SNIA COMPUTE + MEMORY + STORAGE SUMMIT

Architectures, Solutions, and Community
VIRTUAL EVENT, APRIL 11-12, 2023

Introduction to CXL Fabrics

Presented by

Vincent Haché

Director of Systems Architecture, Rambus

Introduction to CXL Fabrics

- CXL 3.0 Overview
- Transport Level Details
- Routing Model
- Fabric Management Architecture
- Specification Roadmap

CXL 3.0 Specification

Fabric capabilities and management

Improved memory sharing and pooling

Symmetric coherency

Peer-to-peer

Expanded capabilities for increasing scale and optimizing resource utilization

- Fabric capabilities and fabric attached memory
- Enhance fabric management framework
- Memory pooling and sharing
- Peer-to-peer memory access
- Multi-level switching
- Near memory processing
- Multi-headed devices
- Multiple Type 1/Type 2 devices per root port
- Fully backward compatible to CXL 2.0, 1.1, and 1.0
- Supports PCIe® 6.0

CXL 3.0 Spec Feature Summary

Features	CXL 1.0 / 1.1	CXL 2.0	CXL 3.0
Release date	2019	2020	2022
Max link rate	32GTs	32GTs	64GTs
Flit 68 byte (up to 32 GTs)	✓	✓	✓
Flit 256 byte (up to 64 GTs)			✓
Type 1, Type 2 and Type 3 Devices	✓	✓	✓
Memory Pooling w/ MLDs		✓	✓
Global Persistent Flush		✓	✓
CXL IDE		✓	✓
Switching (Single-level)		✓	✓
Switching (Multi-level)			✓
Direct memory access for peer-to-peer			✓
Symmetric coherency (256 byte flit)			✓
Memory sharing (256 byte flit)			✓
Multiple Type 1/Type 2 devices per root port			✓
Fabric capabilities (256 byte flit)			✓

