh	7-0	L-Power Hi-Lo Alarm Rx11 - Rx08	Rx Optical Power Hi-Lo Alarm Latched: Coded 10 when asserted for High Rx Optical power alarm, Coded 01 when asserted for Low Rx optical power alarm. Latched, Clears on Read. Thresholds for these alarms are at Upper Page 01, Bytes 176-179. Masking bits for these alarms are at Rx Lower Page, Bytes 102-104.	RO	-	-	O
15	7-0	L-Power Hi-Lo Alarm Rx07 - Rx04					
16 10h	7-0	L-Power Hi-Lo Alarm Rx03 - Rx00					

Table 143 CXP Rx Lower Page Memory Map (Optional) (Sheet 3 of 7)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
17 11h	7	L-Temp High Alarm - Rx	High Internal Temperature Alarm Latched: Coded 1 when asserted, Latched, Clears on Read. Thresholds for all temperature alarms are at Upper Page 01, Bytes 1176-179. Masking bits for all Temperature alarms are at Lower Page, Bytes 128-129.	RO	-	-	O
	6	L-Temp Low Alarm - Rx	Low Internal Temperature Alarm Latched: Coded 1 when asserted, Latched, Clears on Read.		-	-	O
	5-0	Reserved					
18 12h	7	L-Vcc3.3 High Alarm - Rx	High Internal Vcc3.3 Alarm Latched: Coded 1 when asserted, Latched, Clears on Read. Thresholds for all voltage alarms are at Upper Page 01, Bytes 144-151. Masking bits for all voltage alarms are at Lower Page, Byte 105.	RO	O	O	O
	6	L-Vcc3.3 Low Alarm - Rx	Low Internal Vcc3.3 Alarm Latched: Coded 1 when asserted, Latched, Clears on Read.				
	5-4	Reserved					
	3	L-Vcc12 High Alarm - Rx	High Internal Vcc12 Alarm Latched: Coded 1 when asserted, Latched, Clears on Read.		O	O	O
	2	L-Vcc12 Low Alarm - Rx	Low Internal Vcc12 Alarm Latched: Coded 1 when asserted, Latched, Clears on Read.				
	1-0	Reserved					
	All	Reserved	Reserved - Rx module alarm		RO		
20 14h	7-4	Reserved	Loss of Lock, Rx CDR: Coded 1 when asserted (i.e., when CDR is enabled and not locked to data stream), 0 when (CDR is enabled AND locked) OR (CDR is bypassed). Latched.	RO	-	O	O
	3-0	LOL Rx11 - Rx08					
21	7-0	LOL Rx07 - Rx00					

Table 143 CXP Rx Lower Page Memory Map (Optional) (Sheet 4 of 7)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
22 16h	All	1st Rx Temp Monitor MSB	1st Internal Temperature Monitor for Rx MSB: Integer part coded in signed 2's complement. Tolerance is $\pm 3^{\circ}\text{C}$. If implemented, sensor location is vendor-defined - intended to be most temperature-sensitive Rx location.	RO	O	O	O
23 17h	All	1st Rx Temp Monitor LSB	1st Internal Temperature Monitor for Rx LSB: Fractional part in units of $1^{\circ}/256$ coded in binary.				
24-25	All	2nd Rx Temp Monitor	2nd Internal Temperature Monitor for Rx. Same 2 Byte format as 1st. If implemented, sensor location is vendor-defined.	RO	O	O	O
26-27 1A-1Bh	All	Rx Vcc3.3 Monitor MSB Rx Vcc3.3 Monitor LSB	Internal Vcc3.3 Monitor for Rx: Voltage in $100 \mu\text{V}$ units coded as 16 bit unsigned integer, Low byte is MSB. Tolerance is $\pm 0.10\text{V}$.	RO	O	O	O
28-29 1C-1Dh	All	Rx Vcc12 Monitor MSB Rx Vcc12 Monitor LSB	Internal Vcc12 Monitor for Rx: Voltage in $250 \mu\text{V}$ units coded as 16 bit unsigned integer, Low byte is MSB. Tolerance is $\pm 0.1\text{V}$.	RO	O	O	O
30-37	All	Reserved - 8B	Reserved - Module Monitors	RO			
38-39 26h-27h	All	Elapsed Operating Time	Elapsed (Power-on) Operating Time: Elapsed time in 2 hour units coded as 16 bit unsigned integer, Low byte is MSB, Tolerance is $\pm 10\%$	RO	O	O	O
40 28h	All	Rx module application select	Format to be determined as other applications besides InfiniBand arise	RW	-	O	O
41 29h	7-5	Reserved	Reserved - Rate Select	RW			
	4-0	Rx Rate Select	Rx Rate Select / optimization bit-map Bit 4: EDR Bit 3:FDR Bit 2: QDR Bit 1: DDR Bit 0:SDR Examples: 00111: Configured for QDR / DDR / SDR operation 01000: Configured for FDR operation 10000: Configured for EDR operation 00000: no info provided	RW	-	O	O
42 2Ah	All	Reserved	Used in Tx Lower Page to manage devices with >6.0 Watt power utilization	RW			
43 2Bh	7-1	Reserved		RW	-	O	O
	0	RX CDR Bypassed/Enabled	0: RX CDR is Bypassed. Default condition at power-on for InfiniBand modules and active cables, for SDR-FDR host compatibility 1: RX CDR is Enabled.				
44-50	All	Reserved - 8B	Reserved - Module Control	RW			
51 33h	7-1	Reserved		RW			
	0	Reset - Rx	Reset: Writing 1 return all registers on Rx pages (except any vendor-specific non-volatile RW areas) to factory default values. Reads 0 after operation.		R	R	R

Table 143 CXP Rx Lower Page Memory Map (Optional) (Sheet 5 of 7)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
52 34h	7-4	Reserved	Rx Channel Disable: Writing 1 disables the whole channel. For optical module input power reporting, see Rx Upper Page 01h, Bytes 206-229. Default is 0.	RW	-	O	O
	3-0	Channel Disable Rx11 - Rx08					
53	7-0	Channel Disable Rx07 - Rx00					
54 36h	7-4	Reserved	Rx Output Disable: Writing 1 disables only the electrical output for the channel. Default is 0 (Output enabled).	RW	-	O	O
	3-0	Output Disable Rx11 - Rx08					
55	7-0	Output Disable Rx07 - Rx00					
56 38h	7-4	Reserved	Rx Squelch Disable: Writing 1 disables squelch for the channel. Default is 0 (Squelch enabled).	RW	-	O	O
	3-0	Squelch Disable Rx11 - Rx08					
57	7-0	Squelch Disable Rx07 - Rx00					
58 3Ah	7-4	Reserved	Rx Channel polarity Flip: Writing 1 inverts the polarity of outputs relative to inputs. Default is 0 (No polarity flip)	RW	-	O	O
	3-0	Polarity flip Rx11 - Rx08					
59	7-0	Polarity flip Rx07 - Rx00					
60 3Ch	7-4	Reserved	Rx Channel Margin Activation: Writing 1 places Rx in "Margin Mode", reducing receiver sensitivity by a vendor-specific amount (nominally equivalent to roughly 1 dB). Intended use is for testing of link signal integrity margin. May be better implemented on Tx side - both Tx and Rx are optional. Specific implementation is vendor-specific, and may be different for modules vs. active cables (AOCs or AECs). Default is 0.	RW	-	O	O
	3-0	Margin Select Rx11 - Rx08					
61	7-0	Margin Select Rx07 - Rx00					
62 3Eh	7-4	Output Amplitude Rx11	Rx Output Amplitude Control: Four bit code blocks (bits 7-4 or 3-0) are assigned to each channel.	RW	-	O	O
	3-0	Output Amplitude Rx10					
63 3Fh	7-4	Output Amplitude Rx09	Codes 1xxxb are reserved.				
	3-0	Output Amplitude Rx08					
64 40h	7-4	Output Amplitude Rx07	Writing 0111b calls for full-scale signal amplitude.				
	3-0	Output Amplitude Rx06	Writing 0000b calls for minimum signal amplitude.				
65 41h	7-4	Output Amplitude Rx05	Writing intermediate code values calls for intermediate levels of signal amplitude.				
	3-0	Output Amplitude Rx04					
66 42h	7-4	Output Amplitude Rx03					
	3-0	Output Amplitude Rx02					
67 43h	7-4	Output Amplitude Rx01					
	3-0	Output Amplitude Rx00					

Table 143 CXP Rx Lower Page Memory Map (Optional) (Sheet 6 of 7)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
68 44h	7-4	Output De-emphasis Rx11	Rx Output De-emphasis Control: Four bit code blocks (bits 7-4 or 3-0) are assigned to each channel.	RW	-	O	O
	3-0	Output De-emphasis Rx10	Codes 1xxxb are reserved.				
69 45h	7-4	Output De-emphasis Rx09	Writing 0111b calls for full-scale De-emphasis.				
	3-0	Output De-emphasis Rx08	Writing 0000b calls for minimum De-emphasis.				
70 46h	7-4	Output De-emphasis Rx07	Writing intermediate code values calls for intermediate levels of De-emphasis.				
	3-0	Output De-emphasis Rx06					
71 47h	7-4	Output De-emphasis Rx05					
	3-0	Output De-emphasis Rx04					
72 48h	7-4	Output De-emphasis Rx03					
	3-0	Output De-emphasis Rx02					
73 49h	7-4	Output De-emphasis Rx01					
	3-0	Output De-emphasis Rx00					
74-94	All	Reserved - 21B	Reserved - Per-Channel Control	RW			
95 5Fh	7-4	Reserved	Mask Rx LOS Alarm: Writing 1 prevents Int_L on Loss of Signal, Default = 0; mask is required if corresponding optional alarm is implemented.	RW	-	C	C
	3-0	Mask LOS Rx11 - Rx08					
96	7-0	Mask LOS Rx07 - Rx00					
97 61h	7-4	Reserved	Mask Rx Fault Flag: Writing 1 prevents Int_L on Rx Fault. Default = 0; mask is required if corresponding optional alarm is implemented.	RW	-	C	C
	3-0	Mask Rx Fault Flag Rx11 - Rx08					
98	7-0	Mask Rx Fault Flag Rx07 - Rx00					
99-101	All	Reserved - 3B	Reserved - Per Channel Mask	RW			
102 66h	7-0	Mask Pwr Hi-Lo Alarm Rx11 - Rx08	Mask Rx Optical Power Hi-Lo Alarm Writing 10b prevents Int_L on High Rx Optical Power Writing 01b prevents Int_L on Low Rx optical Power Writing 11b prevents Int_L for both High and Low Rx Optical Power alarms. Default = 00b; mask is required if corresponding optional alarm is implemented.	RW	-	-	C
103 67h	7-0	Mask Pwr Hi-Lo Alarm Rx07 - Rx04					
104 68h	7-0	Mask Pwr Hi-Lo Alarm Rx03 - Rx00					

Table 143 CXP Rx Lower Page Memory Map (Optional) (Sheet 7 of 7)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
105 69h	7	Mask Temp High Alarm	Mask High Internal Temperature Alarm: Writing 1 prevents Int_L on High Module Temperature alarm. Default = 0; mask is required if corresponding optional alarm is implemented.	RW	C	C	R
	6	Mask-Temp Low Alarm	Mask Low Internal Temperature Alarm: Writing 1 prevents Int_L on Low Module Temperature alarm. Default = 0; mask is required if corresponding optional alarm is implemented.				
	5-0	Reserved					
106 6Ah	7	Mask Vcc3.3-Rx High Alarm	Mask High Internal 3.3 Vcc Alarm: Writing 1 prevents Int_L on High Vcc3.3-Rx alarm. Default = 0; mask is required if corresponding optional alarm is implemented.	RW	C	C	C
	6	Mask Vcc3.3-Rx Low Alarm	Mask Low Internal 3.3 Vcc Alarm: Writing 1 prevents Int_L on Low Vcc3.3-Rx alarm. Default = 0; mask is required if corresponding optional alarm is implemented.		C	C	C
	5-4	Reserved					
	3	Mask Vcc12-Rx High Alarm	Mask High Internal Vcc12 Alarm: Writing 1 prevents Int_L on High Vcc12-Rx alarm. Default = 0; mask is required if corresponding optional alarm is implemented.		C	C	C
	2	Mask Vcc12-Rx Low Alarm	Mask Low Internal Vcc12 Alarm: Writing 1 prevents Int_L on Low Vcc12-Rx alarm. Default = 0; mask is required if corresponding optional alarm is implemented.		C	C	C
	1-0	Reserved					
107		Vendor Specific	Vendor specific mask	RW			
108 6Ch	7-4	Reserved	Mask Rx LOL Flag: Writing 1 prevents Int_L on Rx Loss of Lock on the Rx CDR. Default = 0, mask is required if corresponding optional alarm is implemented.	RW	-	C	C
	3-0	Mask LOL Flag Rx11 - Rx08					
6D	7-0	Mask LOL Flag Rx07 - Rx00					
110-118 6Eh-76h	All	Vendor Specific - 9B	Vendor Specific Read-Write Registers for Rx	RW			
119-122 77h-7Ah	All	Password Change Entry Area	Password Change Entry Area for Rx register space. See SFF-8636 for definition and usage.	RW			
123-126 7Bh-7Eh	All	Password Entry Area	Password Entry Area for Rx register space. See SFF-8636 for definition and usage.	RW			
127 7Fh	All	Page Select Byte	Selects Upper Page - Required if paging is used on upper page(s). Not required if paging is not used. Writing 00h selects Tx & Rx Upper Page 00h Writing 01h selects Rx Upper Page 01h, etc.	RW	See Description		

8.7.3 CXP MEMORY MAP - Tx & Rx COMMON UPPER PAGE 00H

[Table 144](#) shows the memory map for the first upper page, page 00h, for both Tx and Rx addresses.

Table 144 CXP Tx & Rx Upper Page 00h Memory Map (Sheet 1 of 8)

Byte	Bit	Name	Description	Type	Required/Optional Conditional / - (Not Applicable)		
					Passive	Active copper	Active Optical
128 80h	All	Reserved - Type Identifier	SFF-style Type Identifier code. 2 options defined: 0Eh - CXP or later 12h - CXP28 (aka. CXP2) or later	RO	R	R	R
129 81h	7-5	Power Class	000: 0.25W max - Class 0 001: 1.0W max - Class 1 010: 1.5W max - Class 2 011: 2.5W max - Class 3 100: 4.0W max - Class 4 101: 6.0W max - Class 5 110: >6.0W - Class 6 111: Reserved	RO	R	R	R
	4	Tx CDR Presence	Coded 1 for Tx CDR (clock & data recovery) provided; else coded 0		R	R	R
	3	Rx CDR Presence	Coded 1 for Rx CDR provided; else coded 0		R	R	R
	2-0	Reserved					
130 82h	All	Connector / Cable	00h-0Ch: Not compatible w/CXP, Rsvd.-compatibility 0Dh-1Fh: Reserved 20h-23h: Rsvd.-compatibility 24h-2Fh: Reserved 30h: Passive Copper Cable Assembly 31h: Active Copper Cable Assembly (ref. Byte 147) 32h: Active Optical Cable Assembly 33h: Optical Transceiver w/ optical connector 34h-7Fh: Reserved (ref. Byte 147) 80h-FFh: Vendor Specific	RO	-	R	R
131 83h	7	1	3.3V - Vcc3.3 - Coded 1 if required for the module		R	R	R
	6-4	000b - Reserved	2.5V, 1.8V, Vo supplies - not available in receptacle				
	3	1	12V - Vcc12 - Coded 1 if required for the module		R	R	R
	2-0	000b - Reserved					
132 84h	All	Max Temperature	Maximum Recommended Operating Case Temperature for the module, in Degrees C	RO	R	R	R
133 85h	All	Min. per-channel bit rate	Min. signal rate = binary value x 100 Mb/s (e.g., 25 (00011001b) = 2500 Mb/s, & 100 (01100100b) = 10,000 Mb/s)	RO	-	R	R
134 86h	All	Max. per-channel bit rate	Max. signal rate = binary value x 100 Mb/s	RO	R	R	R

Table 144 CXP Tx & Rx Upper Page 00h Memory Map (Sheet 2 of 8)

Byte	Bit	Name	Description	Type	Required/Optional Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
135-136 87h-88h	All	Optical: Laser Wavelength Copper: Attenuation	<p>Optical transmitter: Nominal Laser Wavelength. Wavelength in nm = value / 20: e.g., 42h 04h = 16,900, 16,900/20 = 845 nm</p> <p>Limiting Active Cable (Optical or Full-Active Copper): Vendor-dependent - value may be zero or non-zero.</p> <p>Copper passive: Nominal attenuation of cable from input to output, not including test board</p> <p>Byte 135: Attenuation at 2.5 GHz in dB</p> <p>Byte 136: Attenuation at 5 GHz in dB</p> <p>Copper Linear Active: Nominal difference in attenuation of cable from input to output</p> <p>Byte 135: Nominal difference in attenuation between 200 MHz and 2.5 GHz in dB</p> <p>Byte 136: Nominal difference in attenuation between 200 MHz and 5 GHz in dB</p> <p>Half-active limiting copper cables (Near- or far-end limiting): Nominal host transmitter equalization at cable input requested by cable</p> <p>Byte 135: gain difference requested between 200 MHz and 2.5 GHz in dB. 00h=no info (no host equalization requested)</p> <p>Byte 136: gain difference requested between 200 MHz and 5 GHz in dB. 00h=no info</p>	RO	R	R	R

Table 144 CXP Tx & Rx Upper Page 00h Memory Map (Sheet 3 of 8)

Byte	Bit	Name	Description	Type	Required/Optional Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
137-138 89-8Ah	All	Optical: Max Wavelength Deviation Tolerance Copper: Attenuation Extended	Optical transmitter: Wavelength tolerance (max deviation from nominal) Wavelength tolerance in nm = +/- value / 200): e.g., 0Bh BBh = 3003, 3003/200= 15 nm Limiting Active Cable (Optical or Full-Active Copper): Vendor-dependent - value may be zero or non-zero. Copper Passive: Nominal attenuation of cable from input to the output extended Byte 137: Attenuation at 7.0 GHz in dB Byte 138: Attenuation at 12.9 GHz in dB Copper Linear Active: Nominal difference in attenuation of cable from input to the output Byte 137: Nominal difference in attenuation between 200 MHz and 7.0 GHz in dB Byte 138 Nominal difference in attenuation between 200 MHz and 12.9 GHz in dB Half-active limiting copper cables (Near- or far-end limiting): Nominal host transmitter equalization at cable input requested by cable Byte 137: gain difference requested between 200 MHz and 7.0 GHz in dB. 00h=no info (no host equalization requested). Byte 138 gain difference requested between 200 MHz and 12.9 GHz in dB. 00h=no info	RO	R	R	R
139 8Bh	7	Support for Tx Fault	Coded 1 if Tx Fault Flag supported, else coded 0		R	R	R
	6	Support for Rx Fault	Coded 1 if Rx Fault Flag supported, else coded 0		R	R	R
	5	Support for Tx LOS	Coded 1 if Tx Loss of Signal Flag supported, else coded 0		R	R	R
	4	Support for Rx LOS	Coded 1 if Rx Loss of Signal Flag supported, else coded 0		R	R	R
	3	Support for Tx Squelch	Coded 1 if Tx Squelch supported, else 0		R	R	R
	2	Support for Rx Squelch	Coded 1 if Rx Squelch supported, else 0		R	R	R
	1	Support for Tx CDR LOL	Coded 1 if Tx CDR Loss of Lock Flag supported, else coded 0		R	R	R
	0	Support for Rx CDR LOL	Coded 1 if Rx CDR Loss of Lock Flag supported, else coded 0		R	R	R

Table 144 CXP Tx & Rx Upper Page 00h Memory Map (Sheet 4 of 8)

Byte	Bit	Name	Description	Type	Required/Optional Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
140 8Ch	7	Support for Tx Bias Monitor	Coded 1 if Tx Bias Monitor supported, else coded 0	RO	R	R	R
	6	Support for Tx LOP Monitor	Coded 1 if Tx Light Output Power Monitor supported, else coded 0		R	R	R
	5	Support for Rx Input Power Monitor	Coded 1 if individual Rx Input Power Monitors supported, coded 0 for single-channel or group monitor		R	R	R
	4	Support for Rx Input Power Format	Coded 1 if Rx Input Power reported as Pave, coded 0 for reported as OMA		R	R	R
	3	Support for Case Temp Monitor	Coded 1 if Case Temperature Monitor supported, else coded 0		R	R	R
	2	Support for Internal Temp Monitor	Coded 1 if Internal Temperature Monitor supported, else coded 0		R	R	R
	1	Support for Peak Temp Monitor	Coded 1 if Peak Temperature Monitor supported, else coded 0		R	R	R
	0	Support for Elapsed Time Monitor	Coded 1 if Elapsed PowerOn Operating Time Monitor supported, else coded 0		R	R	R
141 8Dh	7	BER Monitor	Coded 1 for BER Monitor, else coded 0	RO	R	R	R
	6	Vcc3.3-Tx Monitor	Coded 1 for Internal Vcc3.3-Tx Monitor, else coded 0		R	R	R
	5	Vcc3.3-Rx Monitor	Coded 1 for Internal Vcc3.3-Rx Monitor, else coded 0		R	R	R
	4	Vcc12-Tx Monitor	Coded 1 for Internal Vcc12-Tx Monitor, else coded 0		R	R	R
	3	Vcc12-Rx Monitor	Coded 1 for Internal Vcc12-Rx Monitor, else coded 0		R	R	R
	2	TEC Current Monitor	Coded 1 for TEC current Monitor, else coded 0		R	R	R
	1-0	Reserved					

Table 144 CXP Tx & Rx Upper Page 00h Memory Map (Sheet 5 of 8)

Byte	Bit	Name	Description	Type	Required/Optional Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
142 8Eh	7-6	Tx Channel Disable Capabilities	00: Not provided, or unspecified 01: Global Tx Channel Disable Control implemented 10: Individual & independent Tx Channel Disable Control implemented 11: Reserved	RO	R	R	R
	5-4	Tx Channel Output Disable Capabilities	00: Not provided, or unspecified 01: Tx Global Channel Output Disable Control implemented 10: Individual & independent Tx Channel Output Disable Control implemented 11: Reserved		R	R	R
	3-2	Tx Squelch Disable Capabilities	00: Not provided, or unspecified 01: Global Tx Squelch Disable Control implemented 10: Individual and independent Tx Channel Disable Control implemented 11: Reserved		R	R	R
	1	Tx Polarity Flip Mode	Coded 1 for Tx Channel Polarity Flip Control provided, else coded 0		R	R	R
	0	Tx Margin Mode	Coded 1 for Tx Margin Mode provided, else coded 0		R	R	R
143 8Fh	7-4	Reserved		RO			
	3-2	Tx Input Equalization Control	00: Not provided, or unspecified 01: Global Tx Input Equalization Control implemented 10: Individual and independent Tx Input Equalization Control implemented 11: Reserved		R	R	R
	1-0	Tx Rate Select Control	00: Not provided, or unspecified 01: Global Tx Rate/Application Select Control implemented 10: Reserved (Individual and independent Tx Rate/Application Select control not available except in vendor-specific manner). 11: Reserved		R	R	R

Table 144 CXP Tx & Rx Upper Page 00h Memory Map (Sheet 6 of 8)

Byte	Bit	Name	Description	Type	Required/Optional Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
144 90h	7-6	Rx Channel Disable Capabilities	00: Not provided, or unspecified 01: Global Rx Channel Disable Control implemented 10: Individual & independent Rx Channel Disable Control implemented 11: Reserved	RO	R	R	R
	5-4	Rx Channel Output Disable Capabilities	00: Not provided, or unspecified 01: Rx Global Channel Output Disable Control implemented 10: Individual & independent Rx Channel Output Disable Control implemented 11: Reserved		R	R	R
	3-2	Rx Squelch Disable Capabilities	00: Not provided, or unspecified 01: Global Rx Squelch Disable Control implemented 10: Individual and independent Rx Channel Disable Control implemented 11: Reserved		R	R	R
	1	Rx Polarity Flip Mode	Coded 1 for Rx Channel Polarity Flip Control provided, else coded 0		R	R	R
	0	Rx Margin mode	Coded 1 for Rx Margin Mode provided, else coded 0		R	R	R
145 91h	7-6	Reserved		RO			
	5-4	Rx Output Amplitude Control	00: Not provided, or unspecified 01: Global Rx Output Amplitude Control implemented 10: Individual and independent Rx Output Amplitude Control implemented 11: Reserved		R	R	R
	3-2	Rx Output De-Emphasis Control	00: Not provided, or unspecified 01: Global Rx Output De-Emphasis Control implemented 10: Individual and independent Rx Output De-Emphasis Control implemented 11: Reserved		R	R	R
	1-0	Rx Rate Select Control	00: Not provided, or unspecified 01: Global Rx Rate/Application Select Control implemented 10: Reserved (Individual and independent Rx Rate/Application Select control not available except in vendor-specific manner). 11: Reserved		R	R	R

Table 144 CXP Tx & Rx Upper Page 00h Memory Map (Sheet 7 of 8)

Byte	Bit	Name	Description	Type	Required/Optional Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
146 92h	7	FEC Control	Coded 1 for FEC Control, else coded 0	RO	R	R	R
	6	PEC Control	Coded 1 for PEC Control, else coded 0		R	R	R
	5	JTAG Control	Coded 1 for JTAG Control, else coded 0		R	R	R
	4	AC-JTag Control	Coded 1 for AC-JTAG Control, else coded 0		R	R	R
	3	BIST	Coded 1 for BIST, else coded 0		R	R	R
	2	TEC Temperature Control	Coded 1 for TEC Temperature Control, else coded 0		R	R	R
	1	Sleep Mode Set Control	Coded 1 for Sleep Mode Set Control provided, else coded 0		R	R	R
	0	CDR Bypass Control	Coded 1 for CDR Bypass Control provided, else coded 0		R	R	R
147 93h	7-4	Device Technology	0000: 850 nm VCSEL 0001: 1310 nm VCSEL 0010: 1550 nm VCSEL 0011: 1310 nm FP 0100: 1310 nm DFB 0101: 1550 nm DFB 0110: 1310 nm EML 0111: 1550 nm EML 1000: Copper or others 1001: 1490 nm DFB 1010: Copper cable unequalized 1011: Copper cable passive equalized 1100: Copper cable near & far end active equalizers 1101: Copper cable, far end active equalizer 1110: Copper cable, near end active equalizer 1111: Reserved	RO	R	R	R
	3	Wavelength Control	0: No control, 1: Active wavelength control		Inapplicable - coded 0		R
	2	Transmitter cooling	0: Uncooled transmitter, 1: Cooled transmitter				R
	1	Optical Detector	0: P-I-N Detector, 1: APD detector				R
	0	Optical Tunability	0: Transmitter not tunable, 1: Transmitter tunable				R
148 94h	All	Max. Power Utilization	Maximum power utilization, in units of 0.1 Watts. Range: 0.1W - 25.5 Watts 00h: No information	RO	R	R	R

Table 144 CXP Tx & Rx Upper Page 00h Memory Map (Sheet 8 of 8)

Byte	Bit	Name	Description	Type	Required/Optional Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
149 95h	7-1	Data rates supported by the cable assembly	Extended module codes for InfiniBand 7 = IEEE 802.3 CPPI supported 6 = Reserved 5 = EDR 4 = FDR 3 = QDR 2 = DDR 1 = SDR Populate applicable bit rates with a 1b. Examples: 0001 111: FDR / QDR / DDR / SDR 0011 111: EDR / FDR / QDR / DDR / SDR FDR and EDR modules that perform the SDR portions of link initialization using FDR/EDR electrical high-speed signal specifications may set the SDR bit to either 0 or 1.	RO	R	R	R
					0	12x to 3-4x	Coded 1 for 12x to 3-4x Cable, else, for regular cable without fanout, coded 0
150-151 96h-97h	All	Cable length	Physical length of cable, in units of 0.5 meters Range: 0.5 - 32767 m 0000h: Optical transceiver with demateable optical connector	RO	R	R	R
152-167 98h-A7h	All	Vendor Name	Vendor name in ASCII - 16B	RO	R	R	R
168-170 A8-AAh	All	Vendor OUI	Vendor OUI (IEEE ID): Organization-Unique Identifier - 3B	RO	R	R	R
171-186 AB-BAh	All	Vendor Part Number	Vendor Part Number in ASCII - 16B	RO	R	R	R
187-188 BB-BCh	All	Vendor Rev. Number	Vendor Revision Number in ASCII - 2B	RO	R	R	R
189-204 BD-CCh	All	Vendor Serial Number	Vendor Serial Number (ASCII): varies by unit - 16B	RO	R	R	R
205-212 CD-D4h	All	Vendor Date Code	Vendor Date Code YYYYMMDD (ASCII): Spaces (20h) for unused characters	RO	R	R	R
213-222 215-DEh	All	Lot Code	Customer-specific code or Vendor-specific lot code (ASCII) - 10B. All spaces (20h) if unused	RO	O	O	O
223 DFh		Checksum	Checksum of addresses 128 through 222 inclusive: 8 low-order bits of sum	RO	R	R	R
224-255 E0-EFh		Vendor Specific	Vendor-specific Read-Only registers - 32B	RO	O	O	O

8.7.4 CXP MEMORY MAP - Tx UPPER PAGE 01H

[Table 145](#) shows the memory map for the Tx upper page 01h.

