



What's new – Release 2.0, Vol. 1 and 2

General Overview

Specification update overview – Volume 1



- Volume 1, Release 2.0, published July 31, 2025
- The specification defines **InfiniBand** and **RoCE** (RDMA over Converged Ethernet)
- Available to IBTA Members
 - <https://www.infinibandta.org/ibta-specification/>
- 2156 pages
- New features added by both the LWG and the MgtWG

Specification update overview – Volume 2



- Volume 2, Release 2.0, published July 31, 2025
- The document defines **InfiniBand** physical layer electrical & optical communications interfaces at data rates up through XDR at 200 Gb/s per lane
- Available to IBTA Members
 - <https://www.infinibandta.org/ibta-specification/>
- 843 pages
- New material includes new XDR data rate support, plus additional features regarding link initialization, CMIS 5.3 support, cable types



What's new - Volume 1 Release 2.0

MgtWG (*Management Working Group*)

Support For Large Radix Switches

Added new Class Version 2 MADs in Subnet Management Chapter to support large radix switches (switches with ports up to 64K).

Table 146 Changes between Class Version 1 and Class Version 2

Attribute	AttributeModifier Extended	Attribute Modifier	Attribute Component	Cross Reference
NodeDescription				14.2.6.2 NodeDescription on page 929
NodeInfo			X	14.2.6.3 NodeInfo on page 929
SwitchInfo			X	14.2.6.4 SwitchInfo on page 932
GUIDInfo				14.2.6.5 GUIDInfo on page 937
PortInfo		X	X	14.2.6.6 PortInfo on page 938
P_KeyTable		X	X	14.2.6.7 P_KeyTable on page 973
SLtoVLMapping	X	X	X	14.2.6.8 SLtoVLMappingTable on page 974
VLArbitrationTable		X	X	14.2.6.9 VLArbitrationTable on page 979
LinearForwarding-Table		X	X	14.2.6.10 LinearForwardingTable on page 981
RandomForwarding-Table		X	X	14.2.6.11 RandomForwardingTable on page 982
MulticastForwarding-Table	X	X	X	14.2.6.12 MulticastForwarding-Table on page 984
SMInfo				14.2.6.13 SMInfo on page 988
VendorDiag				14.2.6.14 VendorDiag on page 988
LedInfo				14.2.6.15 LedInfo on page 990

Large Radix Switches - continued

Table 146 **Changes between Class Version 1 and Class Version 2**

Attribute	AttributeModifier Extended	Attribute Modifier	Attribute Component	Cross Reference
LinkSpeedWidth- PairsTable		X	X	14.2.6.16 LinkSpeedWidth- PairsTable on page 990
VendorSpecificMad- sTable				14.2.6.17 VendorSpecificMad- sTable on page 994
CableInfo	X	X		14.2.6.18 CableInfo on page 996
PortInfoExtended		X	X	14.2.6.19 PortInfoExtended on page 999
SwitchPortStateTable		X	X	14.2.6.20 SwitchPortStateTable on page 1012
EnhQosArbiterInfo				14.2.6.21 EnhQoSArbiterInfo on page 1014
EnhPortVLArbiter		X	X	14.2.6.22 EnhPortVLArbiter on page 1015
EnhSLArbiter		X	X	14.2.6.23 EnhSLArbiter on page 1026

Changes in PortInfo & PortInfoExtended between Class Version 1 & Class Version 2

The changes between Class Version 1 and Class Version 2 for PortInfo and PortInfoExtended include changing of specific components, rearranging of several components and moving some components from PortInfo to PortInfoExtended

Table 147 **Changes in Class Version 1 and 2 of PortInfo & PortInfo Extended**

Name	Components	Description
LID space	LID MasterSMLID	Class Version 1 - the components are 16 bits without any adjacent reserved bits.
		Class Version 2 - the components are 16 bits with adjacent 16 reserved bits.
Port Number	LocalPortNum	Class Version 1 - port number is 8 bits. Class Version 2 - port number is 16 bits.
CapabilityMask	CapabilityMask CapabilityMask2	Class Version 1 - the components are two different entities, not adjacent within PortInfo.
		Class Version 2 - the components are merged into a single CapabilityMask and extra reserved bits are added.