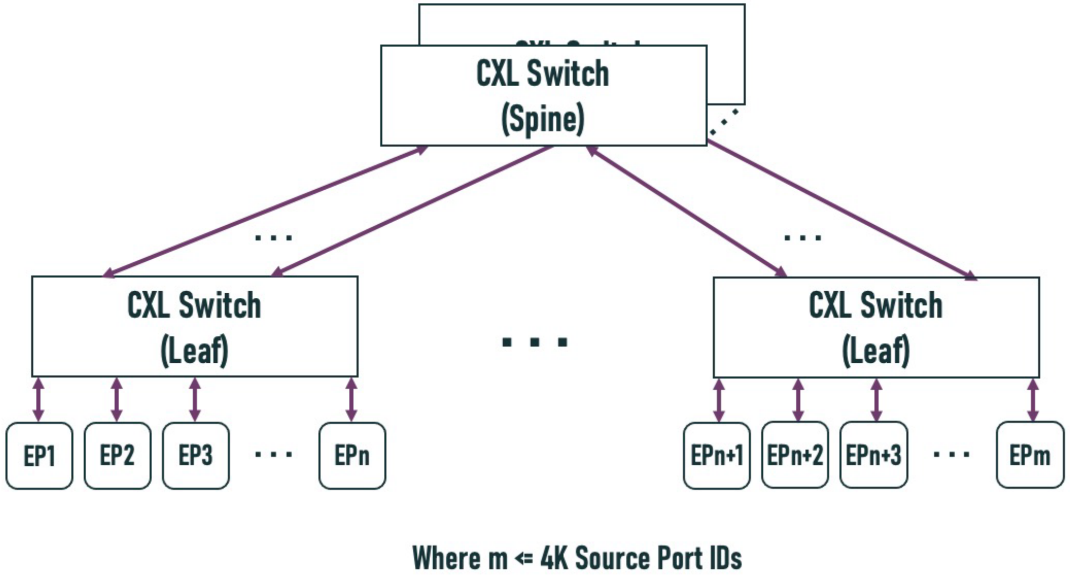
Not supported

✓ Supported

SNIA COMPUTE + MEMORY + STORAGE SUMMIT

CXL Fabrics – Motivation and Overview

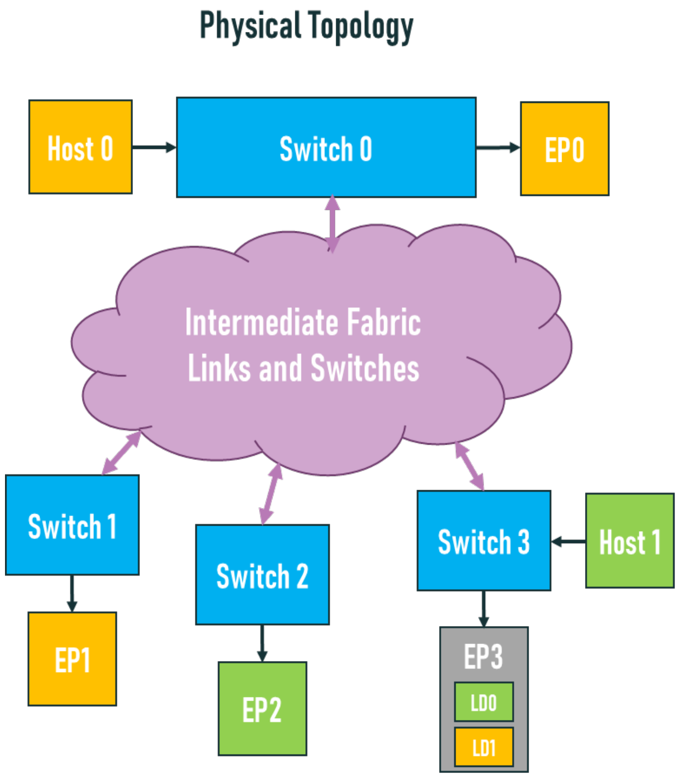
- Disaggregated, Composable Systems – Pooled Host, Device and Memory Resources
- Scale-out Systems – HPC/ML/Analytics
- Add capabilities to expand CXL from node to small number of racks
- Limited by 12b ID space (4k IDs)
- Scale beyond tree-based topologies
- Does not compromise node level properties

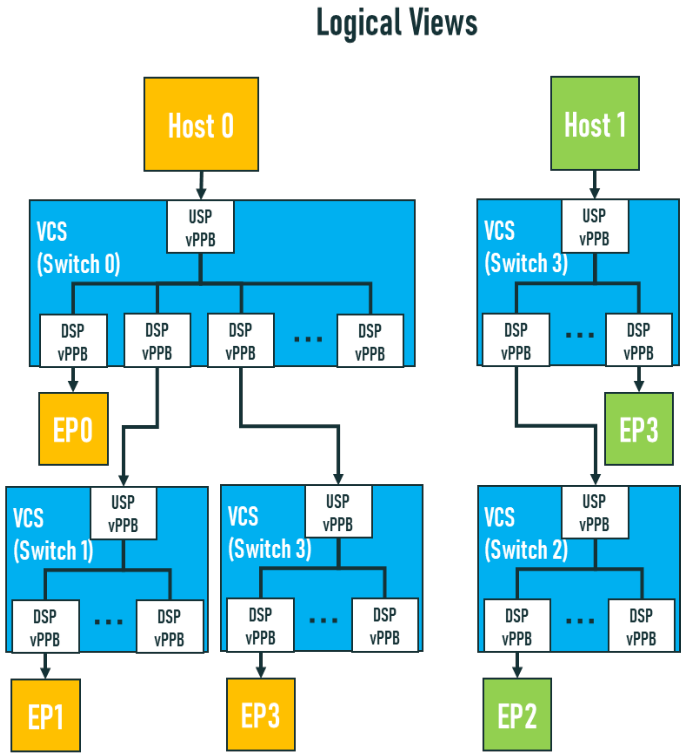


CXL Fabrics – Composability

EP binding from across fabric:

- Host sees up to 2 layers of standard switches: host edge and downstream edge
- Enables re-use of existing host SW