Table 145 CXP Tx Upper Page 01h Memory Map (Sheet 1 of 2)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
128 80h	All	Hi Alarm Threshold for 1st Tx Temperature Monitor MSB	Hi Alarm Threshold for 1st Internal Temperature Monitor for Tx MSB: Integer part coded in signed 2's complement. Tolerance is $\pm 3^{\circ}\text{C}$.	RO	O	R	R
129 81h	All	Hi Alarm Threshold for 1st Tx Temperature Monitor LSB	Hi Alarm Threshold for 1st Internal Temperature Monitor for Tx LSB: Fractional part in units of $1^{\circ}/256$, in binary.				
130-131 82h-83h	All	Lo Alarm Threshold 1st Tx Monitor Temp	Lo Alarm Threshold for 1st Internal Temperature Monitor for Tx. Same 2 Byte format as 128-129	RO	O	R	R
132-133 84h-85h	All	Hi Alarm Threshold 2nd Tx Monitor Temp	Hi Alarm Threshold for 2nd Internal Temperature Monitor for Tx. Same 2 Byte format	RO	O	O	O
134-135 86h-87h	All	Lo Alarm Threshold 2nd Tx Monitor Temp	Lo Alarm Threshold for 2nd Internal Temperature Monitor for Tx. Same 2 Byte format	RO	O	O	O
136-143 88h-8Fh	All	Reserved - 8B	Reserved - Alarm Thresholds for Module Monitors				
144-145 90h-91h	All	Hi Alarm Threshold Tx Vcc3.3 Monitor	Hi Alarm Threshold for Internal Vcc3.3 Monitor for Tx: Voltage in 100 μV units coded as 16 bit unsigned integer, Low byte is MSB.	RO	-	O	O
146-147 92h-93h	All	Lo Alarm Threshold Tx Vcc3.3 Monitor	Lo Alarm Threshold for Internal Vcc3.3 Monitor for Tx: Voltage in 100 μV units coded as 16 bit unsigned integer, Low byte is MSB.	RO	-	O	O
148-149 94h-95h	All	Hi Alarm Threshold Tx Vcc12-Monitor	Hi Alarm Threshold for Internal Vcc12 Monitor for Tx: Voltage in 100 μV units coded as 16 bit unsigned integer, Low byte is MSB.	RO	-	O	O
150-151 96h-97h	All	Lo Alarm Threshold Tx Vcc12 Monitor	Lo Alarm Threshold for Internal Vcc12 Monitor for Tx: Voltage in 100 μV units coded as 16 bit unsigned integer, Low byte is MSB.	RO	-	O	O
152-167 98h-A7h	All	Reserved - 16B	Reserved - Alarm Thresholds for Module Monitors	RO			
168-169 A8h-A9h	All	Hi Alarm Threshold, Tx Bias Current	High Alarm Threshold on Tx Bias current: in 2 μA units coded as 16 bit unsigned int, Low byte is MSB.	RO	-	-	O
170-171 AA-ABh	All	Lo Alarm Threshold, Tx Bias Current	Low Alarm Threshold on Tx Bias current in 2 μA units coded as 16 bit unsigned integer, Low byte is MSB.	RO	-	-	O
172-173 AC-ADh	All	Hi Alarm Threshold, Tx Optical Power	High Alarm Threshold on Transmitted Optical Power in 0.1 μW units. 16 bit unsigned int, Low byte is MSB.	RO	-	-	O
174-175 AE-AFh	All	Lo Alarm Threshold, Tx Optical Power	Low Alarm Threshold on Transmitter Optical Power in 0.1 μW units. 16 bit unsigned int, Low byte is MSB.	RO	-	-	O
176-179 B0h-B3h	All	Reserved - 4B	Reserved - Alarm Thresholds for Channel Monitors				
180-181 B4h-B5h	All	Checksum	Checksum: Low order 16 bits of the sum of all pairs of bytes from 128 through 179 inclusive, as unsigned integers.	RO	R	R	R

Table 145 CXP Tx Upper Page 01h Memory Map (Sheet 2 of 2)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
182-205 B5h- CDh	All	Bias Current Monitor Tx11 Bias Current Monitor Tx00	Per-channel Tx Bias current: Monitor. 2B per channel, each measured in 2 μA units coded as 16 bit unsigned integer, Low byte within each byte pair is MSB. Tolerance is ± 0.50 mA.	RO	-	-	O
206-229 CEh- E5h	All	Output Optical Power Monitor Tx11... Output Optical Power Monitor Tx00	Per-channel Tx Light Output Monitor in 0.1μW units coded as 16 bit unsigned integer, Low byte within each pair is MSB. Tolerance is +/- 3 dB across vendor-specified range. While channel or channel output is disabled or squelched (Bytes 52-57), or output power is below range of accurate power monitoring, reported value should be 0.1μW (-40 dBm), to distinguish from non-operational channel.	RO	-	-	O
230-255 E6h-FFh	All	Vendor Specific - 26B	Vendor Specific Tx Functions				

8.7.5 CXP MEMORY MAP - Rx UPPER PAGE 01H[Table 146](#) shows the memory map for the Rx upper page 01h.**Table 146 CXP Rx Upper Page 01h Memory Map (Sheet 1 of 2)**

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
128 80h	All	Hi Alarm Threshold for 1st Rx Temperature Monitor MSB	Hi Alarm Threshold for 1st Internal Temperature Monitor for Rx MSB: Integer part coded in signed 2's complement. Tolerance is ± 3°C.	RO	O	O	O
129 81h	All	Hi Alarm Threshold for 1st Rx Temperature Monitor LSB	Hi Alarm Threshold for 1st Internal Temperature Monitor for Rx LSB: Fractional part in units of 1°/256 coded in binary.				
130-131 82h-83h	All	Lo Alarm Threshold 1st Rx Temp Monitor	Lo Alarm Threshold for 1st Internal Temperature Monitor for Rx. Same 2 Byte format as 128-129	RO	O	O	O
132-133 84h-85h	All	Hi Alarm Threshold 2nd Rx Temp Monitor	Hi Alarm Threshold for 2nd Internal Temperature Monitor for Rx. Same 2 Byte format	RO	O	O	O
134-135 86h-87h	All	Lo Alarm Threshold 2nd Rx Temp Monitor	Lo Alarm Threshold for 2nd Internal Temperature Monitor for Rx. Same 2 Byte format	RO	O	O	O
136-143 88h-8Fh	All	Reserved - 8B	Reserved - Alarm Thresholds for Module Monitors				

Table 146 CXP Rx Upper Page 01h Memory Map (Sheet 2 of 2)

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
144-145 90h-91h	All	Hi Alarm Threshold Rx Vcc3.3 Monitor	Hi Alarm Threshold for Internal Vcc3.3 Monitor for Rx: Voltage in 100 µV units coded as 16 bit unsigned integer, Low byte is MSB.	RO	-	O	O
146-147 92h-93h	All	Lo Alarm Threshold Rx Vcc3.3 Monitor	Lo Alarm Threshold for Internal Vcc3.3 Monitor for Rx: Voltage in 100 µV units coded as 16 bit unsigned integer, Low byte is MSB.	RO	-	O	O
148-149 94h-95h	All	Hi Alarm Threshold Rx Vcc12-Monitor	Hi Alarm Threshold for Internal Vcc12 Monitor for Rx: Voltage in 100 µV units coded as 16 bit unsigned integer, Low byte is MSB.	RO	-	O	O
150-151 96h-97h	All	Lo Alarm Threshold Rx Vcc12 Monitor	Lo Alarm Threshold for Internal Vcc12 Monitor for Rx: Voltage in 100 µV units coded as 16 bit unsigned integer, Low byte is MSB.	RO	-	O	O
152-167 98h-A7h	All	Reserved - 16B	Reserved - Alarm Thresholds for Module Monitors	RO			
168-175 A8h-AFh	All	Reserved - 8B	Reserved Alarm Thresholds for Channel Monitors	RO			
176-177 B0h-B1h	All	Hi Alarm Threshold, Rx Optical Power	High Alarm Threshold on Received Optical Power in 0.1 µW units coded as 16 bit unsigned integer, Low byte is MSB.	RO	-	-	O
178-179 B2h-B3h	All	Lo Alarm Threshold, Rx Optical Power	Low Alarm Threshold on Received Optical Power in 0.1 µW units coded as 16 bit unsigned integer, Low byte is MSB.	RO	-	-	O
180-181 B4h-B5h	All	Checksum	Checksum: Low order 16 bits of the sum of all pairs of bytes from 128 through 179 inclusive, as unsigned integers.	RO	R	R	R
182-205 B5h- CDh	All	Reserved - 24B	Reserved Rx Channel Monitors	RO			
206-229 CEh- E5h	All	Input Optical Power Monitor Rx11... Input Optical Power Monitor Rx00	Per-channel Rx Light Input Monitor in 0.1µW units coded as 16 bit unsigned integer, Low byte within each pair is MSB. Tolerance is +/- 3 dB across vendor-specified range. While channel input power is below range of accurate reporting, reported value should be 0.1µW (-40 dBm), to distinguish from non-operational channel.	RO	-	-	O
248-255 F8h-FFh	All	Vendor Specific - 26B	Vendor Specific Rx Functions				

8.7.6 CXP MEMORY MAP - Tx AND/OR Rx UPPER PAGE 02H

Table 147 shows the memory map for the upper page 02h. This page may be addressable through either the Tx (A0h) or Rx (A8h) address range, or through both, as described in Lower Page Byte 2, bits 5-4.

Table 147 CXP Rx Upper Page 02h Memory Map

Byte	Bit	Name	Description	Type	Required/Optional/ Conditional/ - (Not Applicable)		
					Passive	Active copper	Active Optical
128-247 80h-F7h	All	User EEPROM - 120B	User/Host Writable non-volatile EEPROM, Write Cycles: >5000	RW	O	O	O
230-255 F8h-FFh	All	Vendor Specific - 8B	Vendor Specific control functions				

CHAPTER 9: FIBER ATTACHMENT - 2.5 GB/s, 5.0 GB/s, & 10 GB/s

9.1 INTRODUCTION

This chapter describes the InfiniBand Fiber Attachment for link operation at 2.5 Gb/s (SDR), 5.0 Gb/s (DDR), and 10Gb/s (QDR). This provides effective link bandwidths ranging from 5.0 Gb/s to 120 Gb/s. The Fiber Attachment is a media-level, point-to-point, duplex fiber optic interconnect.

A class of very short reach (VSR) fiber optic interconnect options is defined, and is referred to as -SX. A class of longer reach fiber optic interconnect options is defined, and is referred to as -LX. Specifications are provided for each of the interface widths (1x, 4x, 8x, 12x) at one or more distance classes. Certain optical specifications and fiber connector specifications differ between distance classes, and differ between interface width options within a distance class. Except where noted, the specifications contained herein apply to all distance classes for each of the interface width options.

9.2 SCOPE

This chapter defines the following attributes for the 2.5, 5.0 & 10.0 GTransfers/second Fiber Attachment:

- Optical transmission scheme
- Optical Transmitter mask compliance
- Jitter compliance
- Eye safety
- Optical link budget and distance
- Optical Receptacle and Optical Connector
- Optical Cable Plant
- Bulk and Aux power connections.
- Link encoding
- Pluggable Devices

An InfiniBand Optical Transceiver may be permanently attached with other electronic components on a printed circuit board, or may be fabricated as a component that plugs-in through the connector housing. Implementation notes are included at the end of this chapter describing recommended arrangements to interface an InfiniBand Optical Transceiver to other components on an InfiniBand printed circuit board.

9.3 FIBER ATTACHMENT TECHNOLOGY OPTIONS

Fiber Optic Attachment technology options allowed in this version of the specification are listed in [Table 148](#), [Table 149](#), [Table 150](#). Detailed specifications for these options are

provided in the remainder of this chapter.

Table 148 Fiber Optic Attachment Option for SDR (2.5 Gb/s)

	Very short reach (VSR)	Longer reach
1x Wide		
Designation	IB-1x-SX	IB-1x-LX
Wavelength	850 nm	1300 nm
Connector	dual-LC	dual-LC
Worst-case operating range	2 m - 250 m using 50/125 µm 500 MHz.km fiber 2 m - 500 m using 50/125 µm 2000 MHz.km fiber 2 m - 125 m using 62.5/125 µm 200 MHz.km fiber	2 m - 10 km with single mode fiber
4x Wide		
Designation	IB-4x-SX	IB-4x-LX
Wavelength	850 nm	1300 nm
Connector	single MPO	dual-SC
Worst-case operating range	2 m - 125 m using 50/125 µm 500 MHz.km fiber 2 m - 200 m using 50/125 µm 2000 MHz.km fiber 2 m - 75 m using 62.5/125 µm 200 MHz.km fiber	2 m - 10 km with single mode fiber
8x & 12x Wide		
Designation	IB-8x-SX, IB-12x-SX	see Note 1
Wavelength	850 nm	
Connector	dual MPO	
Worst-case operating range	2 m - 125 m using 50/125 µm 500 MHz.km fiber 2 m - 200 m using 50/125 µm 2000 MHz.km fiber 2 m - 75 m using 62.5/125 µm 200 MHz.km fiber	

1: 8x wide LX and 12x wide LX links are not defined in this version of the specification.

Table 149 Fiber Optic Attachment Options for DDR (5.0 Gb/s)

	Very short reach (VSR)	Longer reach
1x Wide		
Designation	IB-1x-DDR-SX	IB-1x-DDR-LX
Wavelength	850 nm	1300 nm
Connector	dual-LC	dual-LC
Worst-case operating range	2 m - 125 m using 50/125 µm 500 MHz.km fiber 2 m - 200 m using 50/125 µm 2000 MHz.km fiber 2 m - 65 m using 62.5/125 µm 200 MHz.km fiber	2 m - 10 km with single mode fiber
4x Wide		
Designation	IB-4x-DDR-SX	IB-DDR-4x-LX
Wavelength	850 nm	see Note 1
Connector	single MPO	
Worst-case operating range	2 m - 75 m using 50/125 µm 500 MHz.km fiber 2 m - 150 m using 50/125 µm 2000 MHz.km fiber 2 m - 50 m using 62.5/125 µm 200 MHz.km fiber	
8x & 12x Wide		
Designation	IB-8x-DDR-SX, IB-12x-DDR-SX	IB-12x-DDR-LX
Wavelength	850 nm	see Note 1
Connector	dual MPO	
Worst-case operating range	2 m - 75 m using 50/125 µm 500 MHz.km fiber 2 m - 150 m using 50/125 µm 2000 MHz.km fiber 2 m - 50 m using 62.5/125 µm 200 MHz.km fiber	

1: 4x, 8x and 12x wide DDR LX links are not defined in this specification.

Table 150 Fiber Optic Attachment Options for QDR (10.0 Gb/s)

	Very short reach (VSR)	Longer reach
1x Wide Designation Wavelength Connector	IB-1x-QDR-SX 850 nm dual-LC	IB-1x-QDR-LX 1300 nm dual-LC
Worst-case operating range	2 m - 82 m using 50/125 µm 500 MHz.km fiber 2 m - 300 m using 50/125 µm 2000 MHz.km fiber 2 m - 33 m using 62.5/125 µm 200 MHz.km fiber	2 m - 10 km with single mode fiber
4x Wide Designation Wavelength Connector	IB-4x-QDR-SX see Note 1	IB-4x-QDR-LX see Note 1
8x & 12x Wide Designation Wavelength Connector	IB-8x-QDR-SX, IB-12x-QDR-SX see Note 1	IB-8x-QDR-LX, IB-12x-QDR-LX see Note 1

1: 4x, 8x, and 12x wide QDR SX and LX links are not defined in this specification.

9.4 FIBER ATTACHMENT OVERVIEW

This section provides an overview of the structure, concepts, and mechanisms of the InfiniBand 1x, 4x, 8x, and 12x fiber optic links operating at SDR, DDR, and QDR speeds.

9.4.1 FIBER OPTIC SYSTEM OVERVIEW

A fiber optic link in general conveys InfiniBand signals between two InfiniBand Boards. A fiber optic link used as an external loopback on a single InfiniBand Board is also possible. A fiber optic link consists of a Fiber Optic Cable connected by means of an Optical Connector at each end to a pair of Optical Transceivers. Each Optical Transceiver typically resides on an InfiniBand Board. The Fiber Optic Cable consists of one or more segments of optical fiber, joined together using Fiber Optic Adapters. The Optical Transceiver consists of the following elements:

- 1) an Optical Transmitter section which converts intermediate electrical signals to InfiniBand-compliant optical signals,

- 2) an Optical Receiver section which converts InfiniBand-compliant optical signals into intermediate electrical signals,
- 3) an Optical Receptacle (which shall be single or dual as specified in subsequent sections) into which the Optical Connector plugs,
- 4) a Signal Conditioner section to convert the vendor-specific intermediate electrical signals into InfiniBand-compliant electrical signals which are specified in [Chapter 6: High Speed Electrical Interfaces](#),
- 5) in the case of 4x LX - a serializer in the transmitter and a de-serializer in the receiver,
- 6) a power management section.

The InfiniBand-compliant optical signal parameters are specified in [Section 9.5](#). Optical Receptacles and Connectors are specified in [Section 9.6](#). Fiber Optic cables are specified in [Section 9.7](#). Signal Conditioners are specified in [Section 9.8](#).

The Optical Transceiver implements power management functionality (AUX power, beaconing and remote wake-up). Aux power is specified in [Section 9.9](#).

Architecture Note

The intermediate electrical interface within the Optical Transceiver is intended to be vendor-specific and is not specified in this document, although recommended implementations are outlined in [Section 9.8](#). The concept of intermediate electrical signals is intended to facilitate the use of existing optoelectronic components in an InfiniBand architecture with the addition of only a new electrical interface chipset, referred to as a "Signal Conditioner". The Signal Conditioner will typically contain retiming and control functionality to convert the intermediate electrical signals to/from IB-compliant electrical signals. The Signal Conditioner may be physically integrated with the optoelectronic components into a solderable or pluggable physical assembly for attachment to an IB board. The intermediate electrical signals are not intended to be connected through the boundaries of the Optical Transceiver.

Implementations are also envisaged wherein an Optical Transceiver converts directly from IB-compliant optical signals to IB-compliant electrical signals.

9.4.2 1x SYSTEM OVERVIEW - SDR, DDR & QDR

C9-1: All 1x Optical Cables shall comply with the requirements in [Section 9.4.2](#) for respective distance classifications (SX or LX).

1x-SX and 1x-LX links differ only in optical specifications ([Section 9.5.8](#)), fiber specification ([Section 9.7.1](#)) and color coding ([Section 9.6.1.4](#)). This overview section therefore applies to both 1x-SX and 1x-LX links, except where stated.

A 1x optical link carries a duplex 1x link as shown in [Figure 171](#). IB optical signals are generated in the 1x-SX Optical Transmitter using one laser and are detected in the 1x-SX Optical Receiver using one photodetector.

Each segment of the 1x Fiber Optic Cable shall have one fiber for each direction. The 1x-SX fiber shall be multimode, and the 1x-LX fiber shall be single-mode. The 1x Optical Connector shall be a dual LC-type connector.

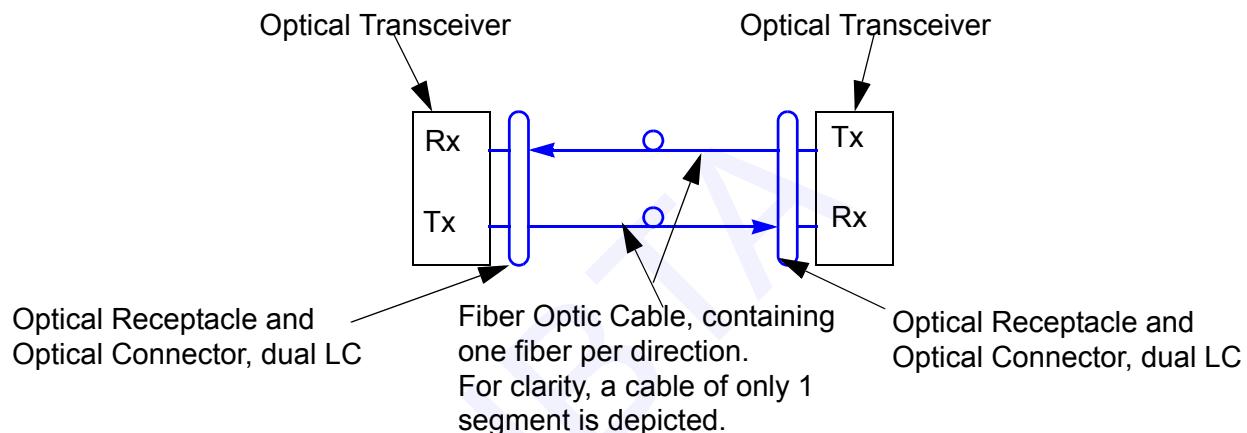


Figure 171 1x Optical Link Overview

9.4.3 4x SYSTEM OVERVIEW

C9-2: This compliance statement is obsolete and has been replaced by [C9-2.1.1](#) and [C9-2.1.2](#):

9.4.3.1 4x-SX OVERVIEW - SDR, DDR, & QDR

C9-2.1.1: All 4x-SX Fiber Optic Cables shall meet [Section 9.4.3.1](#).

A 4x-SX optical link carries a duplex 4x link as shown in [Figure 172](#). Typically the fibers are in a ribbon format. IB optical signals are typically generated in the 4x-SX Optical Transmitter using four lasers and are typically detected in the 4x-SX Optical Receiver using four photodetectors.

Each segment of the 4x SX Fiber Optic Cable shall have four multimode fibers for each direction. The 4x-SX Optical Connector shall be a single MPO connector. To simplify manufacturing, fibers may be present in the four central positions of the connector, but if present these optional fibers shall not be used for IB signals.

4x optical cable links for SDR, DDR, and QDR differ only in supported data rate.

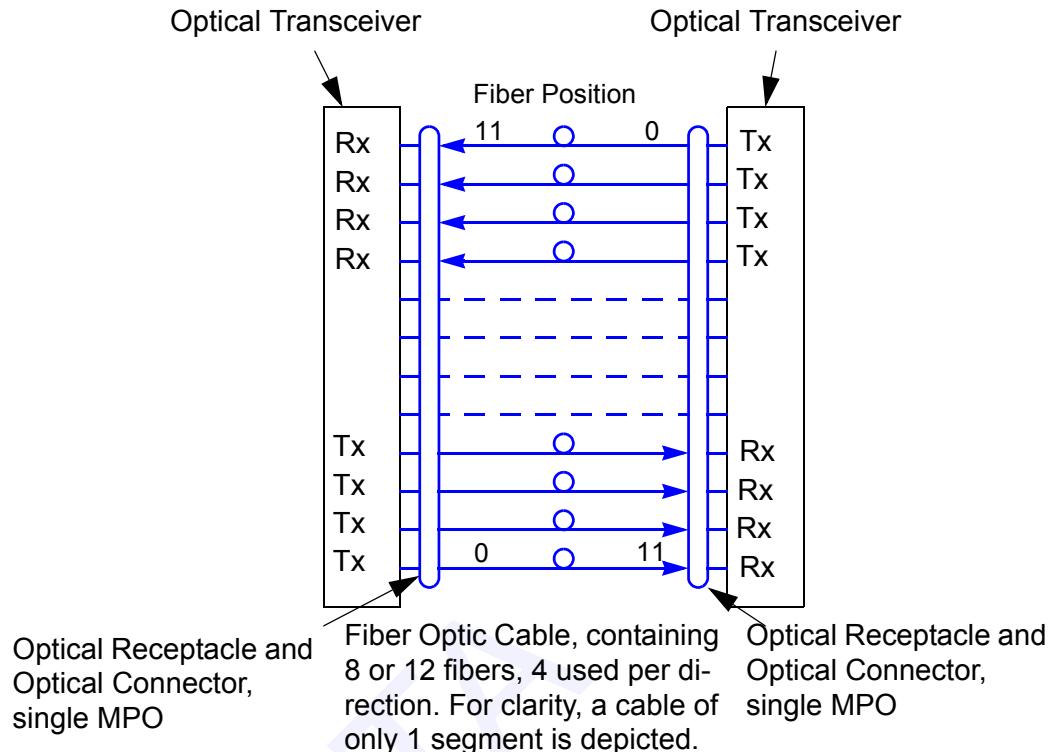


Figure 172 4x-SX Optical Link Overview

9.4.3.2 4x-LX OVERVIEW

Architecture Note

The IB-4x-LX link is somewhat unique, in comparison to the other optical links. The IB-4x-LX operates at a bit rate of 10.0 Gb/s, over a single fiber per direction, and is therefore similar technically similar to an IB-1x-QDR-LX link, but is not identical. It is, rather, a special case of the long-distance SDR 4x link, with unique data serialization, as described in [Section 9.4.3.3 on page 555](#).

C9-2.1.2: All 4x-LX Fiber Optic Cables shall meet Section 9.4.3.2

A 4x-LX optical link carries a duplex 4x link as shown in [Figure 173](#). IB optical signals are generated in the 4x-LX Optical Transmitter using one laser and are detected in the 4x-LX Optical Receiver using one photodetector.

Each segment of the 4x-LX Fiber Optic Cable shall have one fiber for each direction. The 4x-LX fiber shall be single-mode. The 4x LX Optical Connector shall be a dual SC-type connector.

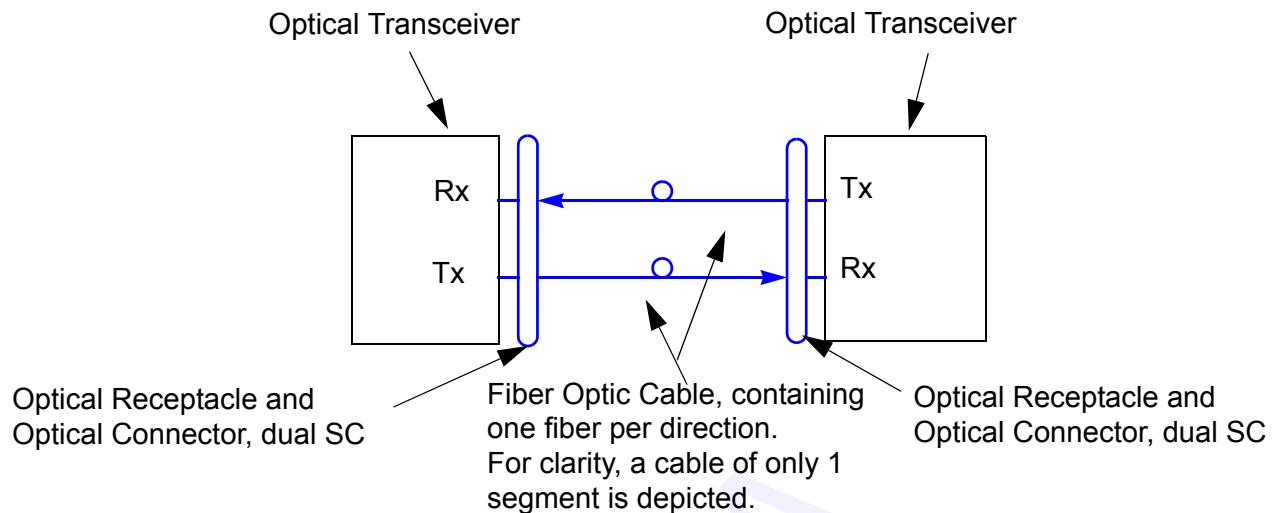


Figure 173 4x-LX Optical Link Overview

9.4.3.3 4x-LX SERIALIZATION

C9-2.1.3: All 4x-LX Optical transceivers shall comply with [9.4.3.3](#)

The 4x-LX optical link is based on the 10Gb/s Ethernet specifications for 10GBASE-LR link. The biggest optical difference is the link speed. 10Gb/s Ethernet serial runs at 10.3125 GBd while the 4x-LX link runs at 10.0 GBd. Both 10GBASE-LR and 4x-LX transmitters accept an 8b/10b encoded byte-striped stream across 4 physical lanes. The 4 lanes for 10Gb/s Ethernet run at 3.125 GBd. To serialize the 4 lanes 10GbE removes the 8b/10b encoding from the bytes on the 4 lanes and then recodes using 64b/66b onto one lane. To keep the optical speed for 4x LX similar to 10GBASE-LR, 4x LX optical transceivers will directly serialize the 4 lanes to achieve 10.0 Gbd.

A 4x-LX Optical Transmitter shall accept an 8b/10b encoded byte striped stream across the 4 physical lanes as described in [Chapter 5: Link/Phy Interface](#). The output of the Optical transmitter shall serialize the traffic in the following manner:

The first byte on the cable shall be byte 0 from lane 0, followed by byte 1 from lane 1, then byte 2 from lane 2, then byte 3 from lane 3, then byte 4 from lane 0,... and so on. Each 10 bit byte shall be serialized in bit order starting with bit 0 and ending with bit 9. Refer to [Figure 174](#) for a diagram detailing the transmitter serialization scheme.

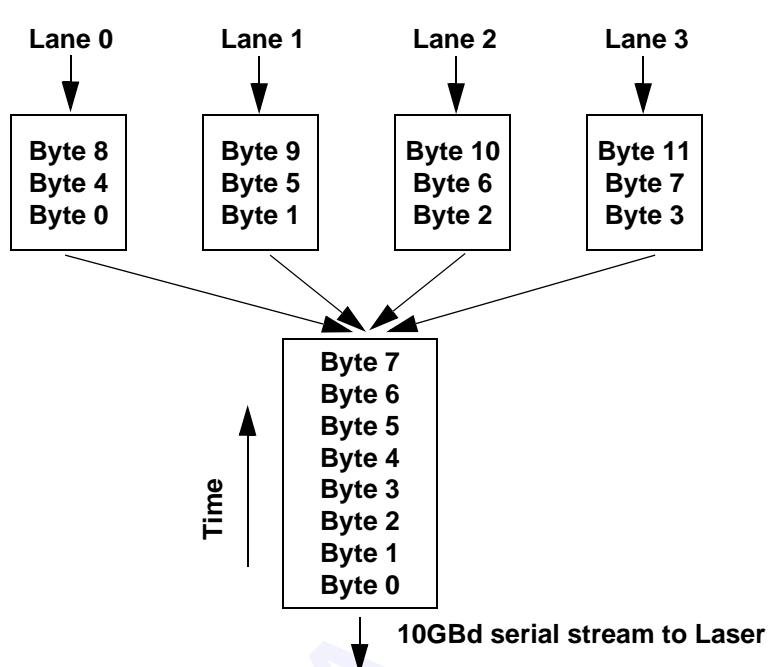


Figure 174 4x LX Transmitter Serialization

A 4x-LX Optical Receiver shall accept a 8b/10b encoded byte serialized stream as described above from a 4x-LX Transmitter. The output of the Optical Receiver shall be a byte-striped stream across 4 lanes as defined in [Chapter 5: Link/Phy Interface](#). Refer to [Figure 175](#) for a diagram detailing de-serialization.

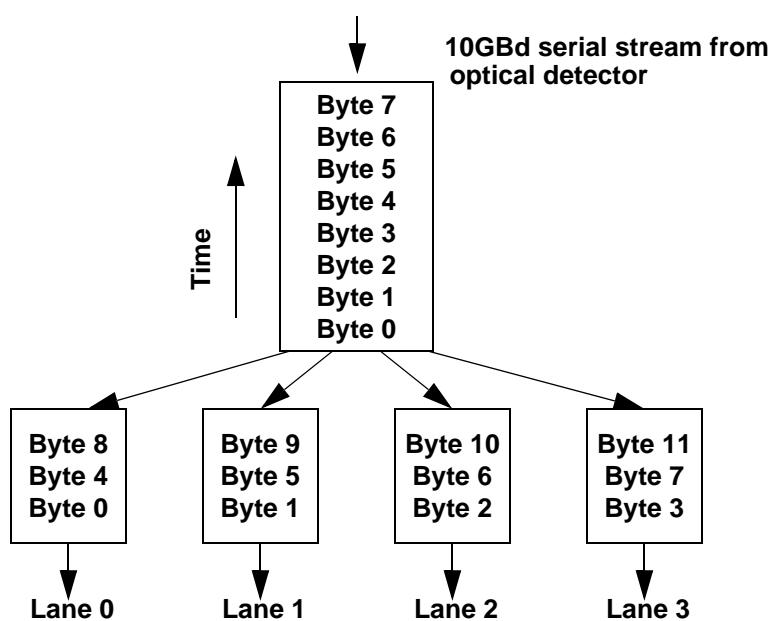


Figure 175 4x LX receiver de-serialization

Architecture Note

The transmitter function to serialize 4 lanes of byte striped traffic is fairly straight forward. The receiver is a little more complicated. One method for correctly de-serializing the symbol-multiplexed stream is to use the comma's to align on bytes. Then use the lane identifier information in TS1s and TS2s to rotate the incoming serial stream to align with the correct physical lane.

9.4.4 8x-SX OVERVIEW - SDR & DDR

C9-2.1.4: All 8x-SX Fiber Optic Cables shall meet [Section 9.4.4](#)

A 8x-SX optical link carries a duplex 8x link as shown in [Figure 176](#). Typically the fibers are in a ribbon format. IB optical signals are typically generated in the 8x-SX Optical Transmitter using eight lasers and are typically detected in the 8x-SX Optical Receiver using eight photodetectors. To simplify manufacturing, fibers may be present in the four extra positions of the connector, but if present these optional fibers shall not be used for IB signals.