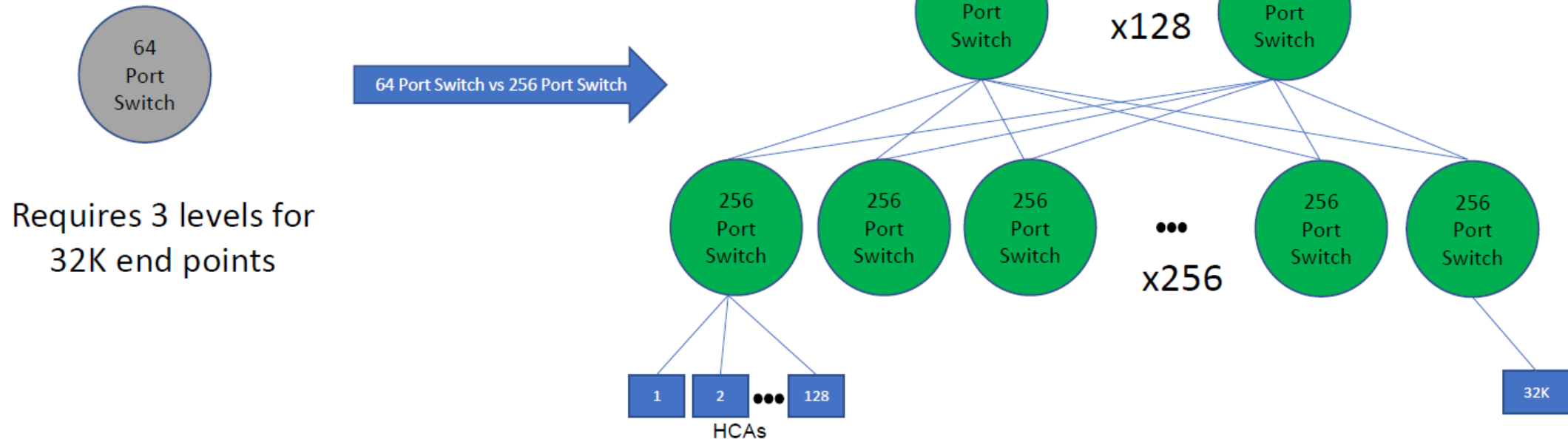
PortInfo & PortInfoExtended Changes Continued

Name	Components	Description
LinkWidth	LinkWidthEnabled LinkWidthSupported, LinkWidthActive	Class Version 1 - the components are scattered across PortInfo. Class Version 2 - the components are adjacent within PortInfo.
LinkSpeed	LinkSpeedSupported LinkSpeedEnabled LinkSpeedExtActive2 LinkSpeedExtSupported2 LinkSpeedExtEnabled2 LinkSpeedExtActive LinkSpeedExtSupported LinkSpeedExtEnabled	Class Version 1 - the components are scattered across PortInfo. Class Version 2 - the components are adjacent within PortInfo.
Components moved from PortInfo to PortInfoExtended	M_Key M_KeyViolations P_KeyViolations Q_KeyViolations M_KeyLeasePeriod MaxRoundTripLatency MulticastPKeyTrapSup- pressionEnabled M_keyProtectBits M_KeyProtectBitsExt PartitionEnforcementIn- bound PartitionEnforcementOut- bound PartitionTopEnabled	Class Version 1 - the components are within PortInfo. Class Version 2 - the components are within PortInfoExtended.

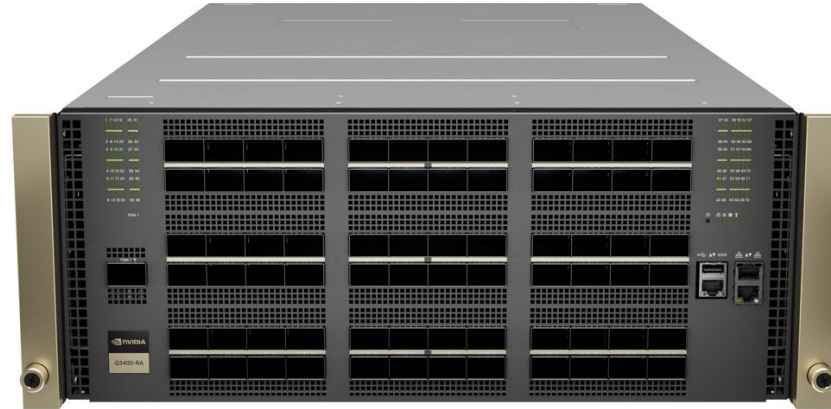
Large Radix Switch Concept

	64 Port Switch	256 Port Switch
Largest scale non blocking 2 level FT	2K HCAs	32K HCA

32K HCAs non blocking topology



Large Radix Switch Example



NVIDIA Quantum InfiniBand Technology Comparison

	Quantum-2 Generation	Quantum-X800 Generation
Network Speed	400Gb/s	800Gb/s
Protocols	InfiniBand	InfiniBand
Radix	Basic system: 64	Basic system: 144
Fat Tree size	2048 ports (2 levels) 65,536 ports (3 levels)	10,368 ports (2 levels) 746,496 ports (3 levels)
Connectivity	Copper between switches (up to 5m) Transceivers (MM, SM, FR4)	Copper between switches (up to 1.5m) Transceivers (SM, FR4)
In-Network Computing	SHARPV3	SHARPV4
Enhancements		Power management



What's new – Volume 1 Release 2.0

LWG (*Link Working Group*)