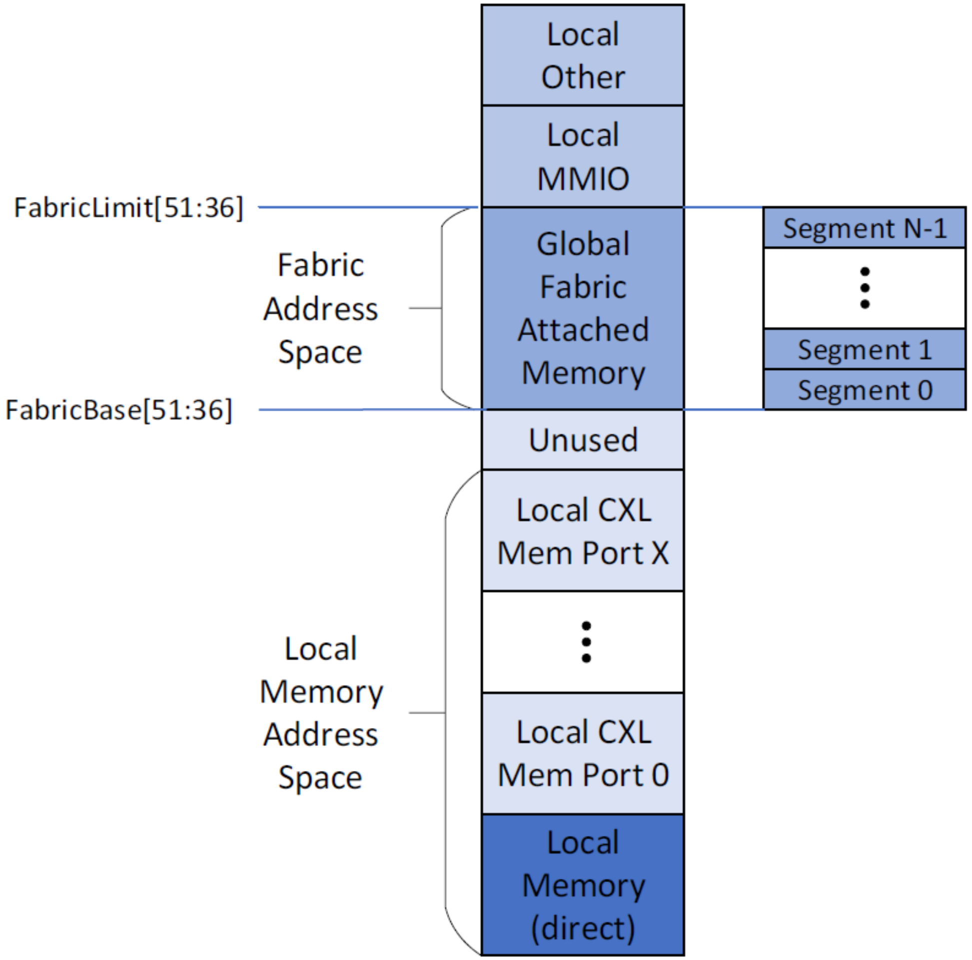




CXL Fabrics – Scale-Out

Global Fabric Attached Memory (G-FAM):

- Highly-scalable memory pool – e.g., 2000+ hosts accessing a memory pool of 2000+ G-FAM devices (GFDs)
- Accessible by all hosts through Fabric Address Segment Table (FAST)



SNIA COMPUTE + MEMORY + STORAGE SUMMIT

Architectures, Solutions, and Community
VIRTUAL EVENT, APRIL 11-12, 2023

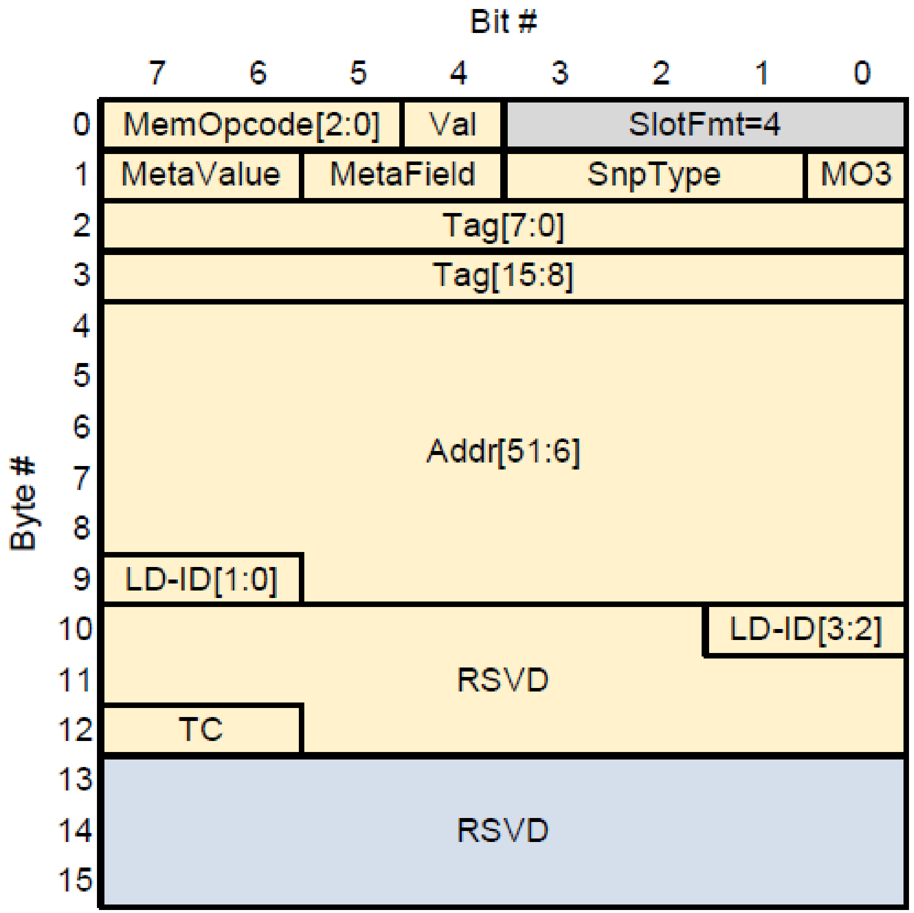
Transport Level Details

Transport Level Details

Host-Based Routing (HBR)