Each segment of the 8x SX Fiber Optic Cable shall have eight multimode fibers for each direction. The 8x-SX Optical Connector shall be a dual MPO connector

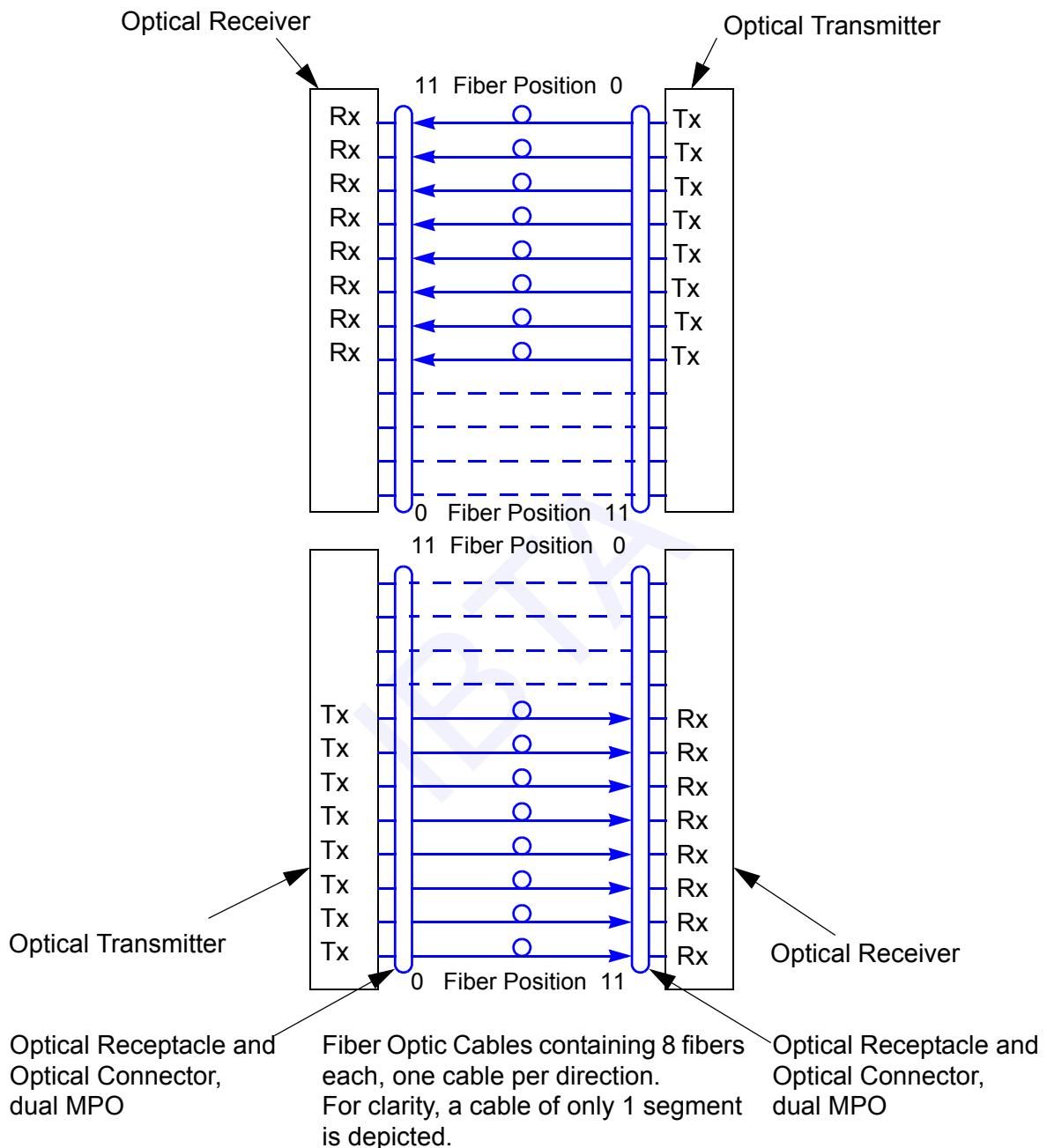


Figure 176 8x-SX Optical Link Overview

9.4.5 12x SYSTEM OVERVIEW - SDR, DDR, & QDR

C9-3: All 12x-SX Fiber Optic Cables shall meet [Section 9.4.5](#).

A 12x-SX optical link carries a duplex 12x link as shown in [Figure 177](#). Typically the fibers are in a ribbon format. IB optical signals are typically generated in the 12x-SX Optical Transmitter using twelve lasers and are typically detected in the 12x-SX Optical Receiver using twelve photodetectors.

Each segment of the 12x SX Fiber Optic Cable shall have twelve multimode fibers for each direction. The 12x-SX Optical Connector shall be a dual MPO connector.

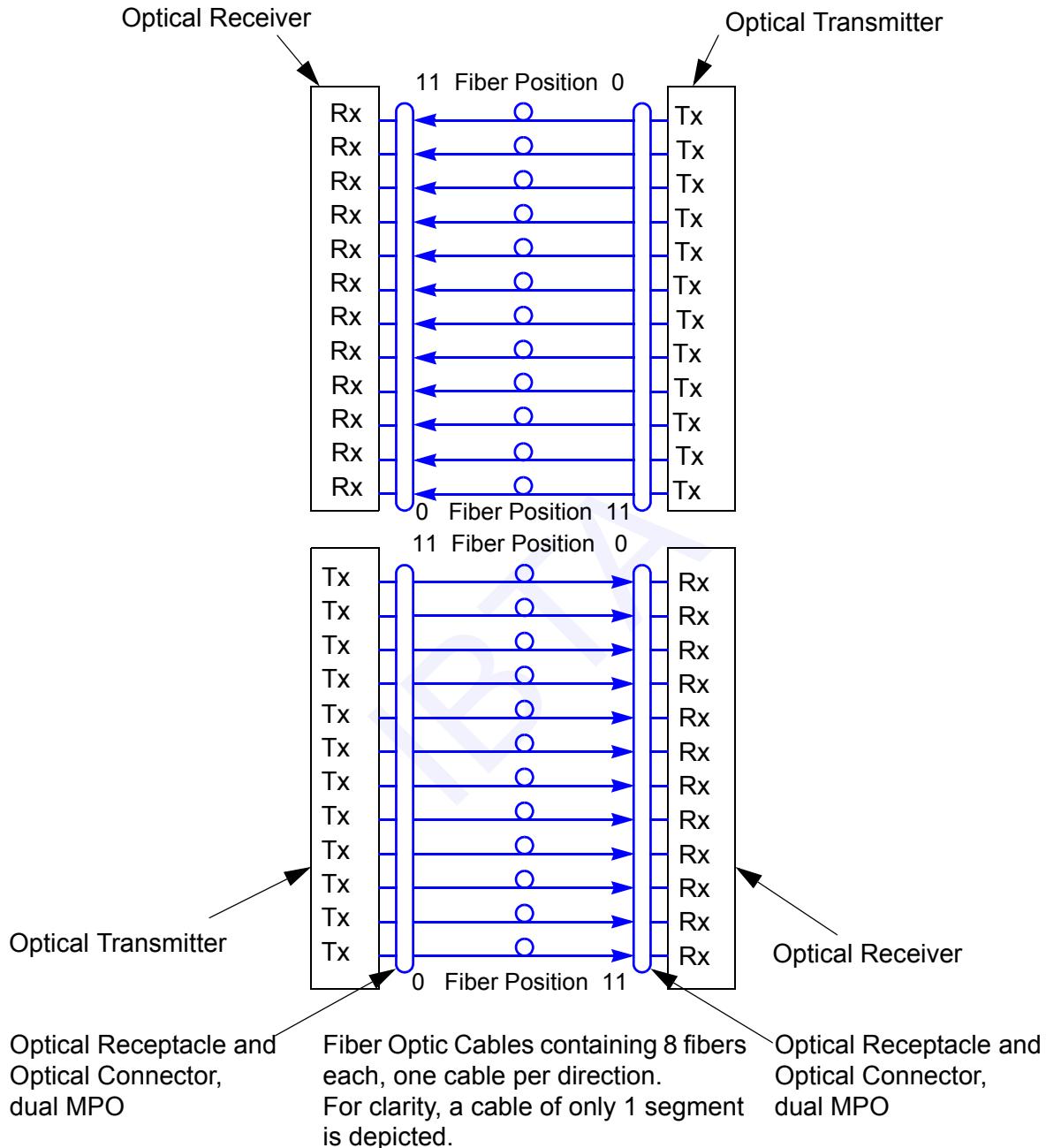


Figure 177 12x-SX Optical Link Overview

9.5 OPTICAL SPECIFICATIONS

This section defines the characteristics of InfiniBand-compliant optical signals. The Optical Transmitter parameters are specified immediately inside the optical fiber adjacent to

the Optical Transmitter (TP2 of [Figure 180](#)). The Optical Receiver parameters are specified immediately inside the optical fiber adjacent to the Optical Receiver (TP3 of [Figure 180](#)). The corresponding fiber optic cable plant specifications are described in [Section 9.7](#).

C9-4: The BER of each lane shall not exceed 10^{-12} under any conditions. The optical specifications in this section support meeting this requirement. In particular the Optical Receiver of each lane is expected to operate at a BER of $\leq 10^{-12}$ over lifetime, temperature, and operating range when driven through a cable plant as specified in [Section 9.7](#) by an Optical Transmitter data stream compliant with the eye mask, jitter and optical parameter specifications.

Architecture Note

The maximum and minimum of the allowed range of average transmitter power coupled into the fiber are worst-case values to ensure that the specified minimum optical modulation amplitude can reliably be launched despite power supply variations, manufacturing variances, drift due to temperature variations, and aging effects.

The minimum received optical modulation amplitude and maximum average received power together define the input power range which is expected to achieve the required BER. The values specified take into account power penalties caused by the use of an Optical Transmitter and Fiber Optic Cable with a worst-case combination of spectral characteristics, optical modulation amplitude, maximum average power, and pulse shape characteristics.

The Gigabit Ethernet optical link model method of IEEE 802.3z was used to estimate the link performance. This model presently uses the parameters extinction ratio and average optical power. Transmitter and Receiver Optical Modulation Amplitudes were calculated from these data. DDR and QDR performance were estimated using 10GBE. This version uses OMA directly.

C9-5: This compliance statement is obsolete and has been replaced by [C9-6.1.1](#): and [C9-6.2.1](#):

C9-6: All Optical Ports shall meet the signal grounding requirements in [Chapter 6: High Speed Electrical Interfaces](#).

9.5.1 QUIESCENT CONDITION

C9-6.1.1: All Optical Ports shall meet the Quiescent conditions and Optical signal polarity specified in [Section 9.5.1](#), and [Section 9.5.2](#)

If an electrical input to an Optical Transmitter is quiescent, then the optical power of that lane shall not be modulated. If there are transitions at the electrical input to the Optical Transmitter, then the optical power shall be modulated according to the pattern of the transitions.

All parameters for each lane shall be met whether the other lanes are individually active or quiescent.

Implementation Note

[Chapter 5: Link/Phy Interface](#) includes a description of a beaconing sequence consisting of bursts of transitions separated by quiescent periods. It is recommended that during the quiescent periods DC current continue to flow through the lasers to bias the lasers at least to threshold and ideally to average optical power level. This will ensure that the optical signals meet specification as quickly as possible when the next beaconing burst starts.

9.5.2 OPTICAL SIGNAL POLARITY

C9-6.1.2: All Optical Ports shall meet the Optical Signal Polarity specified in [Section 9.5.2](#).

A logic Zero high-speed signaling level at the electrical input to an Optical Transmitter shall generate a low level of optical power on the fiber. A logic One high-speed signaling level at the electrical input to an Optical Transmitter shall generate a high level of optical power on the fiber.

9.5.3 OPTICAL TRANSMITTER MASK COMPLIANCE FOR LINKS OPERATING AT 2.5 & 5.0 GB/s

C9-6.2.1: All 1x SX, 1x LX, 4x SX, and 12x SX Optical Ports shall meet the Optical Transmitter Mask specified in [Section 9.5.3](#) while operating at SDR speed.

C9-6.2.2: All 1x SX, 1x LX, 4x SX, and 12x SX Optical ports shall meet the Optical Transmitter Mask specified in [Section 9.5.3](#) while operating at DDR speed. For DDR, the O/E converter 3 dB frequency response shall be scaled up by a factor of 2 to 3.75 GHz.

9.5.3.1 EYE MASK SPECIFICATION

The optical transmitter pulse shape characteristics are specified in the form of a compliance mask on the eye diagram of [Figure 178](#). This transmitter compliance mask is used to verify the overall response of the Optical Transmitter for rise time, fall time, pulse overshoot, pulse undershoot, and ringing. Compliance with this optical mask is a very good indicator that deterministic effects are within generally acceptable limits, but it does not guarantee compliance with IB jitter specifications.

For uniform measurements, the Optical Transmitter eye shall be measured using an O/E converter with a equivalent fourth-order Bessel-Thomson response given by:

$$H_P = \frac{105}{105 + 105y + 45y^2 + 10y^3 + y^4}$$

where

$$y = 2.114p \quad p = \frac{j\omega}{\omega_r} \quad \omega_r = 2\pi f_r \quad f_r = 1.875 \text{ GHz}$$

The O/E converter filter response is based on that described in ITU-T G.957, which provides a physical implementation. The specified O/E converter is only intended to provide uniform measurement and does not represent the noise response of an IB Optical Receiver. An actual SDR (**2.5 Gb/s**) IB Optical Receiver has maximum 3dB bandwidth of 2.8 GHz, as specified in [Table 160](#), [Table 163](#), [Table 172](#) and [Table 178](#). The IB Optical Receiver maximum 3dB bandwidth does not apply to DDR (**5.0 Gb/s**) Receivers.

The reference O/E converter shall have 3 dB frequency response of 1.88 GHz. The equivalent response of the 4th order (or higher) Bessel-Thomson reference O/E converter shall meet the values listed in [Table 151](#) with tolerance **not** exceeding the values listed in [Table 152](#).

Table 151 Equivalent Response of Reference O/E Converter

f/f ₀	f/f _r	Attenuation (dB)	Distortion (UI)
0.15	0.2	0.1	0
0.3	0.4	0.4	0
0.45	0.6	1.0	0
0.6	0.8	1.9	0.002
0.75	1.0	3.0	0.008
0.9	1.2	4.5	0.025
1.0	1.33	5.7	0.044
1.05	1.4	6.4	0.055
1.2	1.6	8.5	0.10
1.35	1.8	10.9	0.14
1.5	2.0	13.4	0.19
2.0	2.67	21.5	0.30

Table 152 Attenuation Tolerance of Reference O/E Converter

Reference Frequency f/f_r	Attenuation Tolerance Δa (dB)
0.1 - 1.00	-0.0... + 0.5
1.00 ... 2.00	+0.5 ... +3.0

NOTE – Intermediate values of Δa shall be linearly interpolated on a logarithmic frequency scale.

The Optical Transmitter compliance mask used for compliance testing shall be as in [Figure 178](#). In this figure the amplitude has been normalized such that an amplitude of 0.0 represents logic ZERO and an amplitude of 1.0 represents logic ONE.

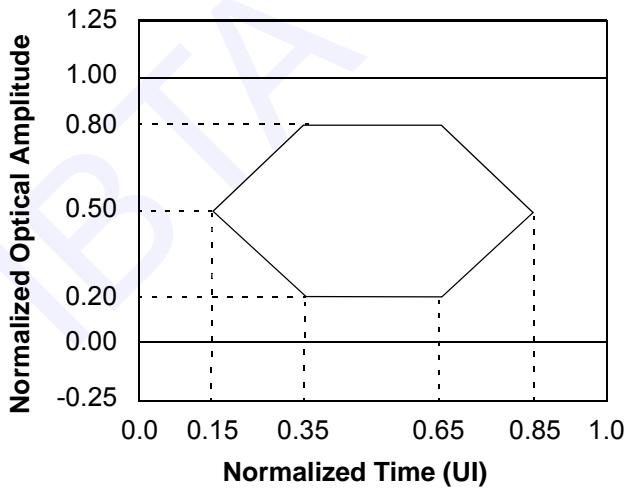


Figure 178 Normalized Optical Transmitter Compliance Mask

9.5.3.2 RISE/FALL TIME MEASUREMENT

Optical rise and fall time specifications are based on unfiltered waveforms. Some lasers have ringing or overshoot, which can reduce the accuracy of 20%-80% rise and fall time measurements. Therefore the 4th-order Bessel-Thomson filter defined in [Section 9.5.3.1](#) is a convenient filter for measurement of the rise and fall time. Since the limited response of the 4th-order Bessel-Thomson filter will adversely impact the measured response, the following equation should be used to remove the filter response from the rise and fall times:

$$T_{rise} = \sqrt{(T_{riseMeasured})^2 - (T_{riseFilter})^2}$$

$$T_{fall} = \sqrt{(T_{fallMeasured})^2 - (T_{fallFilter})^2}$$

The filter 3 dB bandwidth used in the measurement may be different than the specified reference filter, but any filter used in the measurement shall be a fourth order Bessel-Thomson filter.

9.5.3.3 RMS RISE/FALL TIME

Optical rise time and fall time will not be identical in a typical implementation. Optical link models such as the IEEE 802.3z Gigabit Ethernet optical link model are generally defined using the larger of rise time and fall time. This provides an overly pessimistic analysis. Therefore IB optical specifications are defined using T_{rfRMS} , which is the RMS mean of rise time and fall time as defined below:

$$T_{rfRMS} = \sqrt{\frac{(T_{rise})^2 + (T_{fall})^2}{2}}$$

9.5.3.4 OPTICAL MODULATION AMPLITUDE

Optical Modulation Amplitude (OMA) is defined as the absolute difference between the optical power of a logic ONE level and the optical power of a logic ZERO level. OMA is related to Extinction Ratio (ER measured in dB) and Average Optical Power (P_{ave} measured in dBm) by the equation:

$$OMA = 2 \times 10^{P_{ave}/10} \times \frac{(1 - 10^{-ER/10})}{(1 + 10^{-ER/10})}$$

9.5.4 OPTICAL TRANSMITTER MASK COMPLIANCE FOR LINKS OPERATING AT 10.0 GB/S

9.5.4.1 1x QDR OPTICAL TRANSMITTER MASK COMPLIANCE - SX & LX

C9-6.2.3: The 1x SX and 1x LX Optical ports shall meet the Optical Transmitter Mask requirements in [Section 9.5.4.1](#) while operating at QDR speed (**10.0 Gb/s**).

The 1x SX Optical Port shall meet the Optical Transmitter Mask as defined for 10GBASE-SR in IEEE Std 802.3, Clause 52. 1x SX QDR shall use the same transmit and receive optical interface and budget specifications and measurements, with the following notes and exceptions:

- The nominal signaling speed shall be 10.0 GBd ± 100 ppm

The 1x LX QDR Optical Port shall meet the Optical Transmitter Mask as defined for 10GBASE-LR in IEEE Std 802.3, Clause 52. 1x LX QDR shall use the same transmit and receive optical interface and budget specifications and measurements, with the following notes and exceptions:

- The nominal signaling speed shall be 10.0 GBd ± 100 ppm

9.5.4.2 4x LX OPTICAL TRANSMITTER MASK COMPLIANCE

C9-6.1.3: All 4x-LX Optical Ports shall meet the Optical Transmitter Mask specified in [Section 9.5.4.2](#).

Footnote 179. The 4x LX Optical Port shall meet the Optical Transmitter Mask as defined for 10GBASE-LR in IEEE Std 802.3, Clause 52. 4x LX shall use the same transmit and receive optical interface and budget specifications and measurements, with the following notes and exceptions:

- The nominal signaling speed shall be 10.0 GBd ± 100 ppm

9.5.5 OPTICAL JITTER SPECIFICATION FOR LINKS OPERATING AT 2.5 GB/S AND 5.0 GB/S

C9-7: This compliance statement is obsolete and has been replaced by [C9-7.2.1](#):

9.5.5.1 1x SX, 1x LX, 4x SX, AND 12x SX OPTICAL JITTER SPECIFICATIONS FOR 2.5 GB/S AND 5.0 GB/S

C9-7.1.1: This compliance statement is obsolete and has been replaced by [C9-7.2.1](#):

C9-7.1.2: This compliance statement is obsolete and has been replaced by [C9-7.2.4](#):

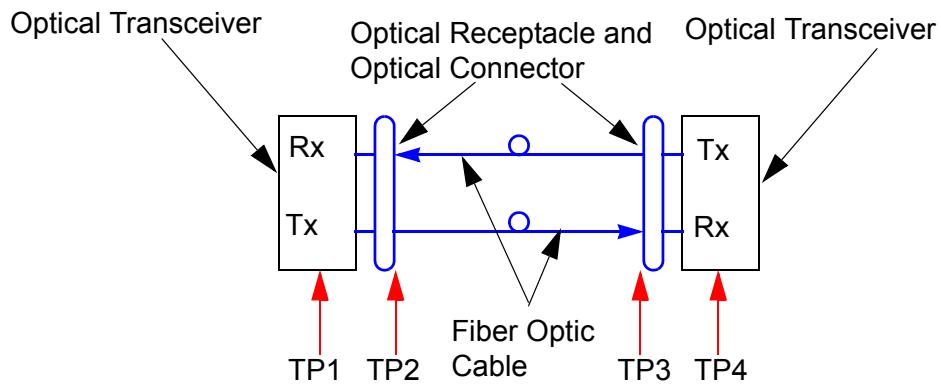
C9-7.2.1: All 1x-SX, 1x-LX, 4x-SX, and 12x-SX Optical Ports shall comply with the Jitter requirements in [Section 9.5.5.1](#) for respective Link widths/distances as defined in [Table 153](#) while operating at SDR speed.

C9-7.2.2: All 1x-SX, 1x-LX, 4x-SX, and 12x-SX Optical Ports shall comply with the Jitter requirements in [Section 9.5.5.1](#) for respective Link widths/distances as defined in [Table 154](#) while operating at DDR speed.

The IB jitter specification is based on the same methodology as the Fibre Channel - Methodologies for Jitter Specification revision 10.0 and the IEEE Std 802.3, Clause 52 Ethernet standard. [Figure 180](#) shows jitter compliance test points TP1, TP2, TP3 and TP4 for an IB-1x optical link.

TP2 is located immediately inside the output end of a test fiber of 2 m length plugged into the Optical Transmitter and wrapped 10 turns around a 25.4 mm-diameter mandrel, and it is the test point at which all Optical Transmitter optical parameters are defined. TP3 is located immediately inside the optical fiber adjacent to the Optical Receiver, and is the location at which all Optical Receiver optical parameters are defined. TP1 is located at the vendor-specific intermediate electrical signals within the Optical Transmitter. TP4 is located at the vendor-specific intermediate electrical signals within the Optical Receiver. The physical existence of TP1 and TP4 is optional.

Test points for 4x-SX, 8x-SX and 12x-SX links are defined similarly to those for 1x.



For clarity, only the jitter compliance test points for the left-to-right portion of the link are shown

Figure 180 Jitter Compliance Test Points

When operating at SDR speed, IB optical links shall not exceed the deterministic and total jitter values listed in [Table 153](#) at TP2 and TP3 when the input jitter to the Optical Transceiver is compliant with the IB electrical specification. Typical jitter values at TP1 and TP4 are listed in [Table 153](#) for reference. The total jitter of an optical component shall be measured at BER of 10^{-12} with a test sequence of K28.5+, K28.5- characters. An IB optical port must provide BER of $\leq 10^{-12}$ under worst case data patterns. Suitable test methods are defined in Fibre Channel - Methodologies for Jitter Specification revision 10.0. The jitter specification shall be met over the entire range of compliant optical parameters and compliant Fiber Optic Cables.

The total jitter listed for TP4 in [Table 153](#) does not include a sinusoidal jitter (SJ) component.

Table 153 Maximum Jitter of Optical Links for SDR

InfiniBand Link	Compliance Point	Deterministic Jitter		Total Jitter	
		UI	ps	UI	ps
1x-SX, 1x-LX	TP1 (input test jitter)	0.10	40	0.25	100
	TP2	0.23	92	0.46	184
	TP3	0.30	120	0.54	216
	TP4	0.40	160	0.70	280
4x-SX, 12x-SX	TP1 (input test jitter)	0.10	40	0.25	100
	TP2	0.25	100	0.48	192
	TP3	0.30	120	0.53	212
	TP4	0.40	160	0.70	280

When operating at DDR speed, IB optical links shall not exceed the deterministic and total jitter values listed in [Table 154](#) at TP2 and TP3 when the input jitter to the Optical Transceiver is compliant with the IB electrical specification.

Table 154 Maximum Jitter of Optical Links for DDR

InfiniBand Link	Compliance Point	Deterministic Jitter		Total Jitter	
		UI	ps	UI	ps
1x-SX, 1x-LX	TP1 (input test jitter)	0.14	28	0.26	52
	TP2	0.26	52	0.452	90.4
	TP3	0.265	53	0.58	116.1
	TP4	0.365	72.6	0.756	151.1
4x-SX, 8x-SX, 12x-SX	TP1 (input test jitter)	0.10	20	0.25	50
	TP2	0.22	44	0.443	89
	TP3	0.22	44.6	0.538	108
	TP4	0.32	64	0.757	151

For all link types, the Signal Conditioner or other component connected at TP4 shall tolerate total jitter of 0.80 UI for SDR (**2.5 Gb/s**) and 0.85 for DDR (**5.0 Gb/s**), which includes 0.10 UI of sinusoidal jitter (SJ) over a swept frequency from 1.5 MHz to 1250 MHz as defined in [Figure 181](#).

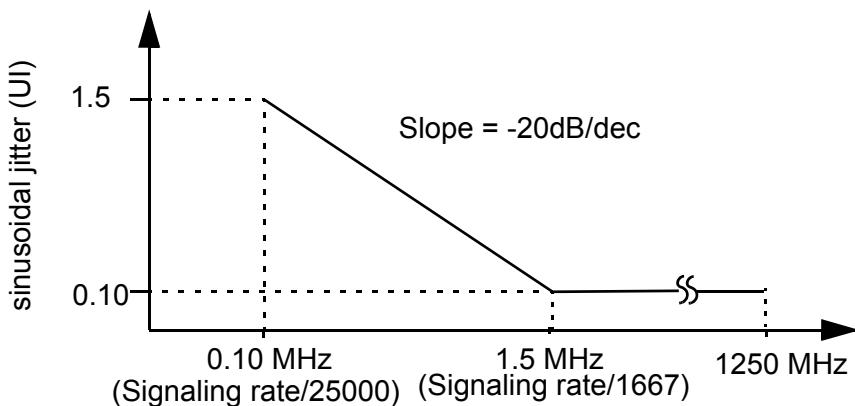


Figure 181 Jitter Tolerance Mask

Architecture Note - Jitter of Optical Links versus Jitter of High Speed Electrical Signaling

The attention of System Designers is drawn to the difference between the jitter specified at TP1 and TP4 in [Table 153](#) compared to the jitter specifications of [Chapter 6: High Speed Electrical Interfaces](#).

In particular, a design which just meets the *informative* jitter specification of TP1 and TP4 will fail the *normative* jitter specification of [Chapter 6: High Speed Electrical Interfaces](#).

Implementation suggestions to deal with this issue are provided herein in [Section 9.8.2](#).

Architecture Note - Jitter

[Table 153](#) and [Table 154](#) specify total jitter (TJ) at BER of 10^{-12} and specifies deterministic jitter (DJ). Random jitter (RJ) can be calculated from TJ and DJ:

$$RJ = TJ - DJ, \text{ where } RJ \text{ is } 14\sigma \text{ for BER of } 10^{-12}.$$

DJ's of successive physical components add linearly. RJ's add in quadrature. TJ values can only be added by breaking them down into DJ and RJ components.

Jitter values are specified at each test point TP1, TP2, TP3, TP4. To determine the amount of deterministic jitter and random jitter that an Optical Transmitter under test adds from TP1 to TP2, the following analysis applies:

$$DJ(\text{Transmitter}) = DJ_2 - DJ_1$$

$$RJ(\text{Transmitter}) = \sqrt{(TJ_2 - DJ_2)^2 - (TJ_1 - DJ_1)^2}$$

where

$DJ_1 = DJ$ at TP1

$DJ_2 = DJ$ at TP2

$TJ_1 = TJ$ at TP1

$TJ_2 = TJ$ at TP2

A similar analysis using TP3 and TP4 applies to testing an Optical Receiver.

9.5.6 OPTICAL JITTER SPECIFICATIONS FOR LINKS OPERATING AT 10.0 GB/S

9.5.6.1 1x SX QDR AND 1x LX QDR OPTICAL JITTER SPECIFICATIONS

C9-7.2.3: All 1x-QDR-SX and 1x-QDR-LX Optical Ports shall comply with the Jitter requirements in [Section 9.5.6.2](#) while operating at QDR speed.

The 1x SX QDR Optical Port shall meet the Optical Jitter specifications as defined for 10GBASE-SR in IEEE Std 802.3, Clause 52. 4x SX QDR shall use the same transmit and receive optical interface and budget specifications and measurements, with the following notes and exceptions:

- The nominal signaling speed shall be 10.0 GBd ± 100 ppm

The 1x LX QDR and Optical Port shall meet the Optical Jitter specifications as defined for 10GBASE-LR in IEEE Std 802.3, Clause 52. 4x LX QDR shall use the same transmit and receive optical interface and budget specifications and measurements, with the following notes and exceptions:

- The nominal signaling speed shall be 10.0 GBd ± 100 ppm

9.5.6.2 4x LX OPTICAL JITTER SPECIFICATIONS

C9-7.2.4: All 4x-LX Optical Ports shall comply with the Jitter requirements in [Section 9.5.6.2](#).

The 4x LX Optical Port shall meet the Optical Jitter specifications as defined for 10GBASE-LR in IEEE Std 802.3, Clause 52. 4x LX shall use the same transmit and receive optical interface and budget specifications and measurements, with the following notes and exceptions:

- The nominal signaling speed shall be 10.0 GBd ± 100 ppm

9.5.7 BIT TO BIT SKEW

C9-8: All IB optical ports shall not exceed the Maximum Skew values allowed across all physical lanes as defined in [Table 155](#) while operating at SDR speed.

C9-8.2.1: All InfiniBand optical ports shall not exceed the Maximum Skew values allowed across all physical Lanes as defined in [Table 156](#) while operating at DDR speed.

C9-8.2.2: All InfiniBand optical ports shall not exceed the Maximum Skew values allowed across all physical Lanes as defined in [Table 157](#) while operating at QDR speed.

[Table 155](#) defines the allowable bit to bit skew across the physical lanes for optical components. All IB optical ports shall limit skew to the maximum values defined in [Table 155](#) when operating at SDR speed. All IB optical ports shall limit skew to the maximum values defined in [Table 156](#) when operating at DDR speed. All IB optical ports shall limit skew to the maximum values defined in [Table 157](#) when operating at QDR speed.

Table 155 Maximum Optical Bit to Bit Skew Values for 2.5 Gb/s

Skew Parameter	Maximum Value
Optical Cable Assembly ^a	3.0 ns
Transmitter ^b	500 ps
Receiver ^c	500 ps

a. An optical cable assembly shall include the optical cable and appropriate optical connectors at each end of the cable.

b. Between any two physical lanes within a transmitter.

c. Between any two physical lanes within a receiver.

Table 156 Maximum Optical Bit to Bit Skew Values for 5.0 Gb/s

Skew Parameter	Maximum Value
Optical Cable Assembly ^a	1.5 ns
Transmitter ^b	250 ps
Receiver ^c	250 ps

- a. An optical cable assembly shall include the optical cable and appropriate optical connectors at each end of the cable.
b. Between any two physical lanes within a transmitter.
c. Between any two physical lanes within a receiver.

Table 157 Maximum Optical Bit to Bit Skew Values for 10.0 Gb/s

Skew Parameter	Maximum Value
Optical Cable Assembly ^a	0.75 ns
Transmitter ^b	125 ps
Receiver ^c	125 ps

- a. An optical cable assembly shall include the optical cable and appropriate optical connectors at each end of the cable.
b. Between any two physical lanes within a transmitter.
c. Between any two physical lanes within a receiver.