Network Probe Updates for 2025



- Network Probes are a generalized mechanism for probing the state of the network for both **InfiniBand** and **RoCE**.
- Congestion Control
 - The prime use case is to create an infrastructure to monitor the network and provide accurate congestion control schemes for high performance networking.
- Extended Telemetry
 - Enabling Congestion Control network probes to gather real time telemetry in order to converge the congestion control algorithm.
 - Telemetry format is extendable and can be driven by the network (switch) and the end points (RNIC or HCA).
 - See section A20.2.3.6 RTTPROBE32EXTENDABLE
 - See section A20.2.3.7 RTTPROBE64EXTENDABLE
- Created security architecture for Network Probing
 - Specifying Key Management Scheme for Network Probing
 - See A20.2.1.1 NP_KEY
- Efficient Network Probe
 - Reduced the packet size of the congestion control probe to enable more optimized telemetry collection for congestion control, due to the low latency and high bandwidth nature of AI system

Network Probe Updates Continued

Table 716 **Network Probe Attributes**

Attribute Name	Attribute ID	Attribute-Modifier	Description	Applicable to
RTTProbe32Extendable	0x0014	0x00000000	Provides the measured round trip time between the NPMgt Agents. See A20.2.3.6 RTTPROBE32Extendable on page 2147	HCA & Switch
RTTProbe64Extendable	0x0015	0x00000000	Provides the measured round trip time between the NPMgt Agents. See A20.2.3.7 RTTPROBE64Extendable on page 2150	HCA & Switch
RTTProbe32Short	0x0016	0x00000000	Provides the measured round trip time between the NPMgt Agents. See A20.2.3.8 RTTPROBE32Short on page 2153	HCA & Switch
RTTProbe64Short	0x0017	0x00000000	Provides the measured round trip time between the NPMgt Agents. See A20.2.3.9 RTTPROBE64Short on page 2155	HCA & Switch

Network Probe Updates Continued

Table 717 **Network Probe Attribute / Method Map**

Attribute Name	NPMgtGet()	NPMgtSet()	NPMgtGetResp()	NPMgtTrap()	NPMgtTrapRepress()
ClassPortInfo	X	X	X		
Notice	X	X	X	X	X
ProbingKeyInfo	X	X	X		
RTTProbe32	X		X		
RTTProbe64	X		X		
RTTProbe32Ex- tendable	X		X		
RTTProbe64Ex- tendable	X		X		
RTTProbe32Short	X		X		
RTTProbe64Short	X		X		

Management Datagram Format



13.4.2 MANAGEMENT DATAGRAM FORMAT

C13-3: This compliance statement is obsolete and has been replaced by [C13-3.1: on page 792](#).

C13-3.1: The data payload (as used in [Chapter 9: Transport Layer on page 258](#)) for all MADs shall be exactly 256 bytes except for the MADs listed in table 123 which have payloads that are below 256B.

Table 123 MADs with payloads of less than 256B

Management Class	Management Class ID	Attribute Name	Attribute ID	Payload size
Network Probe	0x22	RTTProbe32Short	0x0016	46B
Network Probe	0x22	RTTProbe64Short	0x0017	46B



What's new - Volume 2 Release 2.0

EWG (*ElectroMechanical Working Group*)

Next Generation Speed

- Release 2.0 supports XDR speed ~200Gb/s per lane.
 - QSFP → 800 Gb/s
 - OSFP (and QSFP-DD for RoCE) → 1600 Gb/s
- New in 2.0 version
 - Add support for XDR Forward Error Correction (FEC)
 - Added direct support for CMIS-based transceiver and cable management
 - Link auto-negotiation, leveraging 802.3 Clause 73, for XDR and slower speeds
 - Better management of passive and active cables and transceivers, electrical and optical
- Miscellaneous
 - Solved issues found in earlier releases

Number of lanes per port	Port Speed Gb/s
1x	200
2x	400
4x	800
8x	1600

XDR Speeds

Roadmap to New Material – Volume 2



Subject	Section Number	Description
Link Auto-Negotiation	Section 5.20	Link Bringup, aligned with IEEE 802.3, Clause 73
XDR New Data Rate	Section 6.11	Electrical and Management Interfaces at 200 Gb/s per lane
QSFP112	Section 7.10	4x receptacle specification, supporting 112 Gb/s per lane
OSFP	Section 7.11	8x/16x Interface Connectors
CMIS	Section 8.6	New Management Specification for pluggable and on-board modules

Note: QSFP, OSFP, & CMIS specification are developed by SNIA/SFF, the OSFP MSA, & the Optical Internetworking Forum (OIF), respectively.

For more information

- Download IBTA Specifications
 - <https://www.infinibandta.org/ibta-specification/>
- Join the IBTA Working Groups:
 - <https://cw.infinibandta.org/workgroup/index>