9.5.8 1x SDR LINKS - AT 2.5 GB/s

A 1x-SX link operates in the 850 nm wavelength band using multimode (MM) fiber. The optical parameters have been selected so as to allow typical 1x-SX Optical Transceivers to use a GaAs Vertical Cavity Surface Emitting Laser (VCSEL) and a photodetector, trans-impedance pre-amplifier and limiting post-amplifier. Three fiber types are specified for 1x-SX:

- i) 1x-SX/50 link: 500 MHz.km 50 µm / 125 µm MM fiber
- ii) 1x-SX/50 link: 2000 MHz.km 50 µm / 125 µm MM fiber
- iii) 1x-SX/62 link: 200 MHz.km 62.5 µm / 125 µm MM fiber

A 1x-LX link operates in the 1300 nm wavelength band using singlemode (SM) fiber. The optical parameters have been selected to allow a typical 1x-LX Optical Transceiver to use an uncooled InP-based Fabry-Perot (FP) laser, Distributed Feedback (DFB) laser or VCSEL laser, and an InGaAs photodetector, trans-impedance pre-amplifier and limiting post-amplifier. A trade-off curve is provided in [Figure 182](#) between Center Wavelength and RMS Spectral Width to allow designers to select appropriate laser technology while still ensuring link operation. Only one fiber type is specified for 1x-LX: single-mode non-dispersion shifted.

9.5.8.1 1x EYE SAFETY

C9-9: The optical power coupled into the fiber for 1x-SX and 1x-LX shall be limited to a maximum value with Class I laser safety operation in accordance with CDRH and IEC

60825-1 Amendment 2 Radiation Safety of Laser Products: Equipment Classification, Requirements and User Guide.

9.5.8.2 1x-SX OPTICAL PARAMETERS

C9-10: This compliance statement is obsolete and has been replaced by [C9-10.2.1](#).

C9-10.2.1: All 1x-SX Optical Ports shall comply with the Transmitter and Receiver requirements in [Section 9.5.8.2](#).

[Table 158](#) gives the link budgets for 1x-SX fiber optic fiber links running at 2.5 GTransfers/second. Fiber plant specifications are described in [Section 9.7](#).

Optical Passive Loss is the loss resulting from connections between Fiber Optic Segments (adapters or splices), and attenuation attributable to the fiber cable plant. Optical System Penalty includes all other link penalties other than Optical Passive Loss.

Table 158 Link Parameters - 1x-SX

Parameter	IB-1x-SX/50		IB-1x-SX/62		Unit	Note
	Minimum	Maximum	Minimum	Maximum		
Signaling Rate		2500		2500	Mb/s	
Rate tolerance		±100		±100	ppm	
Optical Passive Loss		2.44		2.0	dB	
Optical System Penalty		3.56		4.0	dB	
Total link power budget	6		6		dB	
Worst case operating range		2 - 250 ^a 2-500 ^b		2 - 125	m	c
Fiber mode-field (core) diameter		50		62.5	μm	

a. 1x-SX/50 link: 500 MHz.km 50 μm / 125 μm MM fiber

b. 1x-SX/50 link: 2000 MHz.km 50 μm / 125 μm MM fiber

c. Longer operating distance than the range specified here can be achieved using transmitters, receivers and/or cables meeting specification but performing better than worst-case.

An Optical Transmitter for a 1x-SX link shall meet the parameters specified in [Table 159](#) at TP2.

If any high-speed electrical input signals are less than the V_{RSD} (Signal Threshold) as defined in then the optical modulation amplitude shall not exceed 2.0 μW .

Table 159 Optical Transmitter Parameters - 1x-SX

Parameter	Minimum	Maximum	Unit	Note
Type	Laser			
Center Wavelength	830	860	nm	
RMS spectral width		0.85	nm	
Average launched power		-4.0	dBm	a
Optical Modulation Amplitude (OMA)	0.196		mW	b
RMS mean of 20% - 80% Rise/Fall time		150	ps	c
RIN_{12} (OMA)		-117	dB/Hz	

a. Average launched power, max. is the lesser of the eye safety limit or Average receiver power, max.

b. Optical modulation amplitude values are peak-to-peak.

c. Optical rise and fall time specifications are based on unfiltered waveforms. See [Section 9.5.3.2](#).

An Optical Receiver for a 1x-SX link shall meet the parameters specified in [Table 160](#) at TP3. Conformance testing for a stressed receiver at TP3 shall follow the methods of Annex A of Fibre Channel Physical Interface revision 8.0.

If the average received optical power on any Lane is less than -30 dBm then the corresponding high-speed electrical signaling output shall be squelched to less than V_{RSD} (Signal Threshold) as defined in [Chapter 6: High Speed Electrical Interfaces](#). The electrical outputs of every lane with a optical power greater than -20dBm shall not be squelched.

Table 160 Optical Receiver Parameters - 1x-SX

Parameter	Minimum	Maximum	Unit	Note
Average received power		-1.5	dBm	
Optical modulation amplitude	0.050		mW	a
Return loss of receiver	12		dB	

Table 160 Optical Receiver Parameters - 1x-SX

Parameter	Minimum	Maximum	Unit	Note
Stressed receiver sensitivity (OMA)	0.102		mW	a
Stressed receiver ISI test	2.0		dB	
Stressed receiver DCD component of DJ (at Tx)	40		ps	
Receiver electrical 3 dB upper cutoff frequency		2.8	GHz	
Receiver electrical 10 dB upper cutoff frequency		6.0	GHz	

a. Optical modulation amplitude values are peak-to-peak.

9.5.8.3 1x-LX OPTICAL PARAMETERS

C9-11: This compliance statement is obsolete and has been replaced by [C9-11.2.1](#).

C9-11.2.1: All 1x-LX Optical Ports shall comply with the Transmitter and Receiver requirements in [Section 9.5.8.3](#).

[Table 161](#) gives the link budgets for 1x-LX single-mode fiber optic links running at 2.5GTransfers/second. Fiber plant specifications are described in [Section 9.7](#).

Table 161 Link Parameters - 1x-LX

Parameter	Minimum	Maximum	Unit	Note
Signaling Rate		2500	Mb/s	
Rate tolerance		±100	ppm	
Optical Passive Loss		6.64	dB	
Optical System Penalty		2.36	dB	
Total link power budget	9.0		dB	
Worst case operating range		2 - 10,000	m	a
Fiber mode-field (core) diameter		9	μm	

a. Longer operating distance than the range specified here can be achieved using transmitters, receivers and/or cables meeting specification but performing better than worst-case.

An Optical Transmitter for a 1x-LX link shall meet the parameters specified in [Table 162](#) at TP2.

If any high-speed electrical input signals are less than the V_{RSD} (Signal Threshold) as defined in [Figure 182](#) then the optical modulation amplitude shall not exceed 2.0 μ W.

In addition, the RMS spectral width of a 1x-LX Optical Transmitter shall lie on or below the curve shown in [Figure 182](#) as a function of Optical Transmitter Center Wavelength. The curve was calculated using the worst-case fiber at a given Optical Transmitter Center Wavelength for all operating conditions. Specifically, the zero-dispersion wavelength of the fiber was chosen to be 1324 nm for Optical Transmitter Center Wavelengths below approximately 1312 nm, and the zero-dispersion wavelength of the fiber was chosen to be 1300 nm for Optical Transmitter Center Wavelengths above approximately 1312 nm.

Table 162 Optical Transmitter Parameters - 1x-LX

Parameter	Minimum	Maximum	Unit	Note
Type	Laser			
Center Wavelength	1270	1360	nm	
RMS spectral width			nm	a
Average launched power		-3.0	dBm	b
Optical Modulation Amplitude (OMA)	0.186		mW	c
RMS mean of 20% - 80% Rise/Fall time	150		ps	d
RIN ₁₂ (OMA)		-120	dB/Hz	

a. See text and [Figure 182](#).

b. Average launched power, max. is the lesser of the eye safety limit or Average receiver power, max.

c. Optical modulation amplitude values are peak-to-peak.

d. Optical rise and fall time specifications are based on unfiltered waveforms. See [Section 9.5.3.2](#).

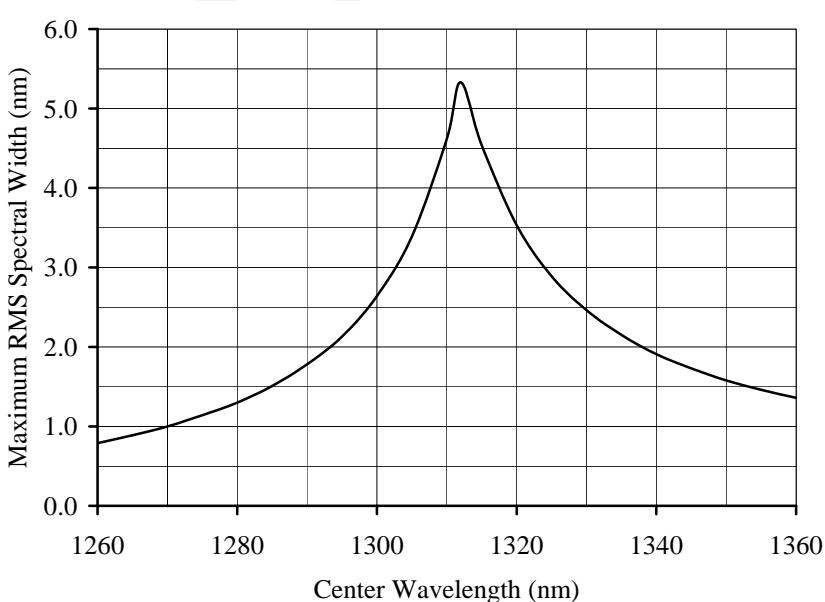
An Optical Receiver for a 1x-LX link shall meet the parameters specified in [Table 163](#) at TP3.

If the average received optical power on any Lane is less than -30 dBm then the corresponding high-speed electrical signaling output shall be squelched to less than V_{RSD} (Signal Threshold) as defined in [Chapter 6: High Speed Electrical Interfaces](#). The electrical outputs of every Lane with a optical power greater than -20dBm shall not be squelched.

Table 163 Optical Receiver Parameters - 1x-LX

Parameter	Minimum	Maximum	Unit	Note
Average received power		-1.5	dBm	
Optical modulation amplitude	0.0234		mW	a
Return loss of receiver	20		dB	
Stressed receiver sensitivity (OMA)	0.0365		mW	a
Stressed receiver ISI test	0.58		dB	
Stressed receiver DCD component of DJ (at Tx)	40		ps	
Receiver electrical 3 dB upper cutoff frequency		2.8	GHz	
Receiver electrical 10 dB upper cutoff frequency		6.0	GHz	

a. Optical modulation amplitude values are peak-to-peak.

**Figure 182 1x-LX Trade-off between RMS Spectral Width and Center Wavelength**

9.5.9 1x DDR LINKS - AT 5.0 GB/S

A 1x-DDR-SX link uses similar transmitter, receiver, and fiber technology as 1x-SX links described in [Section 9.5.8](#), and differ in bit rate, and in the parameters described below. A 1x-DDR-LX link also uses similar technology as 1x-LX links described in [Section 9.5.8](#).

9.5.9.1 EYE SAFETY

C9-11.2.2: The optical power coupled into the fiber for 1x-DDR-SX and 1x-DDR-LX shall be limited to a maximum value with Class I laser safety operation in accordance with CDRH and IEC 60825-1 Amendment 2 Radiation Safety of Laser Products: Equipment Classification, Requirements and User Guide.

9.5.9.2 1x-DDR-SX OPTICAL PARAMETERS

C9-11.2.3: All 1x-DDR-SX Optical Ports shall comply with the Transmitter and Receiver requirements in [Section 9.5.9.2](#).

[Table 164](#) gives the link budgets for 1x-DDR-SX fiber optic fiber links running at 5.0 GTransfers/second. Fiber plant specifications are described in [Section 9.7](#).

Optical Passive Loss is the loss resulting from connections between Fiber Optic Segments (adapters or splices), and attenuation attributable to the fiber cable plant. Optical System Penalty includes all other link penalties other than Optical Passive Loss.

Table 164 Link Parameters - 1x-DDR-SX

Parameter	IB-1x-SX/50		IB-1x-SX/62		Unit	Note
	Minimum	Maximum	Minimum	Maximum		
Signaling Rate		5000		5000	Mb/s	
Rate tolerance		±100		±100	ppm	
Optical Passive Loss		1.97		1.76	dB	
Optical System Penalty		3.6		4.4	dB	
Total link power budget	7.93		7.93		dB	
Worst case operating range		2 - 125 ^a 2-200 ^b		2 - 65	m	c
Fiber mode-field (core) diameter		50		62.5	μm	

a. 1x-SX/50 link: 500 MHz.km 50 μm / 125 μm MM fiber

b. 1x-SX/50 link: 2000 MHz.km 50 μm / 125 μm MM fiber

c. Longer operating distance than the range specified here can be achieved using transmitters, receivers and/or cables meeting specification but performing better than worst-case.

An Optical Transmitter for a 1x-DDR-SX link shall meet the parameters specified in [Table 165](#) at TP2.

Table 165 Optical Transmitter Parameters - 1x-DDR-SX

Parameter	Minimum	Maximum	Unit	Note
Type	Laser			
Center Wavelength	830	860	nm	
RMS spectral width		0.85	nm	
Average launched power		-2.0	dBm	a
Optical Modulation Amplitude (OMA)	0.247		mW	b
RMS mean of 20% - 80% Rise/Fall time		75	ps	c
RIN ₁₂ (OMA)		-123	dB/Hz	

a. Average launched power, max. is the lesser of the eye safety limit or Average receiver power, max.

b. Optical modulation amplitude values are peak-to-peak.

c. Optical rise and fall time specifications are based on unfiltered waveforms. See [Section 9.5.3.2](#).

An Optical Receiver for a 1x-DDR-SX link shall meet the parameters specified in [Table 166](#) at TP3. Conformance testing for a stressed receiver at TP3 shall follow the methods of Annex A of Fibre Channel Physical Interface revision 8.0.

Table 166 Optical Receiver Parameters - 1x-DDR-SX

Parameter	Minimum	Maximum	Unit	Note
Average received power	0.0		dBm	
Optical modulation amplitude (Informative)		0.040	mW	a
Return loss of receiver	12		dB	
Stressed receiver sensitivity (OMA)	0.074		mW	a, b
Stressed receiver ISI test		2.48	dB	
Stressed receiver DCD component of DJ (at Tx)		17.0	ps	

a. Optical modulation amplitude values are peak-to-peak.

b. Stressed receiver entries for Sensitivity and ISI differ for each fiber type. The values here apply to 50u, 500 MHz-km fiber.

9.5.9.3 1x-DDR-LX OPTICAL PARAMETERS

C9-11.2.4: All 1x-DDR-LX Optical Ports shall comply with the Transmitter and Receiver requirements in [Section 9.5.9.3](#).

[Table 167](#) gives the link budgets for 1x-LX single-mode fiber optic links running at 2.5GTransfers/second. Fiber plant specifications are described in [Section 9.7](#).

Table 167 Link Parameters - 1x-DDR-LX

Parameter	Minimum	Maximum	Unit	Note
Signaling Rate		5000	Mb/s	
Rate tolerance		±100	ppm	
Optical Passive Loss		5.66	dB	
Optical System Penalty		2.4	dB	
Total link power budget	9.8		dB	
Worst case operating range		2 - 10,000	m	a
Fiber mode-field (core) diameter		9	μm	

a. Longer operating distance than the range specified here can be achieved using transmitters, receivers and/or cables meeting specification but performing better than worst-case.

An Optical Transmitter for a 1x-DDR-LX link shall meet the parameters specified in [Table 168](#) at TP2.

In addition, the RMS spectral width of a 1x-DDR-LX Optical Transmitter shall lie on or below the curve shown in [Figure 183](#) as a function of Optical Transmitter Center Wavelength and minimum Transmit OMA. The curve was calculated using the worst-case fiber at a given Optical Transmitter Center Wavelength for all operating conditions. Specifically, the zero-dispersion wavelength of the fiber was chosen to be 1324 nm for Optical Transmitter Center Wavelengths below approximately 1312 nm, and the zero-dispersion wavelength of the fiber was chosen to be 1300 nm for Optical Transmitter Center Wavelengths above approximately 1312 nm.

Table 168 Optical Transmitter Parameters - 1x-DDR-LX

Parameter	Minimum	Maximum	Unit	Note
Type	Laser			
Center Wavelength	1270	1360	nm	
RMS spectral width		0.47	nm	a
Average launched power		-1.0	dBm	b
Optical Modulation Amplitude (OMA)	0.29		mW	c
RMS mean of 20% - 80% Rise/Fall time		75	ps	d
RIN ₁₂ (OMA)		-123	dB/Hz	

a. See text and [Figure 183](#).

b. Average launched power, max. is the lesser of the eye safety limit or Average receiver power, max.

c. Optical modulation amplitude values are peak-to-peak.

d. Optical rise and fall time specifications are based on unfiltered waveforms. See [Section 9.5.3.2](#).

An Optical Receiver for a 1x-DDR-LX link shall meet the parameters specified in [Table 169](#) at TP3.

Table 169 Optical Receiver Parameters - 1x-DDR-LX

Parameter	Minimum	Maximum	Unit	Note
Average received power	-1.0		dBm	
Optical modulation amplitude (Informative)		0.029	mW	a
Return loss of receiver	12		dB	
Stressed receiver sensitivity (OMA)	0.030		mW	a
Stressed receiver ISI test	1.45		dB	
Stressed receiver DCD component of DJ (at Tx)	16.3		ps	
Receiver electrical 3 dB upper cutoff frequency		5.3	GHz	b
Receiver electrical 10 dB upper cutoff frequency			GHz	2

a. Optical modulation amplitude values are peak-to-peak.

b. These rows may be deleted since receivers may operate over all three signal rates?

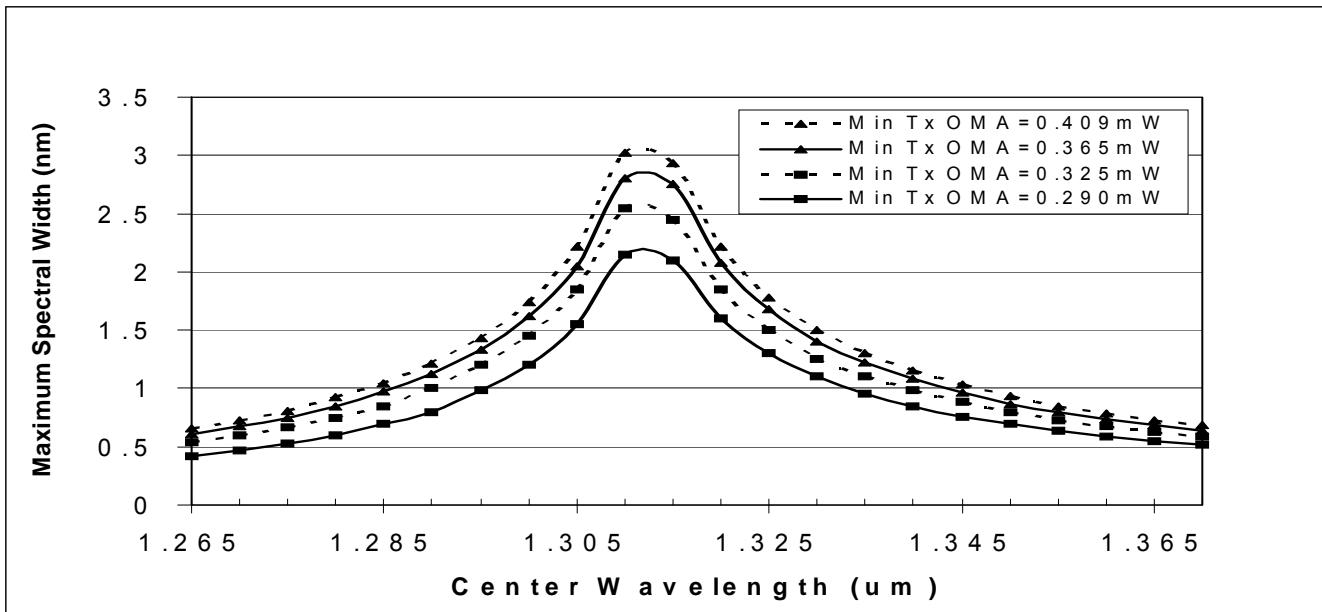


Figure 183 1x-DDR-LX Trade-off between OMA, RMS Spectral Width, and Center Wavelength

9.5.10 4x SDR LINKS - AT 2.5 GB/s

C9-12: This compliance statement is obsolete and has been replaced by [C9-12.2.1](#).

C9-12.1.1: This compliance statement is obsolete and has been replaced by [C9-12.2.2](#).

C9-12.2.1: All 4x-SX Optical Ports shall comply with the Eye Safety, Transmitter, and Receiver requirements in [Section 9.5.10.1](#) and [Section 9.5.10.2](#).

C9-12.2.2: All 4x-LX Optical Ports shall comply with the Eye Safety, Transmitter, and Receiver requirements in [Section 9.5.10.3](#) and [Section 9.5.10.4](#)

A 4x-SX link operates in the 850 nm wavelength band using multimode (MM) fiber. The optical parameters have been selected so as to allow typical 4x-SX Optical Transceivers to use an array of GaAs VCSELs and an array of photodetectors, trans-impedance pre-amplifiers and limiting post-amplifiers. Three fiber types are specified for 4x-SX:

- i) 4x-SX/50: 500 MHz.km 50 μm / 125 μm MM fiber
- ii) 4x-SX/50: 2000 MHz.km 50 μm / 125 μm MM fiber

iii) 4x-SX/62: 200 MHz.km 62.5 μm / 125 μm MM fiber

9.5.10.1 4x-SX EYE SAFETY

The optical power emitted from a 4x-SX port shall be limited to a maximum value for Class 1M laser safety in accordance with IEC/EN 60825-1 Amendment 2 Safety of Laser Products, part 1: Equipment classification, requirements, and user's guide and FDA/CDRH 21 CFR 1040.10

For systems that are required to meet existing IEC Class I laser safety regulations, Open Fiber Control (OFC) or other similar vendor-specific protocols **should** be implemented to meet the CDRH and IEC requirement. Any such implementation shall be transparent to all InfiniBand layers.

Recommendation to System Designer

Without any form of open fiber control, the power levels specified for the 4x-SX Optical Transmitter will meet the new relaxed specification for IEC Class 1M.

9.5.10.2 4x-SX OPTICAL PARAMETERS

[Table 170](#) gives the link budgets for 4x-SX multimode fiber optic links running at 2.5 GTransfers/second. Fiber plant specifications are described in [Section 9.7](#).

Table 170 Link Parameters - 4x-SX

Parameter	IB-4x-SX/50		IB-4x-SX/62		Unit	Note
	Minimum	Maximum	Minimum	Maximum		
Signaling Rate		2500		2500	Mb/s	
Rate tolerance		±100		±100	ppm	
Optical Passive Loss		1.9		1.8	dB	
Optical System Penalty		2.9		3.0	dB	
Total link power budget	4.8		4.8		dB	
Worst case operating range		2 - 125 ^a 2 - 200 ^b		2 - 75	m	c
Fiber mode-field (core) diameter	50		62.5		μm	

a. 4x-SX/50: 500 MHz.km 50 μm / 125 μm MM fiber

b. 4x-SX/50: 2000 MHz.km 50 μm / 125 μm MM fiber

c. Longer operating distance than the range specified here can be achieved using transmitters, receivers and/or cables meeting specification but performing better than worst-case.

An Optical Transmitter for a 4x-SX link shall meet the parameters specified in [Table 171](#) at TP2.

If any high-speed electrical input signals are less than the V_{RSD} (Signal Threshold) as defined in [Chapter 6: High Speed Electrical Interfaces](#) then the optical modulation amplitude shall not exceed 5.0 μ W.

Table 171 Optical Transmitter Parameters - 4x-SX

Parameter	Minimum	Maximum	Unit	Note
Type	Laser			
Center Wavelength	830	860	nm	
RMS spectral width		0.85	nm	
Average launched power		-2.5	dBm	a
Optical Modulation Amplitude (OMA)	0.150		mW	b
RMS mean of 20% - 80% Rise/Fall time		150	ps	c
RIN ₁₂ (OMA)		-117	dB/Hz	

a. Average launched power per fiber, max. is the lesser of the eye safety limit or Average receiver power, max.

b. Optical modulation amplitude values are peak-to-peak.

c. Optical rise and fall time specifications are based on unfiltered waveforms. See [Section 9.5.3.2](#).

An Optical Receiver for a 4x-SX link shall meet the parameters specified in [Table 172](#) at TP3.

If the average received optical power on any Lane is less than -26 dBm then the corresponding high-speed electrical signaling output shall be squelched to less than V_{RSD} (Signal Threshold) as defined in [Chapter 6: High Speed Electrical Interfaces](#). The electrical outputs of every lane with a optical power greater than -18 dBm shall not be squelched.

Table 172 Optical Receiver Parameters - 4x-SX

Parameter	Minimum	Maximum	Unit	Note
Average received power		-1.5	dBm	
Optical modulation amplitude	0.050		mW	a
Return loss of receiver	12		dB	

Table 172 Optical Receiver Parameters - 4x-SX (Continued)

Parameter	Minimum	Maximum	Unit	Note
Stressed receiver sensitivity (OMA)	0.085		mW	a
Stressed receiver ISI test	0.90		dB	
Stressed receiver DCD component of DJ (at Tx)	60		ps	
Receiver electrical 3 dB upper cutoff frequency		2.8	GHz	
Receiver electrical 10 dB upper cutoff frequency		6.0	GHz	

a. Optical modulation amplitude values are peak-to-peak.

9.5.10.3 4x-LX EYE SAFETY

The 4x LX Optical Port shall meet the Eye safety as defined for 10GBASE-LR in IEEE Std 802.3, Clause 52.

9.5.10.4 4x-LX OPTICAL PARAMETERS

The 4x-LX Optical Port shall meet the Optical Parameters as defined for 10GBASE-LR in IEEE Std 802.3, Clause 52. 4x LX shall use the same transmit and receive optical interface and budget specifications and measurements, with the following notes and exceptions:

- The nominal signaling speed for 4x-LX shall be 10.0 GBd \pm 100 ppm

9.5.11 4x-DDR-SX LINK

C9-12.2.3: All 4x-DDR-SX Optical Ports shall comply with the Eye Safety, Transmitter, and Receiver requirements in [Section 9.5.11.1](#) and [Section 9.5.11.2](#).

A 4x-DDR-SX link operates in the 850 nm wavelength band using multimode (MM) fiber. The optical parameters have been selected so as to allow typical 4x-DDR-SX Optical Transceivers to use an array of GaAs VCSELs and an array of photodetectors, trans-impedance pre-amplifiers and limiting post-amplifiers. Three fiber types are specified for 4x-DDR-SX:

- i) 4x-SX/50: 500 MHz.km 50 μ m / 125 μ m MM fiber
- ii) 4x-SX/50: 2000 MHz.km 50 μ m / 125 μ m MM fiber
- iii) 4x-SX/62: 200 MHz.km 62.5 μ m / 125 μ m MM fiber

9.5.11.1 4x-DDR-SX EYE SAFETY

The optical power emitted from a 4x-DDR-SX port shall be limited to a maximum value for Class 1M laser safety in accordance with IEC/EN 60825-1 Amendment 2 Safety of Laser Products, part 1: Equipment classification, requirements, and user's guide and FDA/CDRH 21 CFR 1040.10

For systems that are required to meet existing IEC Class I laser safety regulations, Open Fiber Control (OFC) or other similar vendor-specific protocols **should** be implemented to meet the CDRH and IEC requirement. Any such implementation shall be transparent to all InfiniBand layers.

Recommendation to System Designer

Without any form of open fiber control, the power levels specified for the 4x-DDR-SX Optical Transmitter will meet the new relaxed specification for IEC Class 1M.

9.5.11.2 4x-DDR-SX OPTICAL PARAMETERS

[Table 173](#) gives the link budgets for 4x-DDR-SX multimode fiber optic links running at 5.0 GTransfers/second. Fiber plant specifications are described in [Section 9.7](#).

Table 173 Link Parameters - 4x-DDR-SX

Parameter	IB-4x-SX/50		IB-4x-SX/62		Unit	Note
	Minimum	Maximum	Minimum	Maximum		
Signaling Rate		5000		5000	Mb/s	
Rate tolerance		±100		±100	ppm	
Optical Passive Loss		1.78 ^a 2.06 ^b		1.70	dB	
Optical System Penalty		2.5 ^c 2.3 ^d		3.2	dB	
Total link power budget	6.25		6.25		dB	
Worst case operating range		2 - 75 ^e 2 - 150 ^f		2 - 50	m	^g
Fiber mode-field (core) diameter	50		62.5		μm	

a. 4x-SX/50: 500 MHz.km 50 μm / 125 μm MM fiber

b. 4x-SX/50: 2000 MHz.km 50 μm / 125 μm MM fiber

c. 4x-SX/50: 500 MHz.km 50 μm / 125 μm MM fiber

d. 4x-SX/50: 2000 MHz.km 50 μm / 125 μm MM fiber

e. 4x-SX/50: 500 MHz.km 50 μm / 125 μm MM fiber

f. 4x-SX/50: 2000 MHz.km 50 μm / 125 μm MM fiber

g. Longer operating distance than the range specified here can be achieved using transmitters, receivers and/or cables meeting specification but performing better than worst-case.

An Optical Transmitter for a 4x-DDR-SX link shall meet the parameters specified in [Table 174](#) at TP2.

Table 174 Optical Transmitter Parameters - 4x-DDR-SX

Parameter	Minimum	Maximum	Unit	Note
Type	Laser			
Center Wavelength	830	860	nm	
RMS spectral width		0.85	nm	
Average launched power		-2.0	dBm	a
Optical Modulation Amplitude (OMA)	0.224		mW	b
RMS mean of 20% - 80% Rise/Fall time		75	ps	c
RIN ₁₂ (OMA)		-122	dB/Hz	

a. Average launched power per fiber, max. is the lesser of the eye safety limit or Average receiver power, max.

b. Optical modulation amplitude values are peak-to-peak.

c. Optical rise and fall time specifications are based on unfiltered waveforms. See [Section 9.5.3.2](#).

An Optical Receiver for a 4x-DDR-SX link shall meet the parameters specified in [Table 175](#) at TP3.

Table 175 Optical Receiver Parameters - 4x-DDR-SX

Parameter	Minimum	Maximum	Unit	Note
Average received power	-1.5		dBm	
Optical modulation amplitude (Informative)		0.053	mW	a Informative
Return loss of receiver	12		dB	
Stressed receiver sensitivity (OMA)	0.078		mW	b
Stressed receiver ISI test	1.50		dB	
Stressed receiver DCD component of DJ (at Tx)	16.3		ps	

a. Optical modulation amplitude values are peak-to-peak.

b. Stressed Rx entries for Sensitivity and ISI differ for each fiber type. The entries shown are for 50 μm, 500 MHz.km fiber

9.5.12 8x-SX LINKS - AT 2.5 GB/s

C9-12.2.4: All 8x SX Optical Ports shall meet all the optical specifications for 12x-SX in [Section 9.5.13](#) with the exception that 8x SX has only eight receive and transmit lanes.

9.5.13 12x-SX LINKS - AT 2.5 GB/s

C9-13: This compliance statement is obsolete and has been replaced by [C9-13.2.1](#).

C9-13.2.1: All 12x-SX Optical Ports shall comply with the Eye Safety, Transmitter, and Receiver requirements in [Section 9.5.13](#).

A 12x-SX link operates in the 850 nm wavelength band using multimode (MM) fiber. The optical parameters have been selected so as to allow typical 12x-SX Optical Transceivers to use an array of GaAs VCSELs and an array of photodetectors, trans-impedance pre-amplifiers and limiting post-amplifiers. Two fiber types are specified for 12x-SX:

- i) 12x-SX/50 link: 500 MHz.km 50 μm / 125 μm MM fiber
- ii) 12x-SX/50 link: 2000 MHz.km 50 μm / 125 μm MM fiber
- iii) 12x-SX/62 link: 200 MHz.km 62.5 μm / 125 μm MM fiber

9.5.13.1 12-SX EYE SAFETY

The optical power emitted from a IB-12x-SX port shall be limited to a maximum value for Class 1M laser safety in accordance with IEC/EN 60825-1 Amendment 2 Safety of Laser Products, part 1: Equipment classification, requirements, and user's guide and FDA/CDRH 21 CFR 1040.10.

Without any form of open fiber control, the power levels specified for the 12x-SX Optical Transmitter will meet the new relaxed specification for IEC Class 1M.

9.5.13.2 12x-SX OPTICAL PARAMETERS

[Table 176](#) gives the link budgets for 12x-SX and 8x-SX multimode fiber optic links running at 2.5 GTransfers/second. Fiber plant specifications are described in [Section 9.7](#).

Table 176 Link Parameters - 12x-SX and 8x-SX

Parameter	IB-12x-SX/50		IB-12x-SX/62		Unit	Note
	Minimum	Maximum	Minimum	Maximum		
Signaling Rate		2500		2500	Mb/s	
Rate tolerance		±100		±100	ppm	
Optical Passive Loss		1.9		1.8	dB	
Optical System Penalty		2.9		3.0	dB	
Total link power budget	4.8		4.8		dB	
Worst case operating range		125 ^a 200 ^b		75	m	c
Fiber mode-field (core) diameter		50		62.5	μm	

a. 12x-SX/50 link: 500 MHz.km 50 μm / 125 μm MM fiber

b. 12x-SX/50 link: 2000 MHz.km 50 μm / 125 μm MM fiber

c. Longer operating distance than the range specified here can be achieved using transmitters, receivers and/or cables meeting specification but performing better than worst-case.

An Optical Transmitter for a 12x-SX or 8x-SX link shall meet the parameters specified in [Table 177](#) at TP2.

If any high-speed electrical input signals are less than the V_{RSD} (Signal Threshold) as defined in [Table 177](#) then the optical modulation amplitude shall not exceed 5.0 μ W.

Table 177 Optical Transmitter Parameters - 12x-SX and 8x-SX

Parameter	Minimum	Maximum	Unit	Note
Type	Laser			
Center Wavelength	830	860	nm	
RMS spectral width		0.85	nm	
Average launched power		-2.5	dBm	a
Optical modulation amplitude	0.150		mW	b
RMS mean of 20% - 80% Rise/Fall time		150	ps	c
RIN ₁₂ (OMA)		-117	dB/Hz	

a. Average launched power per fiber, max. is the lesser of the eye safety limit or Average receiver power, max.

b. Optical modulation amplitude values are peak-to-peak.

c. Optical rise and fall time specifications are based on unfiltered waveforms. See [Section 9.5.3.2](#).

An Optical Receiver for a 12x-SX or an 8x-SX link shall meet the parameters specified in [Table 178](#) at TP3.

If the average received optical power on any Lane is less than -26 dBm then the corresponding high-speed electrical signaling output shall be squelched to less than V_{RSD} (Signal Threshold) as defined in [Chapter 6: High Speed Electrical Interfaces](#). The electrical outputs of every lane with a optical power greater than -18 dBm shall not be squelched.

Table 178 Optical Receiver Parameters - 12x-SX and 8x-SX

Parameter	Minimum	Maximum	Unit	Note
Average received power		-1.5	dBm	
Optical Modulation Amplitude (OMA)	0.050		mW	a
Return loss of receiver	12		dB	

Table 178 Optical Receiver Parameters - 12x-SX and 8x-SX

Parameter	Minimum	Maximum	Unit	Note
Stressed receiver sensitivity (OMA)	0.085		mW	a
Stressed receiver ISI test	0.90		dB	
Stressed receiver DCD component of DJ (at Tx)	60		ps	
Receiver electrical 3 dB upper cutoff frequency		2.8	GHz	
Receiver electrical 10 dB upper cutoff frequency		6.0	GHz	

a. Optical modulation amplitude values are peak-to-peak.

9.5.14 8x-DDR-SX AND 12x-DDR-SX LINKS

C9-13.2.2: All 8x-SX-DDR and 12x-SX-DDR Optical Ports shall comply with the Eye Safety, Transmitter, and Receiver requirements in [Section 9.5.14](#)

A 8x-DDR-SX and 12x-DDR-SX links operates in the 850 nm wavelength band using multimode (MM) fiber. The optical parameters have been selected so as to allow typical 8x-DDR-SX and 12x-DDR-SX Optical Transceivers to use an array of GaAs VCSELs and an array of photodetectors, trans-impedance pre-amplifiers and limiting post-amplifiers. Three fiber types are specified for 8x-DDR-SX and 12x-DDR-SX:

- i) 12x-SX/50 link: 500 MHz.km 50 µm / 125 µm MM fiber
- ii) 12x-SX/50 link: 2000 MHz.km 50 µm / 125 µm MM fiber
- iii) 12x-SX/62 link: 200 MHz.km 62.5 µm / 125 µm MM fiber

9.5.14.1 8x-DDR-SX AND 12x-DDR-SX EYE SAFETY

The optical power coupled into the fiber for 8x-DDR-SX and 12x-DDR-SX shall be limited to a maximum value with Class 1M laser safety operation in accordance with CDRH and EN 60825-1 Radiation Safety of Laser Products: Equipment Classification, Requirements and User Guide.

For systems that are required to meet existing IEC Class I laser safety regulations, Open Fiber Control (OFC) or other similar vendor-specific protocols **should** be implemented to meet the CDRH and IEC requirement. Any such implementation shall be transparent to all InfiniBand layers.

9.5.14.2 8x-DDR-SX AND 12x-DDR-SX OPTICAL PARAMETERS

[Table 179](#) gives the link budgets for 8x-DDR-SX and 12x-DDR-SX multimode fiber optic links running at 5.0 GTransfers/second. Fiber plant specifications are described in [Section 9.7](#).

Table 179 Link Parameters - 8x-DDR-SX and 12x-DDR-SX

Parameter	IB-12x-SX/50		IB-12x-SX/62		Unit	Note
	Minimum	Maximum	Minimum	Maximum		
Signaling Rate		5000		5000	Mb/s	
Rate tolerance		± 100		± 100	ppm	
Optical Passive Loss		1.78 ^a 2.06 ^b		1.7	dB	
Optical System Penalty		2.5 ^c 2.3 ^d		3.2	dB	
Total link power budget	6.25		6.25		dB	
Worst case operating range		2-75 ^e 2-150 ^f		2-50	m	^g
Fiber mode-field (core) diameter	50		62.5		μm	

a. 12x-SX/50 link: 500 MHz.km 50 μm / 125 μm MM fiber

b. 12x-SX/50 link: 2000 MHz.km 50 μm / 125 μm MM fiber

c. 12x-SX/50 link: 500 MHz.km 50 μm / 125 μm MM fiber

d. 12x-SX/50 link: 2000 MHz.km 50 μm / 125 μm MM fiber

e. 12x-SX/50 link: 500 MHz.km 50 μm / 125 μm MM fiber

f. 12x-SX/50 link: 2000 MHz.km 50 μm / 125 μm MM fiber

g. Longer operating distance than the range specified here can be achieved using transmitters, receivers and/or cables meeting specification but performing better than worst-case.

An Optical Transmitter for a 8x-DDR-SX and 12x-DDR-SX link shall meet the parameters specified in [Table 180](#) at TP2.

Table 180 Optical Transmitter Parameters - 8x-DDR-SX and 12x-DDR-SX

Parameter	Minimum	Maximum	Unit	Note
Type		Laser		
Center Wavelength	830	860	nm	
RMS spectral width		0.85	nm	
Average launched power		-2.0	dBm	^a
Optical modulation amplitude	0.224		mW	^b
RMS mean of 20% - 80% Rise/Fall time		75	ps	^c
RIN ₁₂ (OMA)		-122	dB/Hz	

a. Average launched power per fiber, max. is the lesser of the eye safety limit or Average receiver power, max.

b. Optical modulation amplitude values are peak-to-peak.

c. Optical rise and fall time specifications are based on unfiltered waveforms. See [Section 9.5.3.2](#).

An Optical Receiver for a 12x-DDR-SX link shall meet the parameters specified in [Table 178](#) at TP3.

Table 181 Optical Receiver Parameters - 12x-DDR-SX and 8x-DDR-SX

Parameter	Minimum	Maximum	Unit	Note
Average received power	-1.5		dBm	
Optical Modulation Amplitude (OMA) (Informative)	0.053	mW		a
Return loss of receiver	12		dB	
Stressed receiver sensitivity (OMA)	0.078	mW		a, b
Stressed receiver ISI test	1.50		dB	2
Stressed receiver DCD component of DJ (at Tx)	16.3		ps	

a. Optical modulation amplitude values are peak-to-peak.

b. Stressed Rx entries for Sensitivity and ISI differ for each fiber type. The entries shown are for 50 μm, 500 MHz.km fiber

9.5.15 1x QDR LINKS

C9-13.2.3: All 1x-QDR-SX Optical Ports shall comply with the Eye Safety, Transmitter, and Receiver requirements in [Section 9.5.15.1](#) and [Section 9.5.15.2](#)

C9-13.2.4: All 1x-QDR-LX Optical Ports shall comply with the Eye Safety, Transmitter, and Receiver requirements in [Section 9.5.15.1](#) and [Section 9.5.15.3](#)

A 1x-QDR-SX link operates in the 850 nm wavelength band using multimode (MM) fiber. The optical parameters have been selected so as to allow typical 1x-SX Optical Transceivers to use a GaAs Vertical Cavity Surface Emitting Laser (VCSEL) and a photodetector, trans-impedance pre-amplifier and limiting post-amplifier. Three fiber types are specified for 1x-QDR-SX:

- i) 1x-SX/50 link: 500 MHz.km 50 μm / 125 μm MM fiber
- ii) 1x-SX/50 link: 2000 MHz.km 50 μm / 125 μm MM fiber
- iii) 1x-SX/62 link: 200 MHz.km 62.5 μm / 125 μm MM fiber

A 1x-QDR-LX link operates in the 1300 nm wavelength band using singlemode (SM) fiber. The optical parameters have been selected to allow a typical 1x-LX Optical Transceiver to use an uncooled InP-based Fabry-Perot (FP) laser, Distributed Feedback (DFB) laser or VCSEL laser, and an InGaAs photodetector, trans-impedance pre-amplifier and limiting post-amplifier. A trade-off curve is provided in [Figure 171](#) between Center Wavelength and RMS Spectral Width to allow designers to select appropriate laser technology while still ensuring link operation. Only one fiber type is specified for 1x-LX: single-mode non-dispersion shifted.

9.5.15.1 1x QDR EYE SAFETY

C9-13.2.5: The optical power coupled into the fiber for 1x-QDR-SX and 1x-QDR-LX shall be limited to a maximum value with Class I laser safety operation in accordance with CDRH and IEC 60825-1 Radiation Safety of Laser Products: Equipment Classification, Requirements and User Guide.

9.5.15.2 1x-QDR-SX OPTICAL PARAMETERS

The 1x QDR SX Optical Port shall meet the Optical Parameters as defined for 10GBASE-SR in IEEE Std 802.3, Clause 52. 1x QDR SX shall use the same transmit and receive optical interface and budget specifications and measurements, with the following notes and exceptions:

- The nominal signaling speed for 1x-QDR-SX shall be 10.0 GBd ± 100 ppm

9.5.15.3 1x-QDR-LX OPTICAL PARAMETERS

The 1x LX QDR Optical Port shall meet the Optical Parameters as defined for 10GBASE-LR in IEEE Std 802.3, Clause 52. 1x LX QDR shall use the same transmit and receive optical interface and budget specifications and measurements, with the following notes and exceptions:

- The nominal signaling speed for 1x-QDR-LX shall be 10.0 GBd ± 100 ppm

9.6 OPTICAL RECEPTACLE AND CONNECTOR

The primary function of the optical fiber connector specification is to define mechanical alignment of the optical fibers to the optical port of an Optical Transceiver.

The objective of this section is to specify the optical interface sufficiently to ensure the following:

- a) Intermateability
- b) Mechanical/Optical Performance
- c) Maximum Supplier Flexibility

In this section, only the dimensions necessary to specify the transmitter-receiver center-center distance of the 12x-SX connector are provided. All other dimensions are included by reference to other standards.

C9-14: All Optical Transceivers shall present the Optical Receptacle specified in [Optical Receptacle and Connector \(Section 9.6\)](#) through the system bulkhead. Typically, this bulkhead will be a connector housing.

Optical Receptacles, Optical Connectors, and Fibre Optic Adapters shall be as specified in this section.

The InfiniBand connector for 1x, 1x DDR and 1x QDR links shall be a duplex LC connector, for 4x SX and 4x SX DDR links shall be an MPO connector, for 4x LX links shall

be a dual-SC connector and for 12x, 8x SX DDR and 12x SX DDR links shall be a dual-MPO connector.

9.6.1 1x CONNECTOR - LC

The connectors, adapters and receptacles described here shall be fully duplex, creating fully bi-directional optical connection with one mating action. [Figure 184](#) and [Figure 185](#) show outline drawings of the Duplex LC Optical Connector and Duplex LC Fiber Optic Adapter respectively.

The 1x Optical Connector defined by this specification shall conform to ANSI/TIA/EIA-604-10 (FOCIS 10), Fiber Optic Connector Intermateability Standard, Type "LC".

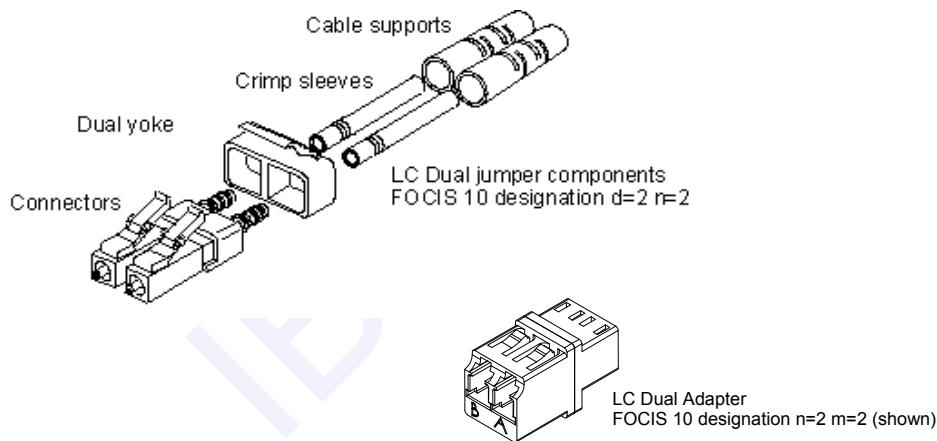


Figure 184 Dual LC Plug

9.6.1.1 1X FIBER OPTIC CONNECTOR

C9-15: All 1x Optical Cable Connectors shall comply with connector specifications of [Section 9.6.1.1](#).

The InfiniBand connector for 1x links shall be a duplex LC connector.

The 1x Optical Connector on each end of the Fiber Optic Cable shall conform to:

- i) ANSI/TIA/EIA-604-10 (FOCIS 10), Fiber Optic Connector Intermateability Standard, Type "LC", and
- ii) Fibre Channel Physical Interface standard (FC-PI), revision 8.0

The implementation of connectors or adapters compliant with the FOCIS 10 standard shall not preclude the intermateability of connectors, adapters or receptacles compliant with the FOCIS 10A standard.

9.6.1.2 1x FIBER OPTIC RECEPTACLE

C9-16: All 1x Optical Port Receptacles shall comply with receptacle and fiber orientation specifications of [Section 9.6.1.2](#) and [Section 9.6.1.3](#).

The 1x (SX & LX) Optical Transceiver Port shall be a Type "LC" receptacle as defined by ANSI/TIA/EIA-604-10 (FOCIS 10), Fiber Optic Connector Intermateability Standard.

It may contain resilient sleeves to optically align the connector plug ferrules. The positioning of the ferrule endfaces to optimize optical coupling can be accomplished in a variety of ways not described here.

The implementation of connectors, adapters or receptacles compliant with the FOCIS 10 standard shall not preclude the intermateability of connectors, adapters or receptacles compliant with the FOCIS 10A standard.

9.6.1.3 1x FIBER OPTIC RECEPTACLE ORIENTATION

1x-SX and 1x-LX Fiber Optic Transceivers shall follow the Transmit/Receive convention detailed in [Figure 185](#). If the Optical Connector is orientated such that the keying features on the LC housings are at the top, then looking into the Fiber Optic Receptacle the fiber on the left shall be used for optical transmit and the fiber on the right shall be used for optical receive.

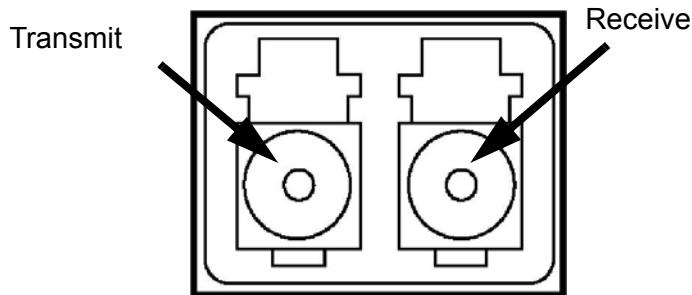


Figure 185 1x-SX and 1x-LX Optical Receptacle orientation looking into the Optical Transceiver

9.6.1.4 1x COLOR

C9-17: The 1x-SX multimode Optical Connector, Adapter and Receptacle, or a visible portion of these, shall be beige in color.

C9-18: The 1x-LX singlemode Optical Connector, Adapter and Receptacle, or a visible portion of these, shall be blue in color.

9.6.2 4x-SX CONNECTOR - SINGLE MPO

The connectors, adapters and receptacles described here shall be fully duplex, creating fully bi-directional optical connection with one mating action.

9.6.2.1 4x-SX FIBER OPTICAL CONNECTOR

C9-19: All 4x SX Optical Cable Connectors shall comply with fiber optical connector specifications of [Section 9.6.2.1](#).

The 4x-SX Optical Connector on each end of the Fiber Optic Cable shall consist of a female MPO plug. The female MPO plug shall conform to IEC 1754-7-4, Push/Pull MPO Female Plug Connector Interface, and shall contain a female MT ferrule.

A female MT ferrule is similar to a male MT ferrule, except that the female ferrule does not have alignment pins. Instead, alignment holes are provided, which accept the alignment pins of the corresponding male MT ferrule. In this manner, precision alignment is achieved between the ferrule in an Optical Receptacle and the ferrule in an Optical Connector.

[Figure 186](#) shows an outline drawing of a typical MPO connector with push-pull coupling mechanism. The MPO housing surrounds a rectangular male MT ferrule. The male MT ferrule has positions for 12 fibers. The male MT ferrule is typically 6.4 mm by 2.5 mm and contains two precision alignment pins of 0.7 mm diameter.

9.6.2.2 4x-SX FIBER OPTICAL RECEPTACLE

C9-20: All 4x SX Optical Port Receptacles shall comply with receptacle and fiber optical orientation specifications of [Section 9.6.2.2](#) and [Section 9.6.2.3](#).

The 4x-SX Optical Transceiver Port shall have a 4x-SX Optical Receptacle which shall be a male MPO receptacle. The male MPO receptacle shall have two fixed pins conforming to IEC 1754-7-5, and shall conform to IEC 1754-7-3, Push/Pull MPO Adapter Interface standard.

Note: In the U.S.A., the MPO connector is also known as the MTP connector.

9.6.2.3 4x-SX FIBER OPTIC RECEPTACLE ORIENTATION

4x-SX Optical Transceivers shall follow the Transmit/Receive convention detailed in [Figure 187](#). If the Optical Connector is orientated such that the keying feature on the MPO housing is at the top, then looking into the Fiber Optic Receptacle fibers are numbered left-to-right as 0-through-11. The 4 fiber positions on the left (fibers 0, 1, 2, 3) shall

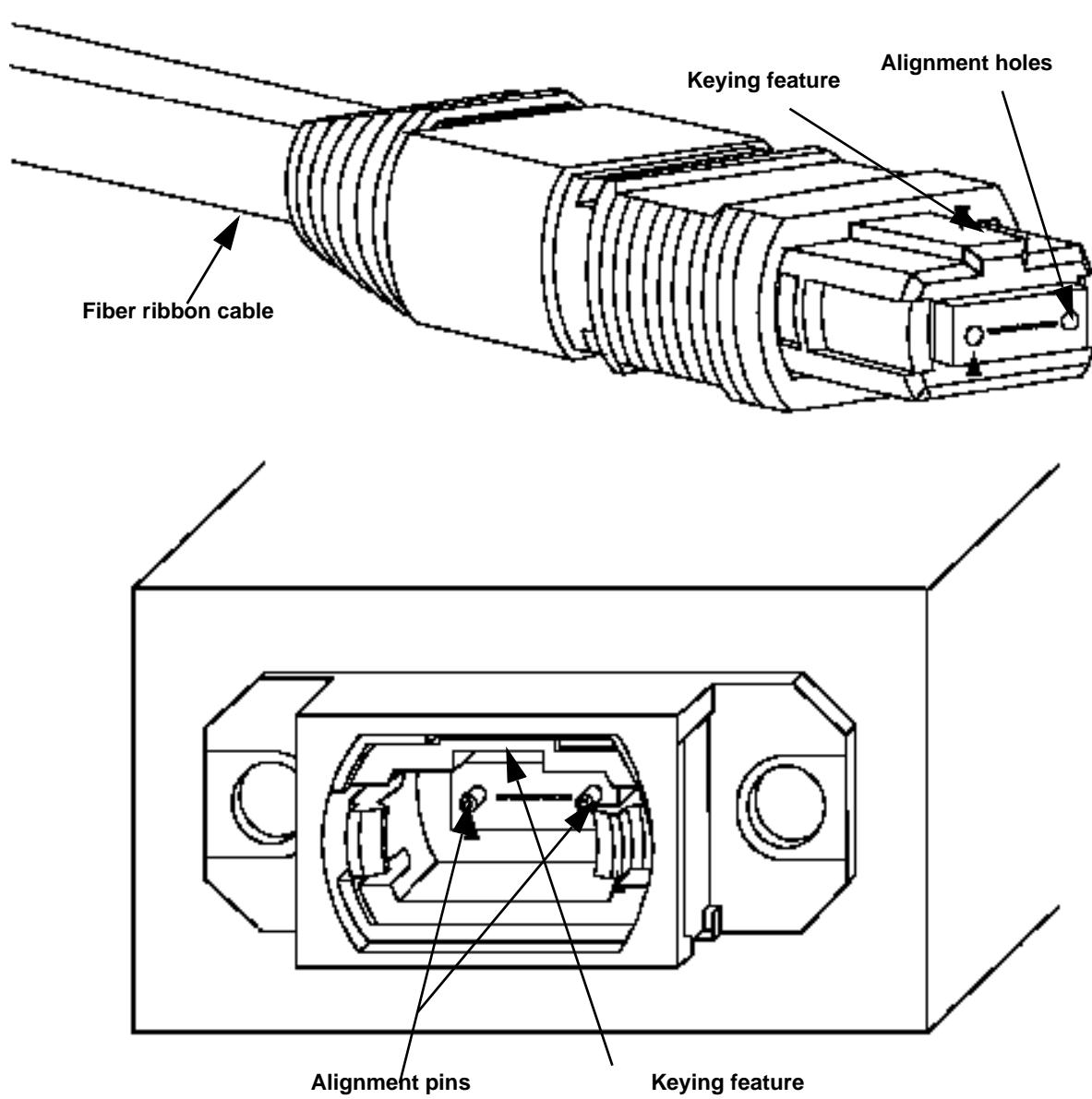


Figure 186 MPO Plug and Receptacle

be used for optical transmit and the 4 fiber positions on the right (fibers 8, 9, 10, 11) shall be used for optical receive. Fibers 0, 1, 2, 3 shall carry transmit Lanes 0, 1, 2, 3 respectively. Fibers 8, 9, 10, 11 shall carry receive Lanes 3, 2, 1, 0 respectively.

The central four fibers (fibers 4, 5, 6, 7) may be physically present. If one or more of the central four fibers is present, then it shall not be used to carry IB signals.

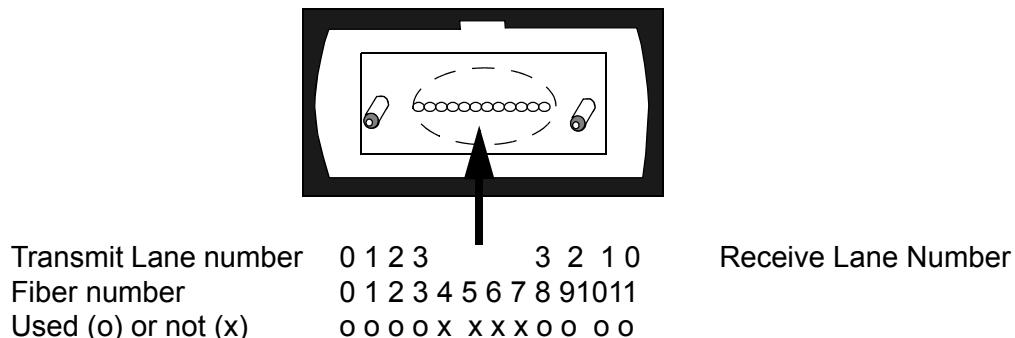


Figure 187 4x-SX Optical Receptacle orientation looking into the Transceiver

9.6.3 4x LX CONNECTOR - SC

The connectors, adapters and receptacles described here shall be fully duplex, creating fully bi-directional optical connection with one mating action.

9.6.3.1 4x LX FIBER OPTIC CONNECTOR

C9-20.1.1: All 1x Optical Cable Connectors shall comply with connector specifications of [Section 9.6.3.1](#).

The InfiniBand connector for 4x LX links shall be a duplex SC connector.

The 4x LX Optical Connector on each end of the Fiber Optic Cable shall conform to the requirements of IEC 61754-4.

Only the Floating Duplex style Connector Plug shall be used. Rigid SC Duplex connector shall not be used.

9.6.3.2 4x LX FIBER OPTIC RECEPTACLE

C9-20.1.2: All 1x Optical Port Receptacles shall comply with receptacle and fiber orientation specifications of [Section 9.6.3.2](#) and [Section 9.6.3.3](#).

The 4x LX Optical Transceiver Port shall be a Type "SC" receptacle as defined by IEC 61754-4-5.

9.6.3.3 4x LX FIBER OPTIC RECEPTACLE ORIENTATION

4x LX Fiber Optic Transceivers shall follow the Transmit/Receive convention detailed in [Figure 188](#). If the Optical Connector is orientated such that the keying features on the SC housings are at the Bottom, then looking into the Fiber Optic Receptacle the fiber on the

right shall be used for optical transmit and the fiber on the left shall be used for optical receive.

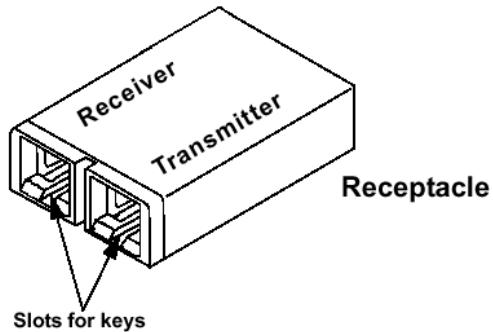


Figure 188 4x-LX Optical Receptacle orientation

9.6.4 8x-SX OPTICAL RECEPTACLE AND CONNECTOR

C9-20.2.1: All 8x SX Optical Cable Connectors and 8x Optical Port Receptacles shall comply with the 12x-SX connector and receptacle specifications in [Section 9.6.4](#) and [Section 9.6.5](#) with the exception of the following exception for Fiber Optic Receptacle Orientation:

8x-SX Optical Transceivers shall follow the transmit/receive convention detailed in [Figure 189](#). If the Optical Connector is orientated such that the keying features on the MPO housings are at the top, then looking into the Optical Receptacle the fibers are numbered left-to-right as 0-through-11 for the left-hand section and 0-through-11 for the right hand section. The 8 fiber positions on the left shall be used for optical transmit and the 8 fiber positions on the right shall be used for optical receive. On the left, fibers 0, 1, 2, 3, 4, 5, 6, 7 shall carry transmit Lanes 0, 1, 2, 3, 4, 5, 6, 7 respectively. On the right, fibers 4, 5, 6, 7, 8, 9, 10, 11 shall carry receive Lanes 7, 6, 5, 4, 3, 2, 1, 0 respectively. On the left fibers 8, 9, 10, 11 may be present but shall not be used to carry IB signals. On the right fibers 0, 1, 2, 3 may be present but shall not be used to carry IB signals.

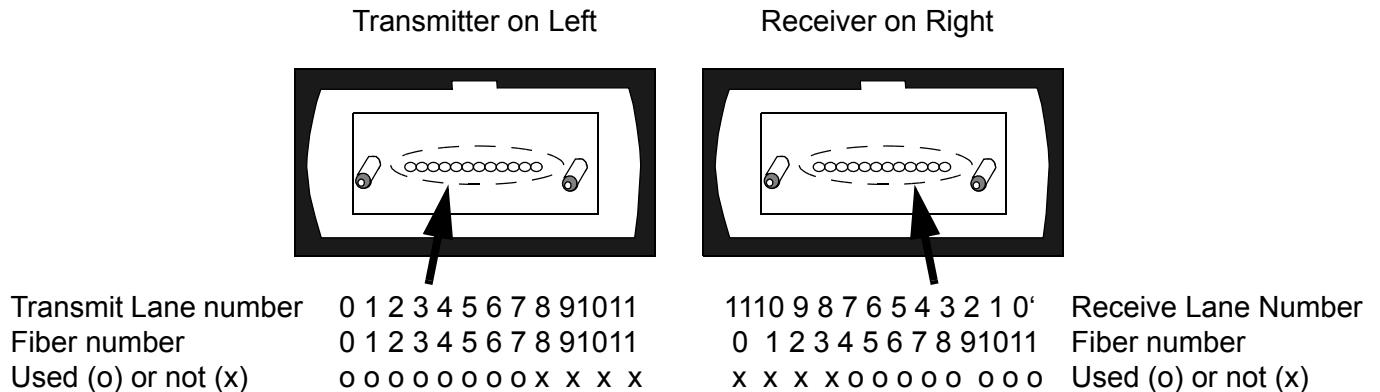


Figure 189 8x-SX Optical Receptacle orientation looking into the Transceiver

9.6.5 12x-SX CONNECTOR - DUAL MPO

The connectors, adapters and receptacles described here shall be fully duplex, creating fully bi-directional optical connection.

9.6.5.1 12x-SX FIBER OPTIC CONNECTOR

C9-21: All 12x-SX Optical Cable Connectors shall comply with connector specifications of [Section 9.6.5.1](#).

The 12x-SX Optical Connector on each end of the Fiber Optic Cable shall consist of a double female MPO plug. Each female plug shall conform to IEC 1754-7-4, Push/Pull MPO Female Plug Connector Interface, and shall contain a female MT ferrule.

9.6.5.2 12x-SX FIBER OPTIC RECEPTACLE

C9-22: This compliance statement is obsolete and has been replaced by [C9-22.2.1](#):

C9-22.2.1: All 12x-SX Optical Port Receptacles shall comply with receptacle and fiber optical orientation specifications of [Section 9.6.5.2](#) and [Section 9.6.2.3](#)

The 12x-SX Optical Transceiver Port shall have a 12x-SX Optical Receptacle which shall consist of a double male MPO receptacle. Each male MPO receptacle shall be as described in [Section 9.6.2](#), shall have two fixed pins conforming to IEC 1754-7-5, and shall conform to IEC 1754-7-3, Push/Pull MPO Adapter Interface standard.

A 12x Optical Transceiver may consist of physically separate Transmit and Receive Modules or a single physical Transceiver.

If a 12x solution is based on separate transmit and receive modules, then these modules shall support a centerline to centerline receptacle spacing of 20.0 mm +/-0.5 mm as defined in [Figure 190](#).

If a 12x solution is based on a single transceiver module, then these modules shall support a centerline to centerline receptacle spacing of 16.0 mm +/- 0.1 mm as defined in [Figure 190](#).

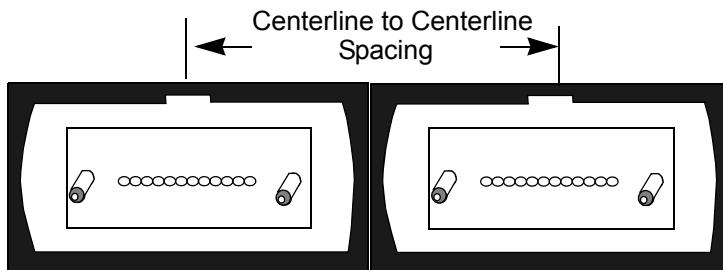


Figure 190 Double MPO Optical Receptacle Configuration

Implementation Note

To simplify cable management, in typical installations a duplex Fiber Optic Cable will be mated to a 12x-SX Optical Transceiver in a single mating action. This reduces the probability of mis-configuration. A plastic clip would serve to join two individual MPO housings together to form a double-MPO connector. Alternatively, a common housing could be moulded to house two sets of MPO actuation mechanisms plus two MT ferrules that float independently. In all cases, the two MT ferrules would float independently of each other to some extent.

The Optical Transceiver within the IB module could be constructed either as one transceiver component. Alternatively it could be constructed as separate Optical Transmitter and Optical Receiver components mounted to the board independently, but mutually aligned to within the total travel of the two MT ferrules.

9.6.5.3 12x-SX FIBER OPTIC RECEPTACLE ORIENTATION

12x-SX Optical Transceivers shall follow the transmit/receive convention detailed in [Figure 191](#). If the Optical Connector is orientated such that the keying features on the MPO housings are at the top, then looking into the Optical Receptacle the fibers are numbered left-to-right as 0-through-11 for the left-hand section and 0-through-11 for the right hand section. The 12 fiber positions on the left shall be used for optical transmit and the 12 fiber positions on the right shall be used for optical receive. On the left, fibers 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 shall carry transmit Lanes 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 respectively. On the right, fibers 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 shall carry receive Lanes 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 respectively.

	Transmitter on Left	Receiver on Right	
Transmit Lane number	0 1 2 3 4 5 6 7 8 9 10 11	11 10 9 8 7 6 5 4 3 2 1 0'	Receive Lane Number
Fiber number	0 1 2 3 4 5 6 7 8 9 10 11	0 1 2 3 4 5 6 7 8 9 10 11	Fiber number
Used (o) or not (x)	o o o o o o o o o o o o	o o o o o o o o o o o o	Used (o) or not (x)

Figure 191 12x-SX and 8x-SX Optical Receptacle orientation looking into the Transceiver

9.7 FIBER OPTIC CABLE PLANT SPECIFICATIONS

9.7.1 OPTICAL FIBER SPECIFICATION

C9-23: This compliance statement is obsolete and has been replaced by [C9-23.2.1](#):

C9-23.2.1: All Optical Cables shall comply with the fiber cable requirements in [Section 9.7.1](#) and [Section 9.7.2](#) for respective Fiber Modes (Multimode or Single mode) as defined in [Table 182](#).

C9-23.2.2: All 8x-SX Optical Ports shall meet all the optical cable plant specifications for 12x-SX in [Section 9.7](#) with the exception that 8x-SX has only eight receive and transmit lanes.

Fiber Optic Cables for links in the -SX class shall use either 50/125 μm optical fiber or 62.5/125 μm multimode optical fiber compliant respectively with the “MMF 50/125 um” or “MMF 62.5/125 μm ” specification in this section.

Fiber Optic Cables for links in the -LX class shall use non-dispersion-shifted single mode fiber compliant with the “SMF” specification in this section.

SMF shall conform to TIA/EIA-492CAAA-98 “Dispersion-Unshifted Single-Mode Optical Fibers”.

MMF 500 MHz.km 50/125 μm shall conform to TIA/EIA-492AAAB-98 “Detail Specification for 50- μm Core Diameter/125- μm Cladding Diameter Class Ia Graded-Index Multi-mode Optical Fibers” or IEC 60793-2 Type A1a.

MMF 2000 MHz.km 50/125 μm shall conform to TIA/EIA-492AAC. TIA/EIA-492AAC is presently in ballot.

MMF 62.5/125 μm shall conform to TIA/EIA-492AAAA-A-97 “Detail Specification for 62.5- μm Core Diameter/125- μm Cladding Diameter Class Ia Graded-Index Multimode Optical Fibers” or IEC 60793-2 Type A1b.

Recommendation to Optical System Designer

For new -SX installations, MMF 50/125 μm fiber **should** be used because of the increased range of 1x-SX/50 links compared to 1x-SX/62 links.

4x-SX, 8x-SX and 12x-SX Fiber Optic Cables may use ribbonized fibers.

SMF, MMF 50 $\mu\text{m}/125 \mu\text{m}$ and MMF 62.5 $\mu\text{m}/125 \mu\text{m}$ fibers shall also conform to the respective columns of [Table 182](#).

9.7.2 MODAL BANDWIDTH

The Modal Bandwidth for the 500 MHz.km and 200 MHz.km cables with Overfilled Launch specified in [Table 182](#) is the worst case modal bandwidth, measured according to the methods of TIA2.2.1 working specification TIA/EIA-455-204-FOTP204 Measurement method for Multimode Fiber Bandwidth. The 2000 MHz.km cable is measured using TIA/EIA-492AAC. Worst case modal bandwidth is defined as the lowest bandwidth that can occur in a fiber under reasonable launch conditions.

In practice, worst case modal bandwidth is used to account for differences in multimode fiber bandwidth that can occur under restricted launch conditions relative to bandwidth observed using an overfilled launch condition. The Optical System Penalty limits presented in [Section 9.5.8](#), [Section 9.5.10](#) and [Section 9.5.13](#) represent conservative calculations based on an overfilled launch model. In practice, better link performance is typically expected.

Table 182 Optical Fiber Specifications

Parameter	SMF (9 μm)	MMF 50/125 μm	MMF 62.5/125 μm	Unit
Nominal Fiber Specification Wavelength	1310	850	850	nm
Fiber Cable Attenuation (Max)	0.5	3.5	3.5	dB/km
Modal Bandwidth with Overfilled Launch (Min)	not applicable	500 ^a and 2000 ^b	200	MHz.km
Zero Dispersion Wavelength λ_0	1300 $\leq \lambda_0 \leq$ 1320	1295 $\leq \lambda_0 \leq$ 1320	1320 $\leq \lambda_0 \leq$ 1365	nm
		0.11 for $1300 \leq \lambda_0 \leq 1320$ and $95 \leq \lambda_0 \leq 1300$	0.11 for $1320 \leq \lambda_0 \leq 1348$ and $48 \leq \lambda_0 \leq 1365$	ps/nm ² .km
Zero Dispersion Slope S_0 (Max)	0.093			

a. Overfilled launch bandwidth per IEC 60793-1-41 or TIA/EIA-455-204.

b. Effective modal bandwidth for fiber meeting TIA /EIA-492AAC

9.7.3 OPTICAL PASSIVE LOSS OF FIBER OPTIC CABLE

C9-24: This compliance statement is obsolete and has been replaced by [C9-24.1.1](#):

C9-24.1.1: All optical cables shall comply with the passive loss specifications of [Section 9.7.3](#)

Optical Passive Loss shall be not exceed than the values specified in [Table 158](#), [Table 161](#), [Table 170](#) and [Table 176](#) for 1x-SX, 1x-LX, 4x-SX and 12x-SX links respectively. Optical Passive Loss shall not exceed the values specified in [Table 164](#), [Table 167](#), [Table](#)

[173](#), [Table 179](#), and [Table 179](#) for 1x-DDR SX, 1x-DDR LX, 4x-DDR SX, 8x-DDR-SX and 12x-DDR-SX links respectively. The loss of the fiber plant shall be verified by the methods of OFSTP-14A. The Optical Passive Loss of a Fiber Optic Cable is the sum of attenuation losses due to the fiber, Fiber Optic Adapters and splices.

Connection Insertion Loss for 4x-LX shall not exceed the values specified for 10GBASE-L optical cables in IEEE Std 802.3, Clause 52. The loss of the fiber plant shall be verified by the methods defined in IEEE Std 802.3, Clause 52. The Connection Insertion Loss for 4x-LX of a Fiber Optic Cable is the sum of attenuation losses due to the fiber, Fiber Optic Adapters and splices.

A Fiber Optic Cable may contain one or more Fiber Optic Adapters and/or splices, provided that the total Optical Passive Loss conforms to the optical budget of this specification. In calculating worst case operating range values for [Table 158](#), [Table 161](#), [Table 164](#), [Table 167](#), [Table 170](#), [Table 173](#) and [Table 176](#), a total loss budget of 1.5dB was assigned to Fiber Optic Adapters and splices. For 4x-LX, 10G Ethernet assumed a budget of 2.0dB for Fiber Optic Adapters and splices.

9.7.4 FIBER OPTIC ADAPTERS AND SPLICES

9.7.4.1 1x (SX & LX) FIBER OPTIC ADAPTERS AND SPLICES

C9-25: All 1x (SX & LX) Optical Cable adaptors and splices shall comply with [Section 9.7.4.1](#)

If the 1x Fiber Optic Cable consists of more than one Fiber Optic Segment, then the Fiber Optic Adapter used to join the Fiber Optic Segments shall conform to ANSI/TIA/EIA-604-10 (FOCIS 10), Fiber Optic Connector Intermateability Standard, Type "LC".

Fiber Optic Adapters and splices for 1x-SX (multimode) links shall have a return loss of 20dB minimum as measured by the methods of FOTP-107 or equivalent. Fiber Optic Adapters and splices for 1x-LX (single-mode) links shall have a return loss of 26dB minimum as measured by the methods of FOTP-107 or equivalent.

9.7.4.2 4x-SX FIBER OPTIC ADAPTERS AND SPLICES

If the 4x -SX Fiber Optic Cable consists of more than one Fiber Optic Segment, then Optical Adapters are used to join the Fiber Optic Segments. These Optical Adapters are not detailed in this specification.

C9-26: Fiber Optic Adapters and splices for 4x-SX links shall have a return loss of 20dB minimum as measured by the methods of FOTP-107 or equivalent.

9.7.4.3 4x-LX FIBER OPTIC ADAPTERS AND SPLICES

If the 4x -LX Fiber Optic Cable consists of more than one Fiber Optic Segment, then Optical Adapters are used to join the Fiber Optic Segments. These Optical Adapters are not detailed in this specification.

C9-26.1.1: Fiber Optic Adapters and splices for 4x-LX links shall have a return loss of 26dB minimum as measured by the methods of FOTP-107 or equivalent.

9.7.4.4 12x-SX FIBER OPTIC ADAPTERS AND SPLICES

If the 12x -SX Fiber Optic Cable consists of more than one Fiber Optic Segment, then Optical Adapters are used to join the Fiber Optic Segments. These Optical Adapters are not detailed in this specification.

C9-27: Fiber Optic Adapters and splices for 12-SX links shall have a return loss of 20dB minimum as measured by the methods of FOTP-107 or equivalent.

9.8 SIGNAL CONDITIONER IN OPTICAL TRANSCEIVER

C9-28: All Optical Ports shall comply with the Signal Conditioner requirements in [Section 9.8](#).

C9-28.2.1: All 8x SX Signal conditioners shall meet all the signal conditioner specifications for 12x-SX in [Section 9.8](#) with the exception that 8x SX has only eight receive and transmit lanes.

9.8.1 MOTIVATION FOR SIGNAL CONDITIONER

At the time of preparing this specification, optical functionality is relatively expensive compared to electronic functionality. The overall system cost is generally minimized by using optical components (lasers, fiber optic cables and photoreceivers) that have as low performance as can be tolerated, while including relatively sophisticated signal conditioning and control functionality in the electrical domain in the optical transmitter and optical receiver. Hence, optical links in general, including those specified for InfiniBand, are generally designed to operate with higher jitter than the electrical interconnects found on printed circuit boards.

The optical jitter specification of [Section 9.5.5](#) is relaxed from the jitter specification for high-speed electrical signaling specified in [Chapter 6: High Speed Electrical Interfaces](#). However, the electrical inputs (Optical Transmitter side) and electrical outputs (Optical Receiver side) of an Optical Transceiver shall meet the high-speed signaling specification of [Chapter 6: High Speed Electrical Interfaces](#).

These high speed electrical signaling jitter limits should be compared to the optical jitter specification of [Table 153](#) for the different links widths and distance classes.

If an optical transceiver does not perform signal conditioning, and an IB electrical transceiver driver output is connected directly through a short board trace to the intermediate electrical interface TP1 in an optical transceiver, then the recommended input jitter specification will be exceeded at TP1. The use of appropriate pre-emphasis can reduce deterministic jitter (DJ), making it more likely that the system will meet the mandatory jitter compliance at TP2 and TP33.

There is a related problem on the optical receiver side, assuming that the jitter at the intermediate electrical interface TP4 meets the recommended typical value. The typical TP4 jitter exceeds the specified input jitter tolerance of an IB Electrical Transceiver. Hence the system may not work correctly.

9.8.2 SIGNAL CONDITIONER IMPLEMENTATION

This section describes possible implementations of the Optical Transceiver to include the signal conditioning functionality.

[Figure 192](#), [Figure 193](#), [Figure 194](#), [Figure 195](#) and [Figure 196](#) show typical IB Optical Transceiver implementations for 1x, 1x-DDR-SX, 1x-QDR-SX, 4x-SX, 4x-DDR-SX, 4x-LX, 8x-DDR-SX, 12x-SX, and 12x-DDR-SX respectively. In each case an optional Signal Conditioner is shown. This Signal Conditioner shall be present in an Optical Transceiver if the optoelectronic components do not directly provide IB-compliant high-speed electrical signals.

The function of the Signal Conditioner is to convert between IB-compliant high-speed electrical signaling and the intermediate electrical interface of the optoelectronic components. The Signal Conditioner **should** be implemented as a repeater, IB-compliant Retiming Repeater, or an adaptive equalizer.

Implementation Note

The system designer may choose to integrate the Signal Conditioner with the optoelectronic components, or may choose to install it separately on the Board.

Retiming functionality generally dissipates considerable power. For thermal considerations it may be advantageous to physically separate the signal conditioner from the optoelectronic components. This is especially true for 4x-SX and 12-SX links. In this case the intermediate electrical interface requires careful design to ensure signal integrity.

Implementation Note

In some implementations it may be expedient for an Optical Transceiver not to present an IB-compliant electrical interface. For instance, the Signal Conditioner functionality could be collapsed within the SerDes block of a Switch chip to reduce system power and cost and to improve functional density. In this case the “Optical Transceiver” no longer presents an IB-compliant high-speed electrical signaling interface. Nonetheless the TP2 and TP3 optical specifications can still be met. Hence the overall IB Module and the optical ports may still be able to claim IB-compliance, since the non-compliant electrical interface is hidden within the IB Module.

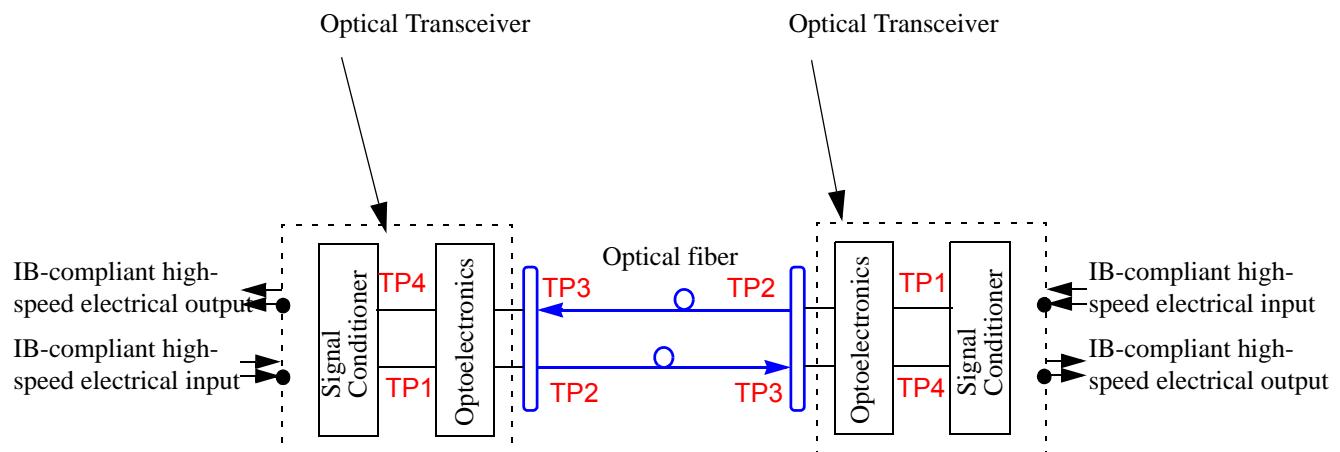


Figure 192 Recommended 1x, 1x DDR, and 1x QDR Optical Link Implementation

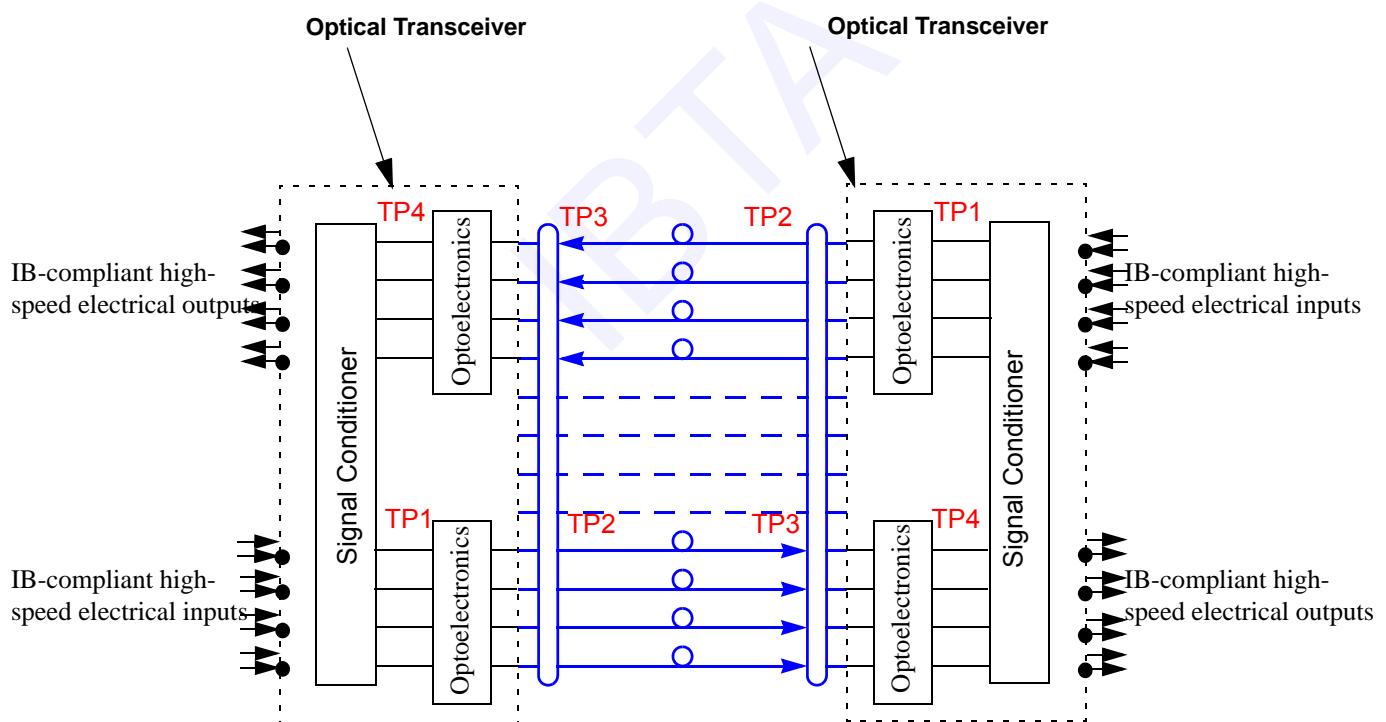


Figure 193 Recommended 4x-SX & 4x-DDR-SX Optical Link Implementation

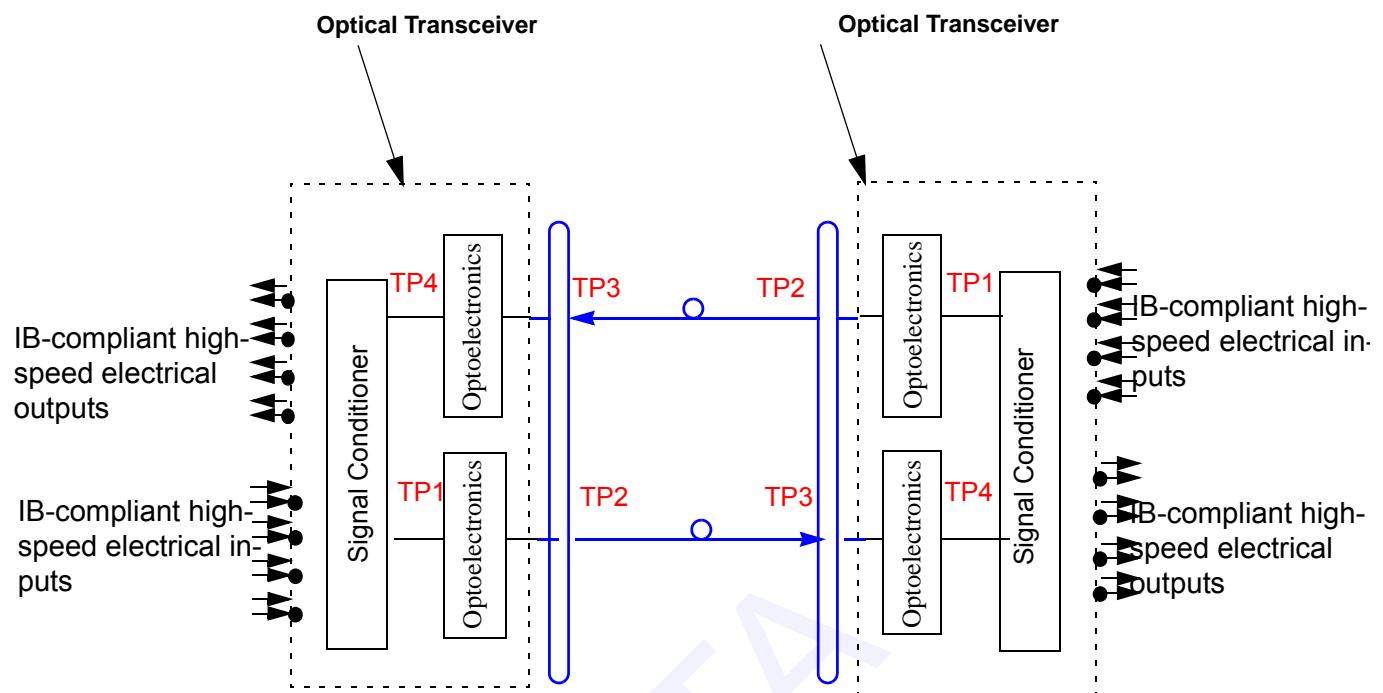


Figure 194 Recommended 4x-LX Optical Link Implementation

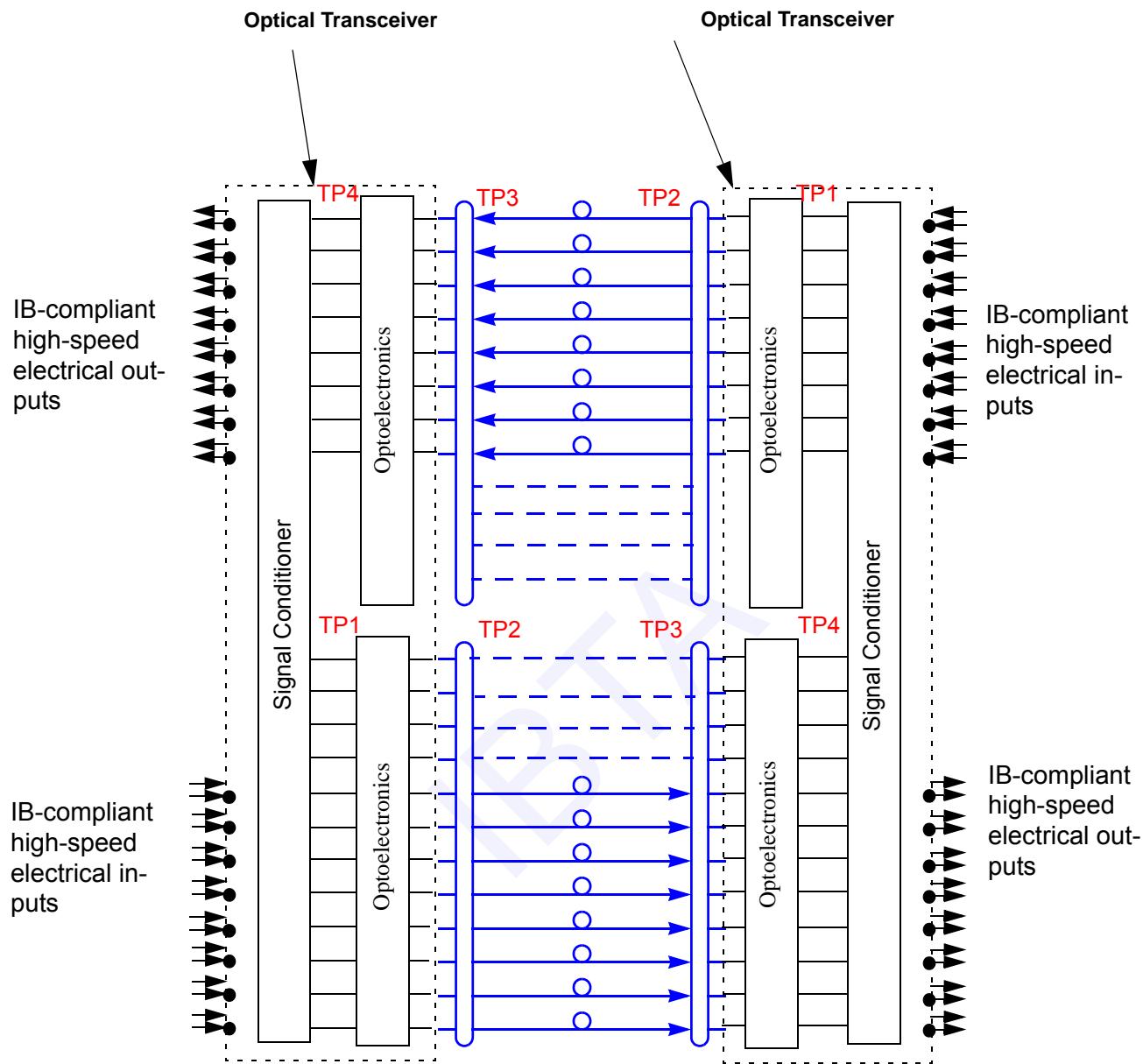


Figure 195 Recommended 8x-SX and 8x-DDR-SX Optical Link Implementation

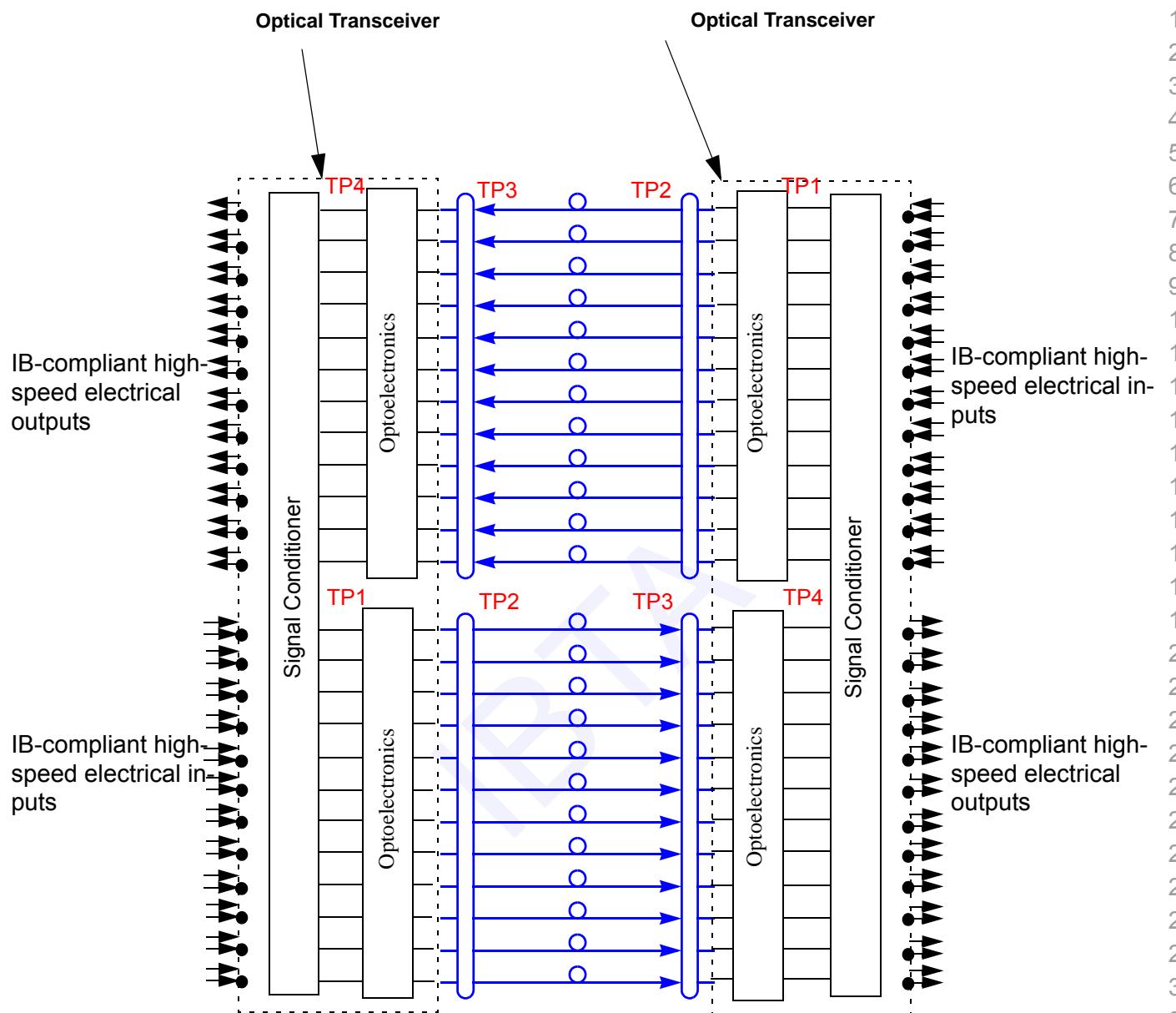


Figure 196 Recommended 12x-SX & 12x-DDR-SX Optical Link Implementation

9.9 AUX POWER

C9-29: All Optical Ports shall comply with Aux power behavior in [Section 9.9](#)

C9-29.2.1: All 8x-SX Optical Ports shall meet all the Aux power specifications for 12x-SX in [Section 9.9](#) with the exception that 8x SX has only eight receive and transmit lanes.

9.9.1 BEHAVIOR IN AUX POWER MODE

To facilitate compliance with the overall Auxiliary power budget, an Optical Transceiver **should not** draw more than 150 mW in Aux power mode. This limit depends on how much other functionality the IB Module is required to provide in Aux power mode. More than 150 mW of Aux power may be available to an Optical Transceiver in certain implementations.

The high-speed electrical outputs of an Optical Receiver operating under Aux power shall be squelched to less than V_{RSD} (Signal Threshold) as defined in [Chapter 6: High Speed Electrical Interfaces](#).

9.9.2 BEACONING AND WAKE-UP

An Optical Receiver operating only under Aux power shall detect the presence of a beaconing sequence (refer to the appropriate receiver specifications in [Section 9.5](#) for the minimum valid optical levels) on the Fiber Optic Cable and trigger the appropriate response. See and [Chapter 5: Link/Phy Interface](#).

9.10 OPTICAL PLUGGABLE MODULES

9.10.1 1X OPTICAL PLUGGABLE MODULES

C9-29.1.1: All 1x SX and 1x LX Optical Pluggable Devices shall comply with [Section 9.10.1](#).

Any 1x Optical Pluggable module shall comply with electrical and mechanical 1x pluggable requirements in and [Section 7.2, "1X Interface," on page 355](#), and the respective 1x optical distance classifications (SX or LX) in [Section 9.4.2](#), [Section 9.5](#), [Section 9.6.1](#), [Section 9.7](#), and [Section 9.9](#)

9.10.2 4X OPTICAL PLUGGABLE MODULES

C9-29.1.2: All 4x SX and 4x LX Optical Pluggable Modules: shall comply with [Section 9.10.2](#).

Any 4x Optical Pluggable module shall comply with electrical and mechanical 4x pluggable requirements in and [Section 7.5, "4X QSFP+ Interface connectors," on page 373](#), and the respective optical distance classifications (SX or LX) in [Section 9.4.2](#), [Section 9.5](#), [Section 9.6.1](#), [Section 9.7](#), and [Section 9.9](#)

9.10.3 8X AND 12X OPTICAL PLUGGABLE MODULES

C9-29.1.3: All 8x or 12x SX and 8x or 12x LX Optical Pluggable Modules shall comply with [Section 9.10.3](#).

Any 8x or 12x Optical Pluggable module shall comply with electrical and mechanical 8x and 12x pluggable requirements in and [Section 7.8, “CXP Interface,” on page 413](#), and the respective optical distance classifications (SX or LX) in [Section 9.4.2](#), [Section 9.5](#), [Section 9.6.1](#), [Section 9.7](#), and [Section 9.9](#)

IBTA

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

ANNEX A1: FDR AND EDR COMPLIANCE BOARDS AND TEST SETUPS

A1.1 BACKGROUND

Compliance boards are used in the measurement of device and cable electrical characteristics. These include a Host Compliance Board (HCB) and a Module Compliance Board (MCB). Many of the electrical specification values are set based on the compliance board characteristics since they are measured using those boards without de-embedding. Therefore, boards with different characteristics may be used if the measured parameter values are corrected to compensate for the differences between the board used and those defined herein. The usage model for the boards is described in the following sections.

A1.1.1 COMPLIANCE BOARD DESIGN

Although the physical size and trace lengths of the Host Compliance Boards described herein vary between the QSFP+ and CXP designs, different board materials and trace geometries are used to ensure that the electrical characteristics are nearly identical. This allows the use of common channel and test specifications for the two interfaces since de-embedding of the board insertion losses is not used in the test process as had been done at QDR and slower speeds.

A1.1.2 COMPLIANCE BOARD CALIBRATION

The compliance boards are calibrated by mating one HCB with one MCB and measuring the various electrical parameters on the combined set.

A1.2 TEST SETUPS

This section provides a high level summary description of test setups for transmitters, receivers, and cables. More complete and detailed description of test setup and procedures are described in MOI (Method of Implementation) documents, available from the CIWG (Compliance and Interoperability Working Group) of the InfiniBand Trade Association.

A1.2.1 TRANSMITTER CHARACTERIZATION

Host board transmitter outputs are characterized using a Host Compliance Board mated to host board port on the device under test (DUT) as shown in [Figure 197](#). The host transmitter will be required to be set to use specific Preset states for some of these measure-

ments, either under the control of an InfiniBand data source generating TS-T ordered sets or through vendor specific means of transmitter configuration.

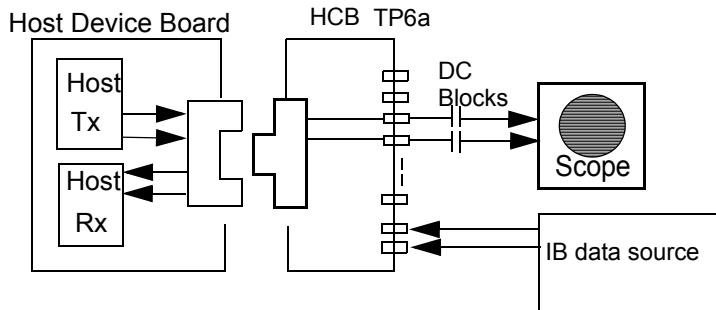


Figure 197 Host transmitter output characterization setup using HCB

A1.2.2 RECEIVER CHARACTERIZATION

Host board receiver inputs are characterized using a Host Compliance Board mated to the device under test (DUT) as shown in [Figure 198](#), using the method described in IEEE 802.3-2015 Annex 86A.

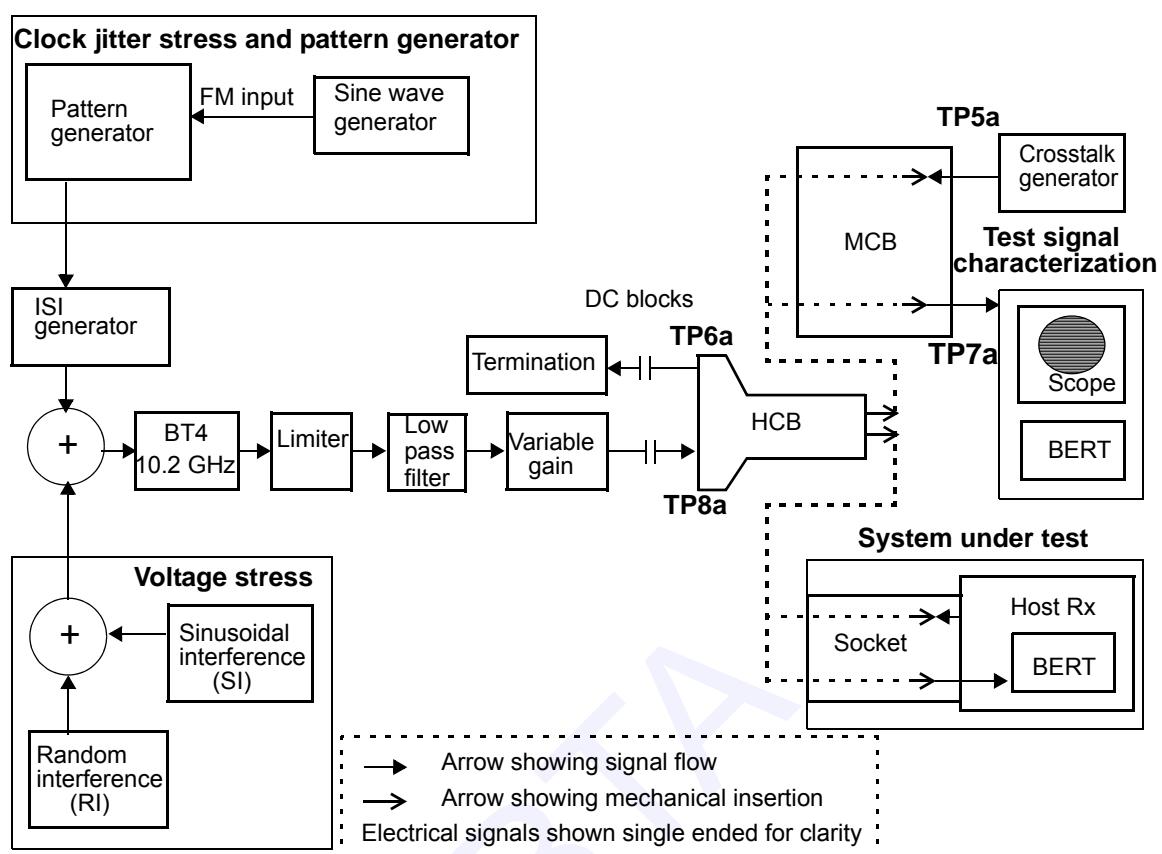


Figure 198 Host receiver input characterization setup

A1.2.3 CABLE CHARACTERIZATION

Cables are characterized using two Module Compliance Boards, one on each end of the cable under test (CUT) as shown in [Figure 199 on page 617](#), using the method described in IEEE 802.3-2015 Annex 86A. Note that the MCBS supply power to the connected cables and provide access to the Management Interface. Passive cable testing requires only the use of the vector network analyzer.

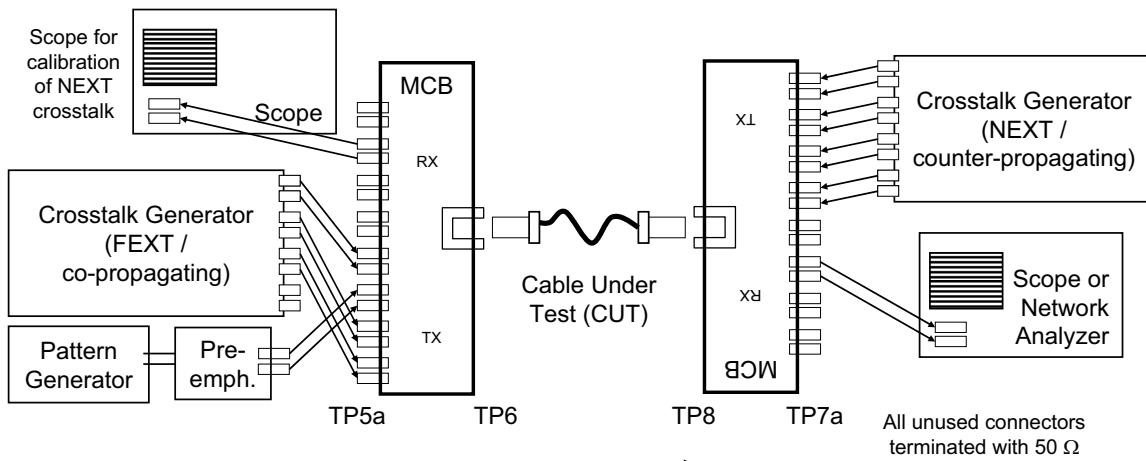


Figure 199 Cable characterization setup using MCB

A1.3 COMPLIANCE BOARDS - ELECTRICAL SPECIFICATIONS

The compliance boards described herein are designed for a specific physical interface (QSFP or CXP). Users are free to design their own boards if desired and should attempt to meet the electrical performance characteristics described here. In the event that there are differences or corrections, they shall be documented by the user to the Compliance and Interoperability Group (CIWG) upon request, in order to demonstrate compliance.

[Table 183](#) lists the various trace pair responses at 7.03125 and 12.891 GHz for reference. These data include estimated variations due to board process variation, and are specified at the Nyquist frequencies of both FDR and EDR data rates. Note that the response for the individual HCB and MCB do not include the contribution of the module connector, however, the mated HCB+MCB response does include the contribution of the mated module connector. All listed values include SMA connector and trace losses.

Architecture Notes

Note that the values of the parameters in [Table 183](#) in Rel. 1.3 of this specification are subject to modification.

It is the intention of the InfiniBand Trade Association that the specification parameters for both module and host compliance boards should match the specifications used by other groups (e.g., IEEE and OIF) defining interoperability standards in the 25 - 28 Gb/s range. At time of writing of Rel. 1.3, host compliance boards with improved loss parameters (Target S_{DD21} , HCB = $2*(0.001-0.096\sqrt{f})-0.046f$, equivalent to -1.15 dB at 7.03125 GHz and -1.87 dB at 12.890625 GHz) were anticipated but not yet conclusively demonstrated.

Updates may appear in MOI documents available from the Compliance and Interoperability Working Group.

Table 183 Compliance board response

Board	Target response at f (GHz)		Maximum response at f		Minimum response at f		Unit	Comment
	7.03125 GHz	12.891 GHz	7.03125 GHz	12.891 GHz	7.03125 GHz	12.891 GHz		
S _{DD21} , HCB QSFP or CXP	-1.15	-1.87					dB	see Eq. 27
S _{DD21} , MCB QSFP or CXP	-0.72	-1.17					dB	see Eq. 28
S _{DD21} , Mated HCB + MCB, QSFP and CXP			-1.62	-2.87	-2.93	-4.67	dB	see Eq. 29 and Eq. 30 , includes DUT connector loss
S _{DD11} , S _{DD22} , Mated HCB+MCB QSFP and CXP			-14.48	-11.55			dB	see Eq. 31 , includes DUT connector loss
S _{DC11} , Mated HCB + MCB, QSFP and CXP			-24.98	-20.79			dB	see Eq. 32
S _{CD21} , Mated HCB + MCB, QSFP and CXP			-27.48	-21.21			dB	see Eq. 33
S _{CCxx} , Mated HCB+MCB QSFP and CXP			-3	-3			dB	
	Maximum Value							
ILD _{RMS} , Mated HCB+MCB	0.1						dB	50 MHz-19.5 GHz
Total Integrated crosstalk noise (ICN) voltage	3.9						mV	see Table 184 on page 632 for crosstalk calculation parameters
MDNEXT voltage	1.35						mV rms	
MDFEXT voltage	3.6						mV rms	

A1.3.1 Host COMPLIANCE BOARDS

Host Compliance Boards (HCBs) are designed for testing transmitter output and receiver input characteristics. The reference insertion loss of the QSFP+ and CXP HCBs is defined by [Equation 27](#) and plotted in [Figure 200 on page 620](#). Use of boards with response characteristics that depart from these limits will require adjustment of the test limits. The specific adjustments needed are beyond the scope of this specification.

$$S_{DD21Ref}(f) = 0.002 - 0.192\sqrt{f} - 0.092f, \quad 0.01 \leq f \leq 26 \quad \text{Eq. 27}$$

where f = frequency in GHz

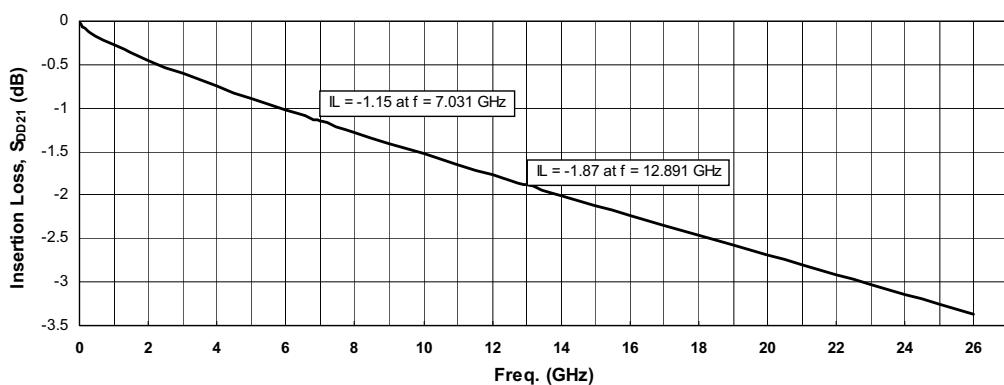


Figure 200 HCB trace pair reference through response

Note that DC blocking capacitors are not utilized in the HCBs. The user is encouraged to use external blocking capacitors to protect the test equipment from damage due to DC offset voltages that may be present during testing.

A1.3.2 MODULE COMPLIANCE BOARDS

Module Compliance Boards (MCBs) are designed to provide an interface for testing cables and modules. MCBs provide power, power supply decoupling, low speed control and management interface connections to the module or cable end. MCBs also provide high-speed signal interface for testing transmitters and receivers with a fixed known passive channel loss, when mated with the appropriate HCB. The reference insertion loss of the MCBs is defined by [Equation 28](#) and plotted in [Figure 201 on page 621](#). Use of boards with insertion loss characteristics that depart from these limits is beyond the scope of this specification and is discouraged.

$$S_{DD21Ref}(f) = 0.00125 - 0.12\sqrt{f} - 0.0575f, \quad 0.01 \leq f \leq 26 \quad \text{Eq. 28}$$

where f = frequency in GHz

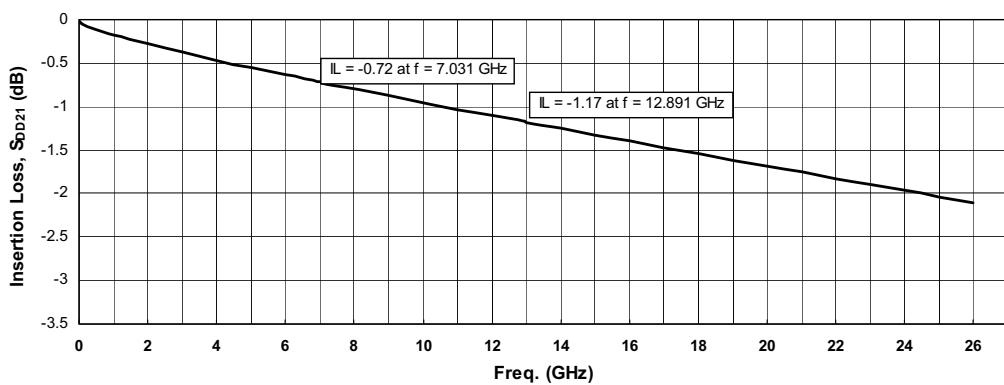


Figure 201 MCB trace pair reference through response

A1.3.3 COMBINED HOST & MODULE COMPLIANCE BOARDS

The through response of the mated QSFP+ HCB and MCB combination is bounded by [Equation 29](#) and [Equation 30 on page 622](#). Use of boards with insertion loss characteristics that depart from these limits will require adjustment of the limits. The specific adjustments needed are beyond the scope of this specification. [Figure 202](#) shows a plot of the reference through response limits for each pair of the HCB mated with the MCB.

$$S_{DD21}(f) < -0.08\sqrt{f} - 0.2f, \quad 0.01 \leq f \leq 26 \quad \text{Eq.29}$$

where f = frequency in GHz

$$S_{DD21}(f) > -0.12 - 0.475\sqrt{f} - 0.221f, \quad 0.01 \leq f \leq 14 \quad \text{Eq. 30}$$

$$S_{DD21}(f) > 4.24 - 0.66f, \quad 14 \leq f \leq 26$$

where f = frequency in GHz

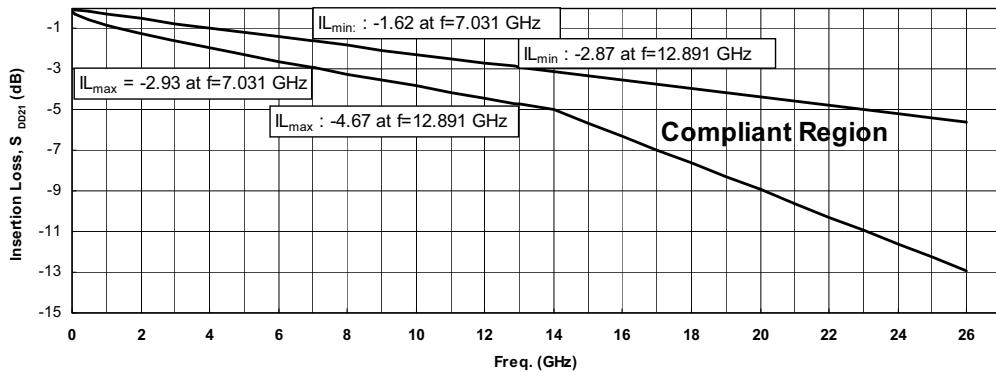


Figure 202 Mated HCB/MCB trace pair through response

The differential return loss of the combined QSFP+ host compliance board and module compliance boards, measured at either board's test interface, shall meet or exceed the

limits in [Eq. 31](#).

$$S_{DD11}, S_{DD22} \geq \begin{cases} -20 + f, & 0.01 \leq f < 4 \\ -18 + 0.5f, & 4 \leq f \leq 28 \end{cases} \quad \text{Eq.31}$$

where f = frequency in GHz

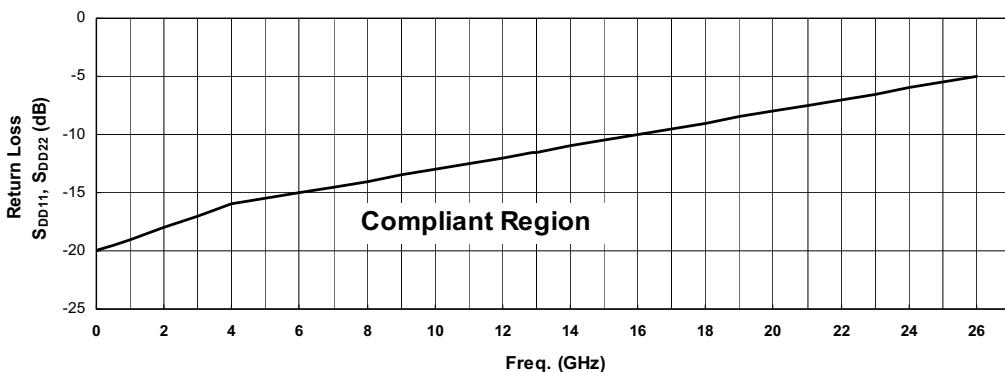


Figure 203 Limits on Return Loss for QSFP+ Host Compliance Boards

The common mode to differential reflection and differential to common mode reflection of the mated HCB/MCB combination, from both ends, including SMA connector and trace loss and DUT connector loss, shall be given by [Equation 32 on page 623](#).

$$S_{DC11}, S_{DC22}, S_{CD11}, S_{CD22} < -30 + (5/7) \cdot f, \quad 0.01 \leq f \leq 14 \quad \text{Eq.32}$$

$$S_{DC11}, S_{DC22}, S_{CD11}, S_{CD22} < -25 + (5/14) \cdot f, \quad 14 \leq f \leq 25$$

where f = frequency in GHz

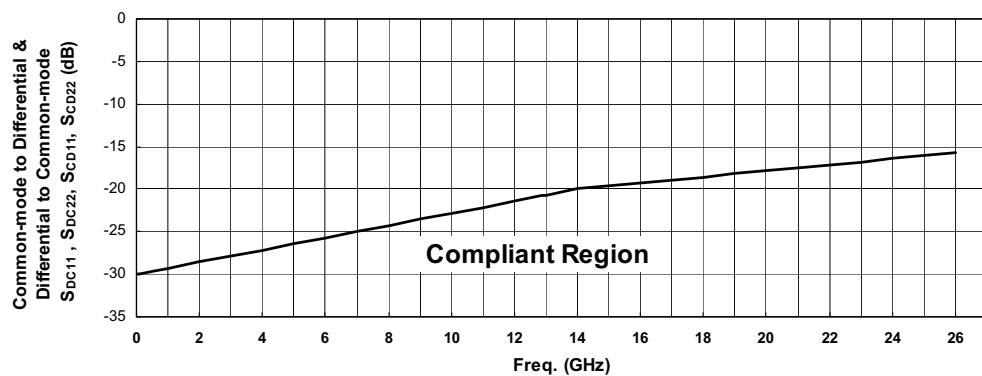


Figure 204 Limits on HCB/MCB common-mode/differential return loss

The differential to common-mode conversion of the mated HCB/MCB combination, including SMA connector and trace loss and DUT connector loss, shall be given by [Equation 33](#).

$$S_{CD21}, S_{CD12} \leq -35 + 1.07f, \quad 0.01 \leq f \leq 14 \\ -20, \quad 14 < f \leq 26 \quad \text{Eq.33}$$

where f = frequency in GHz

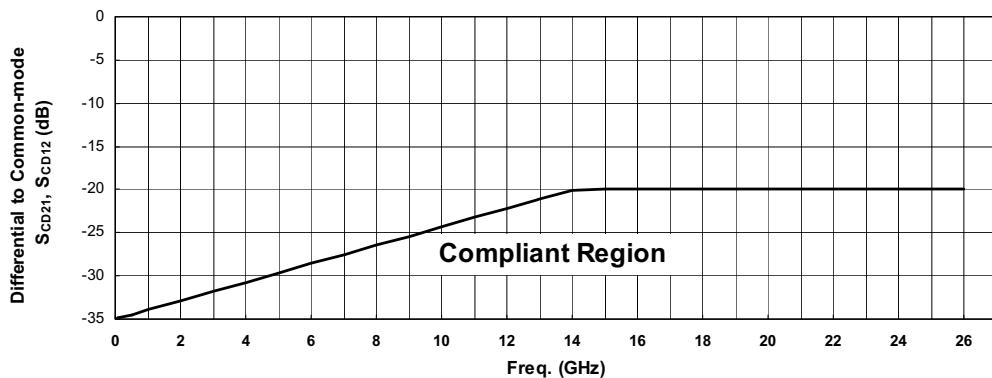


Figure 205 Limits on HCB/MCB differential to common-mode conversion

The common mode reflection (S_{CC11} or S_{CC22}) of the mated HCB/MCB combination, including SMA connector and trace loss and DUT connector loss shall be a maximum of -3 dB from 10 MHz to 25 GHz.

The ILDRMS for the mated HCB and MCB pair, across the frequency range 50 MHz to 19.5 GHz is ≤ 0.1 dB. The Integrated Crosstalk Noise (ICN) shall be less than 3.9 mV. MDNEXT shall be less than 1.35 mV RMS. MDFEXT shall be less than 3.6 mV RMS.

A1.4 QSFP28 COMPLIANCE BOARDS - MECHANICAL DESCRIPTION

A1.4.1 QSFP28 HOST COMPLIANCE BOARD

A sketch view of an exemplary HCB test fixture is shown in [Figure 206 on page 625](#). Other boards may be used with appropriate adjustments to the measured parameters to account for differences in those board designs from those of the one described herein. The plug tongue is plugged into the host port QSFP28 connector for testing. Mechanical support for the card and plug tongue should be employed to prevent board breakage when the board is plugged into the host port.

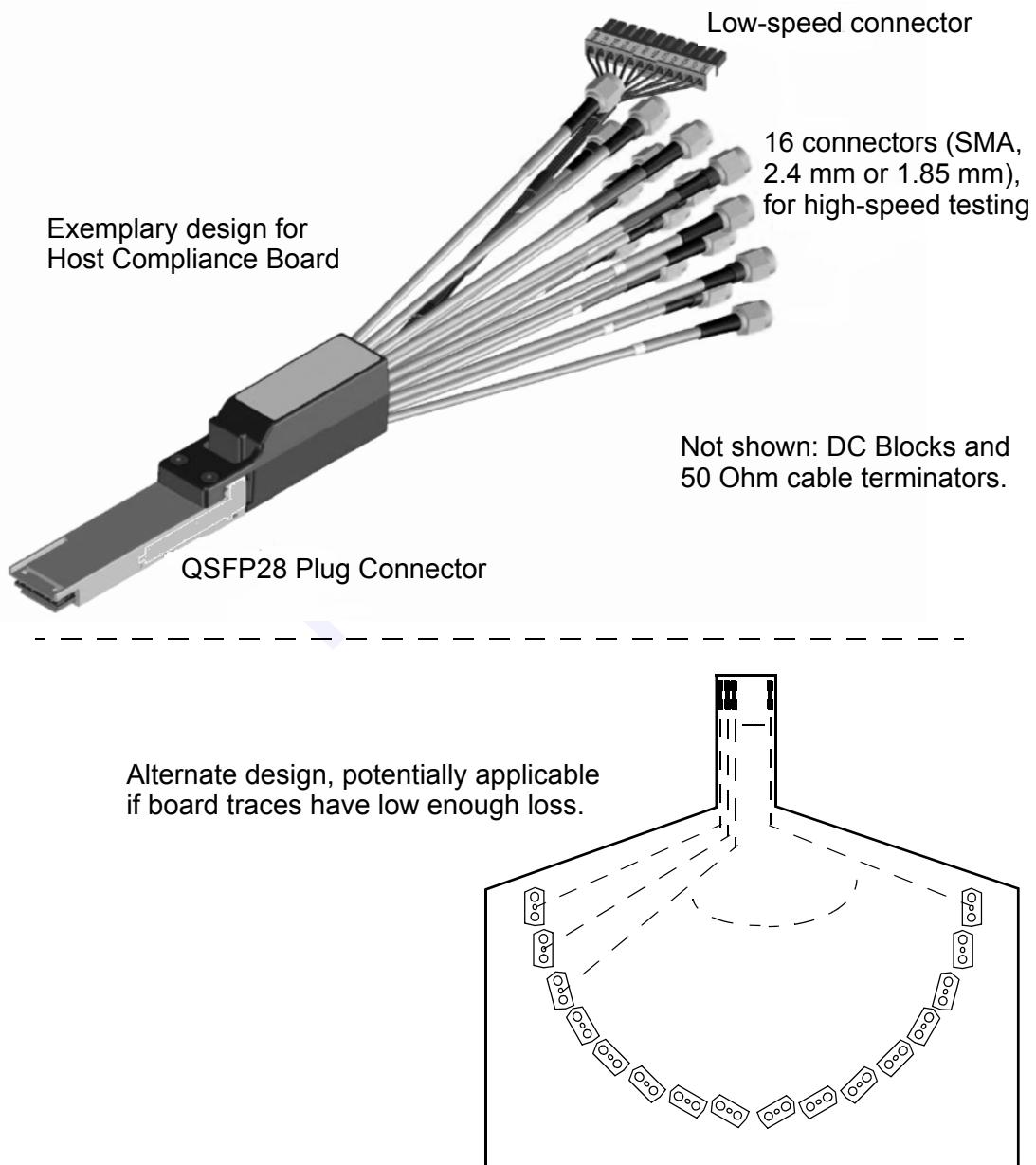


Figure 206 QSFP28 Host Compliance Board layout

A1.4.2 QSFP28 MODULE COMPLIANCE BOARD

A sketch view of an exemplary QSFP28 MCB is shown in [Figure 207 on page 626](#).

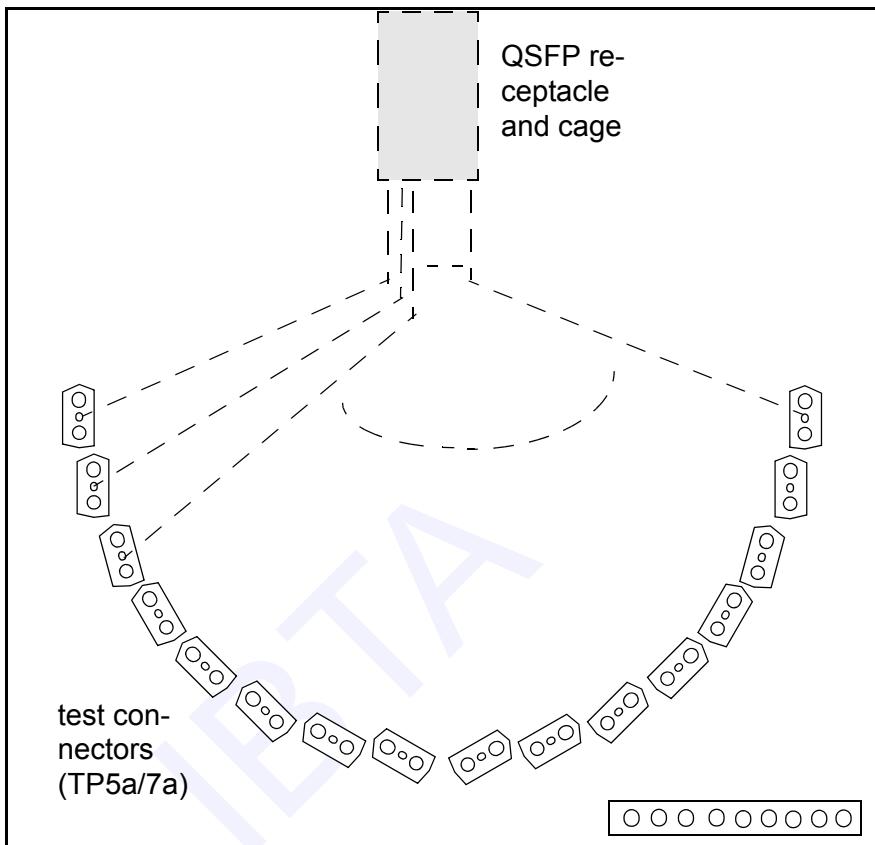


Figure 207 QSFP28 Module Compliance Board layout

A1.5 CXP COMPLIANCE BOARDS - MECHANICAL DESCRIPTION

A1.5.1 CXP HOST COMPLIANCE BOARD

A sketch view of an exemplary board is shown in [Figure 208 on page 627](#) and [Figure 209 on page 628](#). Other boards may be used with appropriate adjustments to the compliance parameters to account for differences in those board designs from those of the one described herein. The plug tongue is plugged into the host port CXP connector for testing. Mechanical support for the card and plug tongue should be employed to prevent board breakage when the board is plugged into the host port.

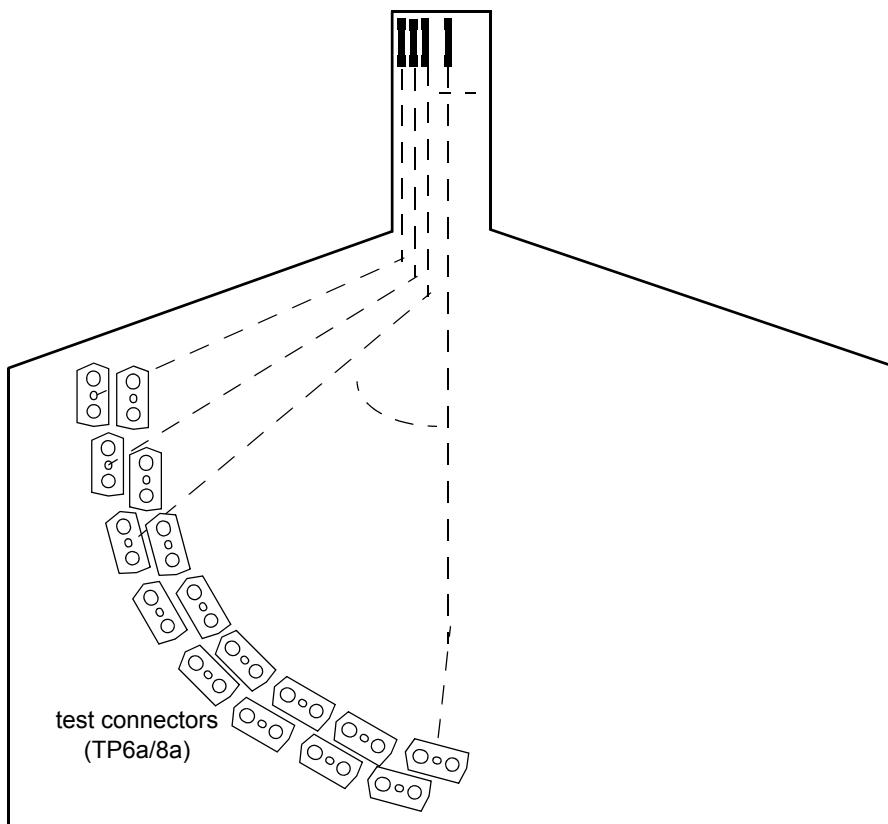


Figure 208 CXP Host Compliance Board layout, side A

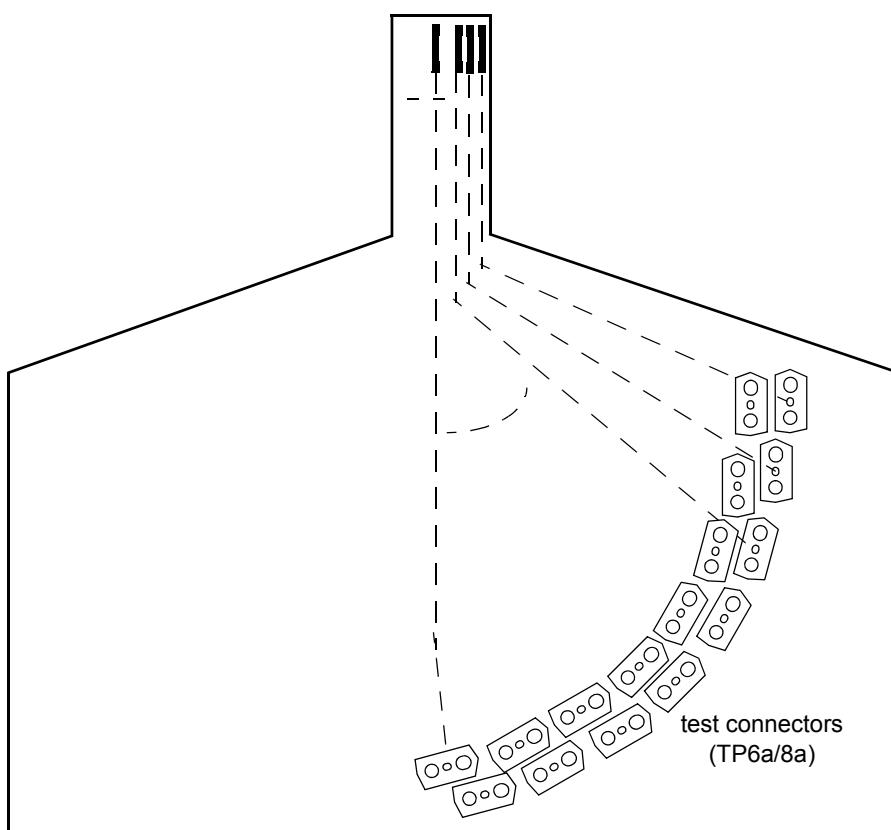


Figure 209 CXP Host Compliance Board layout, side B

A1.5.2 CXP MODULE COMPLIANCE BOARD

A sketch view of the layout of an exemplary CXP MCB is shown in [Figure 210 on page 629](#).

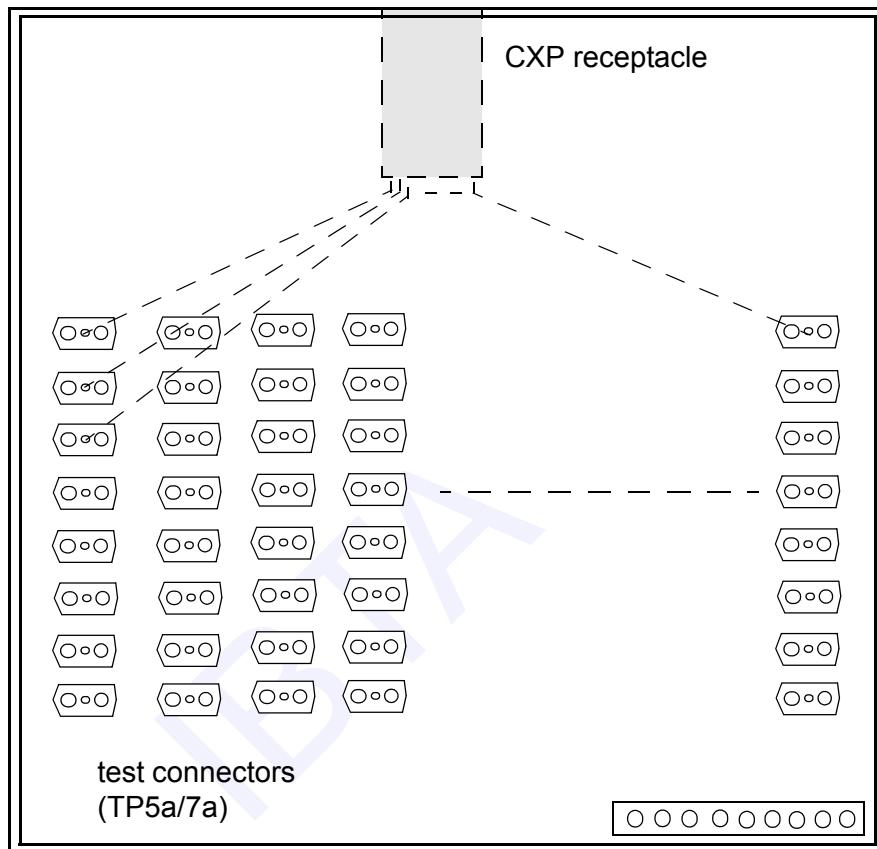


Figure 210 CXP Module Compliance Board layout

ANNEX A2: CABLE ELECTRICAL PARAMETERS FOR FDR AND EDR**A2.1 INSERTION LOSS FITTING**

The limits on cable assembly insertion loss for FDR and EDR interfaces is not a simple limit curve as was used at the lower InfiniBand data rates. The measured data are curve fitted and the coefficients of the fitted curve are compared to the limits specified in [Section 6.8.7 on page 337](#) to determine compliance. This is the method described in clause 12-2 of OIF CEI-03.0 shall be used, with the limits scaled according to the difference in data rates. The fitted insertion loss is given by [Eq. 13 on page 337](#).

The cable insertion loss between test points TP5A and TP7a is measured at frequencies f_n from 50 MHz to 14.1 GHz (FDR) or 19.5 GHz (EDR), at 10 MHz intervals, and the frequency matrix F in [Eq. 34](#) is populated with the values. The quantity “ $\text{mag}(IL_f)$ ” is the magnitude of the measured insertion loss at each frequency point [$\text{mag}(IL_{fx}) = 10^{-(IL_{fx}/20)}$], and is a real number between 0 and 1.

$$F = \begin{bmatrix} \text{mag}(IL_{f_1}) & \text{mag}(IL_{f_1}) \times \sqrt{\frac{f_1}{f_b}} & \text{mag}(IL_{f_1}) \times \frac{f_1}{f_b} & \text{mag}(IL_{f_1}) \times \left(\frac{f_1}{f_b}\right)^2 \\ \text{mag}(IL_{f_2}) & \text{mag}(IL_{f_2}) \times \sqrt{\frac{f_2}{f_b}} & \text{mag}(IL_{f_2}) \times \frac{f_2}{f_b} & \text{mag}(IL_{f_2}) \times \left(\frac{f_2}{f_b}\right)^2 \\ \dots & \dots & \dots & \dots \\ \text{mag}(IL_{f_N}) & \text{mag}(IL_{f_N}) \times \sqrt{\frac{f_N}{f_b}} & \text{mag}(IL_{f_N}) \times \frac{f_N}{f_b} & \text{mag}(IL_{f_N}) \times \left(\frac{f_N}{f_b}\right)^2 \end{bmatrix} \quad \text{Eq.34}$$

The coefficients a_i of the fitted insertion loss equation are then calculated using [Eq. 35 on page 630](#). In that equation, T represents the transpose operator and IL is a column vector of the measured insertion loss values at the frequencies f . The maximum and minimum values for the polynomial coefficients a_i are listed in [Table 83. “FDR compliant linear cable specifications.” on page 338](#), or [Table 86. “EDR compliant linear cable specifications.” on page 345](#), as appropriate. If any of the values of the polynomial coefficients exceed the maximum or are less than the minimum, then the iterative fitting method described in clause 12.2 of OIF-CEI-0.30 shall be used.

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_4 \end{bmatrix} = (F^T F)^{-1} F^T [\text{mag}(IL_f) \times IL_f] \quad \text{Eq. 35}$$

A2.2 INTEGRATED CROSSTALK NOISE (ICN)

In order to limit the signal to noise ratio at the receiver input to an acceptable value, a maximum limit is specified on the integrated crosstalk noise. The ICN is calculated from the combination of the measured insertion loss, MDNEXT, and MDFEXT using [Eq. 36 on page 631](#) through [Eq. 40 on page 632](#) respectively.

First, the rms crosstalk noise is calculated at the output of a specified receive filter with a particular transmitter waveform and the measured crosstalk data. Weighting functions defined in [Eq. 36](#) and [Eq. 37](#) are applied to the multiple disturber crosstalk data,

$$W_{nt}(f_n) = \left(A_{nt}^2 / f_b \right) \text{sinc}^2(f_n/f_b) \left[\frac{1}{1 + (f_n/f_{nt})^4} \right] \left[\frac{1}{1 + (f_n/f_r)^8} \right] \quad \text{Eq. 36}$$

$$W_{ft}(f_n) = \left(A_{ft}^2 / f_b \right) \text{sinc}^2(f_n/f_b) \left[\frac{1}{1 + (f_n/f_{ft})^4} \right] \left[\frac{1}{1 + (f_n/f_r)^8} \right] \quad \text{Eq. 37}$$

where $\text{sinc}(x)=\sin(\pi x)/(\pi x)$, f_n is the frequency in MHz, and f_b is the signaling frequency. The rise/falltime equivalent frequency f_{nt} of the near-end aggressors is inversely proportional to the 20% and 80% rise time T_{nt} such that $f_{nt} * T_{nt} = 0.2365$ with f_{nt} in Hz and T_{nt} in seconds. The rise/falltime equivalent frequency f_{ft} of the far-end aggressors is calculated the same way. The frequency f_r is the 3 dB reference receiver bandwidth which is $0.75*f_b$ GHz. The amplitude of the aggressors A_{nt} and A_{ft} are specified in [Table 184 on page 632](#). The maximum ICN limits are listed in [Table 83 FDR compliant linear cable specifications on page 338](#) and [Table 86, “EDR compliant linear cable specifications,” on page 345](#).

The calculation parameters for ICN are listed in [Table 184](#).

Table 184 ICN Calculation Parameters for FDR and EDR cables

Symbol	Parameter	Specification Value	Unit	Comment
A _t	Victim peak differential output amplitude	400	mV	
A _{nt}	Near end aggressor peak differential output amplitude	600	mV	
A _{ft}	Far end aggressor peak differential output amplitude	600	mV	
T _{nt}	Near end aggressor rise time	24 - FDR 13 - EDR	ps	20-80%
T _{ft}	Far end aggressor rise time	24 - FDR 13 - EDR	ps	20-80%

The quantity MDNEXT, as defined in [Equation 18 on page 340](#), in units of dB, is measured at multiple frequencies spanning the range $0.05\text{GHz} \leq f \leq f_b$, and the near end integrated crosstalk noise N_{nx} is calculated as a weighted sum across frequencies n , as shown in [Eq. 38](#).

$$N_{nx} = \left(2\Delta f \sum_n W_{nt}(f_n) 10^{\frac{MDNEXT(f_n)/10}{10}} \right)^{1/2} \quad \text{Eq. 38}$$

Similarly, the quantity MDFEXT, as defined in [Equation 19 on page 340](#), in units of dB, is measured at multiple frequencies spanning the range $0.05\text{GHz} \leq f \leq f_b$, and the far end integrated crosstalk noise N_{fx} is calculated as a weighted sum across frequencies n , as shown in [Eq. 39](#).

$$N_{fx} = \left(2\Delta f \sum_n W_{ft}(f_n) 10^{\frac{MDFEXT(f_n)/10}{10}} \right)^{1/2} \quad \text{Eq. 39}$$

The total ICN N_x is then calculated using [Eq. 40](#)

$$N_x = \sqrt{N_{nx}^2 + N_{fx}^2} \quad \text{Eq. 40}$$

A2.3 INTEGRATED COMMON MODE CONVERSION NOISE (ICMCN)

The procedure for calculating ICMCN is analogous to the procedure for calculating Integrated Crosstalk Noise (ICN). The differential to common mode conversion amplitude $S_{CD21}(f_n)$, in units of dB, is measured at n frequencies spanning the range $0.05\text{GHz} \leq f \leq f_b$, at 10 MHz intervals. A weighting function is generated using the signal parameters as shown in [Eq. 41](#),

$$W_{cmt}(f_n) = \left(A_c^2 / f_b \right) \operatorname{sinc}^2(f_n/f_b) \left[\frac{1}{1 + (f_n/f_{nct})^4} \right] \left[\frac{1}{1 + (f_n/f_r)^8} \right] \quad \text{Eq. 41}$$

where f_{nct} is the equivalent frequency of the signal rise/falltime as described in [A2.2 Integrated Crosstalk Noise \(ICN\)](#), f_r is the 3 dB reference receiver bandwidth which is 0.75*f_b GHz, and A_c is the differential amplitude of the signal, 600 mV.

The integrated common mode conversion noise is then calculated as a weighted sum across frequencies n , as shown in [Eq. 42](#).

$$ICMCN = \left(2\Delta f \sum_n W_{cmt}(f_n) 10^{S_{CD21}(f_n)/10} \right)^{1/2} \quad \text{Eq. 42}$$

A2.4 FDR OVERALL LINK BUDGET FOR LINEAR CHANNELS (INFORMATIVE)

Figure 211 shows nominal insertion loss parameters for various elements and combinations of elements for the FDR link budget using linear channels. Some parameters, such as the HCB and MCB target trace insertion losses and the mated MCB+HCB insertion loss, are specified in [Table 183](#). Other parameters, such as the host trace IL and the end-to-end link budgets, are assumed target parameters for FDR systems. All insertion losses are described at 7.0325 GHz. These targets and assumptions were instrumental in developing the specifications for signaling over linear channels described in [Chapter 6: High Speed Electrical Interfaces](#).

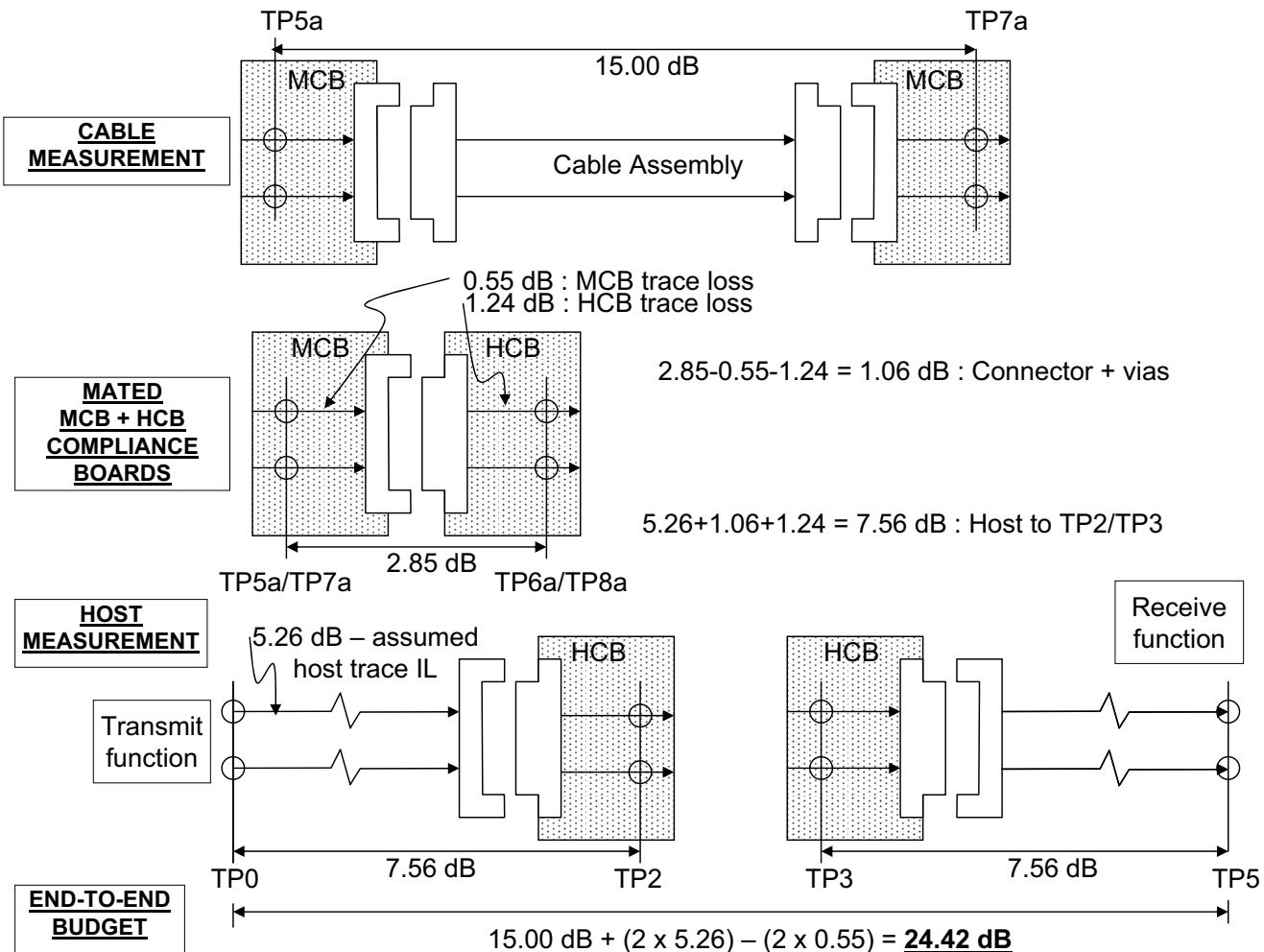


Figure 211 FDR Overall Link Budget (Informative)

A2.5 EDR OVERALL LINK BUDGET FOR LINEAR CHANNELS (INFORMATIVE)

Figure 212 shows nominal insertion loss parameters for various elements and combinations of elements for the EDR link budget using linear channels. Some parameters, such as the HCB and MCB target trace insertion losses and the mated MCB+HCB insertion loss, are specified in [Table 183](#). Other parameters, such as the host trace IL and the end-to-end link budgets, are assumed target parameters for EDR systems. All insertion losses are described at 12.89 GHz, as shown in the figure. These targets and assumptions were instrumental in developing the specifications for signaling over linear channels described in [Chapter 6: High Speed Electrical Interfaces](#).

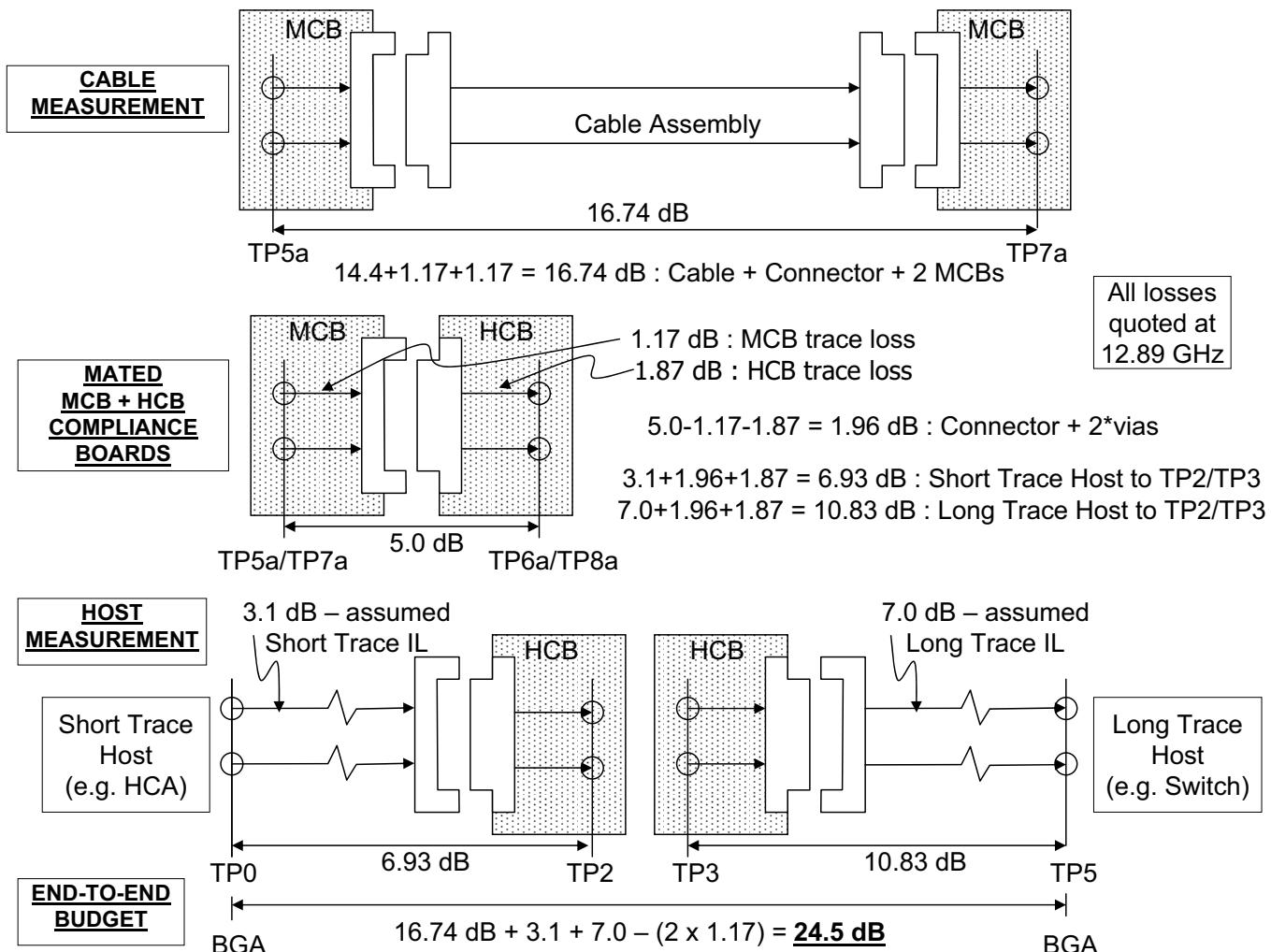


Figure 212 EDR Overall Link Budget (Informative)

ANNEX A3: MANAGEMENT INTERFACE MODIFICATIONS FOR RoCE SUPPORT

A3.1 ROCE MANAGEMENT INTERFACE - OVERVIEW

This document is part of the specification for the InfiniBand Architecture, not for Ethernet. However, the InfiniBand Trade Association has adopted responsibility for managing and promoting the architecture for RoCE (RDMA over Converged Ethernet), which uses IB RDMA protocols over Converged Ethernet switches. In August of 2015, the IBTA published results from the first RDMA over Converged Ethernet plugfest, which tested interoperability between RoCE adapters, cables, and switches. The results of this plugfest made clear that it is worthwhile to highlight the similarities and differences between cables and transceivers designed for Ethernet vs. InfiniBand. This is particularly important for IB-EDR-4X cables vs. Ethernet 100GBase-CR4 / 100GBASE-SR4, because they share a common connector (QSFP28) and common 25G signaling rate, but have some small but significant differences. Note that the IEEE does not specify connectors, and a variety of connectors are used, through multi-source agreements, but the QSFP and CXP form factors are commonly used for copper and short optical cables.

This appendix summarizes the management interface differences between IB-EDR-4x and Ethernet 100G (100GE) cables that use the QSFP+/QSFP28 interface.

A3.2 INFINIBAND VS. ETHERNET MEMORY MAP DIFFERENCES - QSFP/QSFP+

A3.2.1 INFINIBAND VS. ETHERNET MEMORY MAP DIFFERENCES - LOWER PAGE

- Byte 98: CDR Enable
 - InfiniBand: Default value is 0 - CDR disabled, on all channels, for backward compatibility with FDR/QDR.
 - Ethernet: Default value is 1 - CDR enabled on all channels. CDR is assumed as part of the 100GE PMA. Since there is no explicit definition of an initialization flow, CDRs are expected to be enabled for 100GE.
- Byte 113: Far End Implementation, Near End Implementation
 - InfiniBand: reserved Reserved Read-Write
 - Ethernet: Bits 6-4: used to describe Far End (# of connectors, # of channels per connector). Bits 3-0: used to describe Near End (implemented channels).

A3.2.2 INFINIBAND VS. ETHERNET MEMORY MAP DIFFERENCES - UPPER PAGE 00

- Byte 128: Identifier
 - InfiniBand: recommends to use QSFP+ identifier (0Dh) for backward compatibility with 4x FDR and QDR host devices.
 - Ethernet: No explicit definitions. QSFP28 (11h) is typically used for 100GE cables.

- Bytes 131-138: Specification Compliance
 - InfiniBand: Set to 00h for InfiniBand. Electronic compatibility is described in Byte 164.
 - Ethernet: Byte 131 specifies 10/40G/100G Ethernet Compliance Codes. Byte 131: bit7 specifies use of Extended Specification Compliance Codes, in Byte 192.
- Byte 139: Encoding
 - InfiniBand: Set to 00h (Unspecified) for InfiniBand FDR & faster, since must support both 64b/66b & 8b10b.
 - Ethernet: Typically set to 00h, but may be set to value specified in SFF-8024, if transceiver supports a specific encoding.
- Byte 164: Extended module codes for InfiniBand
 - InfiniBand: Set to bit-map of supported rates: EDR, FDR, QDR, DDR, SDR
 - Ethernet: Not used. Set to 00h.
- Byte 192: Extended Specification Compliance Codes
 - InfiniBand: Set to 00h (Unspecified) for InfiniBand products.
 - Ethernet: Defined in SFF-8024, Table 4-4. Examples: 02h: 100GBASE-SR4, 0Bh: 100GBASE-CR4, 18h: 100G AOC or 25GAUI C2M AOC. 19h: 100G ACC or 25GAUI C2M ACC.

A3.2.3 INFINIBAND VS. ETHERNET MEMORY MAP DIFFERENCES - UPPER PAGE 03

- Bytes 224: Tx Input Equalization Magnitude, and Rx Output Emphasis magnitude
 - InfiniBand: Used by host to manage module equalization
 - Ethernet: Not used by host.
- Bytes 234-239: Transmitter Input equalization levels, Receiver Output Emphasis levels, Output amplitude level with no equalization enabled.
 - InfiniBand: Used by host to control module equalization
 - Ethernet: Defined implicitly as part of the host / module interface requirements.

A3.3 INFINIBAND VS. ETHERNET MEMORY MAP DIFFERENCES - CXP/CXP+

A3.3.1 INFINIBAND VS. ETHERNET MEMORY MAP DIFFERENCES - Tx & Rx LOWER PAGE

- Tx Byte 3; Rx Byte 3: Version Control
 - InfiniBand: Version control - used to identify to which version of the CXP specification the cable or optical transceiver is compliant
 - Ethernet: May or may not be identical.
- Tx Byte 41; Rx Byte 42: Tx Rate Select , Rx Rate Select
 - InfiniBand: Allows host to optimize module operation speed: EDR, FDR, QDR, DDR, or SDR.
 - Ethernet: Not specified.

- Tx Byte 43, bit 0; Rx Byte 43, bit 0: CDR Bypass / Enable
 - InfiniBand: Default value is 0 - CDR disabled, on all channels, for backward compatibility with FDR/QDR.
 - Ethernet: Not specified.
- Tx Bytes 62-67; Rx Byte 43, bit 0: Tx Input Equalization Control, Rx Output Amplitude Control, Rx Output De-emphasis Control
 - InfiniBand: Used by host to control module equalization
 - Ethernet: Defined implicitly as part of the host / module interface requirements.

A3.3.2 INFINIBAND VS. ETHERNET MEMORY MAP DIFFERENCES - Tx & Rx UPPER PAGE 00

- Byte 128: Identifier
 - InfiniBand: recommends to use CXP identifier (0Eh) for backward compatibility with 12x FDR and QDR host devices.
 - Ethernet: Either CXP (0Eh) or CXP28 (12h), implementation dependant.
- Byte 149: Identifier
 - Data rates supported by the cable assembly: EDR, FDR, QDR, DDR, SDR
 - Ethernet: Not specified.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

End of Document

IBTA