- Covers CXL 1.1/2.0 transport protocol definitions
- Reads and writes requests are address routed
- Restricts links to a single VH – cannot resolve routing otherwise

256B Packing: G4/H4/HS4
HBR Message

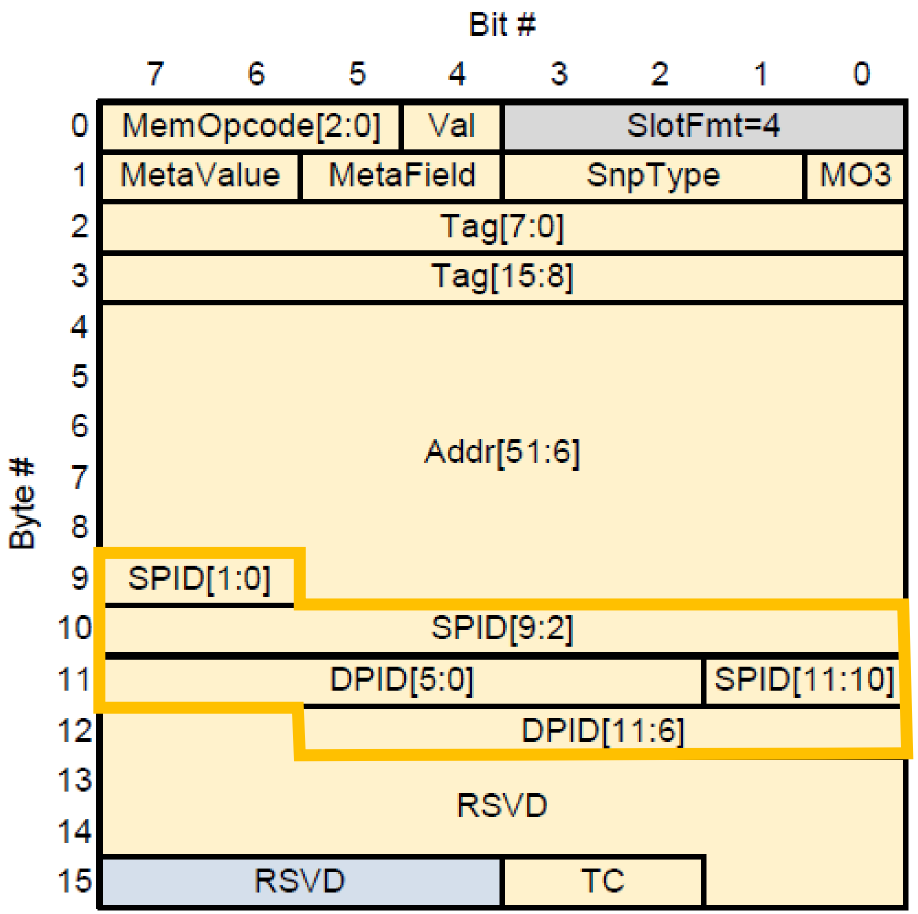


Transport Level Details

Port-Based Routing (PBR)

- Brand new flit mode in 3.0
- Transactions routed by PBR-ID:
  - Destination PBR-ID (DPID) carried by all transactions
  - Source PBR-ID (SPID) carried by select transactions as needed
- Inter-switch links can carry traffic from multiple VHs
- Supported only in 256B flit mode

256B Packing: G4/H4 PBR Message



COMPUTE + MEMORY + STORAGE SUMMIT

Transport Level Details

PBR mode negotiated during Alternate Protocol Negotiation based on capabilities advertised in Symbols 12-14:

12-14	See PCIe Base Specification Specific proprietary usage when Usage = 010b	• Bits[7:0]: **Flex Bus Mode Selection**: — Bit[0]: **PCIe Capable/Enable** — Bit[1]: **CXL.io Capable/Enable** — Bit[2]: **CXL.mem Capable/Enable** — Bit[3]: **CXL.cache Capable/Enable** — Bit[4]: **CXL 68B Flit and VH Capable/Enable** (formerly known as CXL 2.0 Capable/Enable) — Bits[7:5]: **Reserved** • Bits[23:8]: **Flex Bus Additional Info**: — Bit[8]: **Multi-Logical Device Capable/Enable** — Bit[9]: **Reserved** — Bit[10]: **Sync Header Bypass Capable/Enable** — Bit[11]: **Latency-Optimized 256B Flit Capable/Enable** — Bit[12]: **Retimer1 CXL Aware**[1] — Bit[13]: **Reserved** — Bit[14]: **Retimer2 CXL Aware**[2] — Bit[15]: **CXL.io Throttle Required at 64 GT/s** — Bits[17:16]: **CXL NOP Hint Info[1:0]** — Bit[18]: **PBR Flit Capable/Enable** — Bits[23:19]: **Reserved** See Table 6-11 for more information.

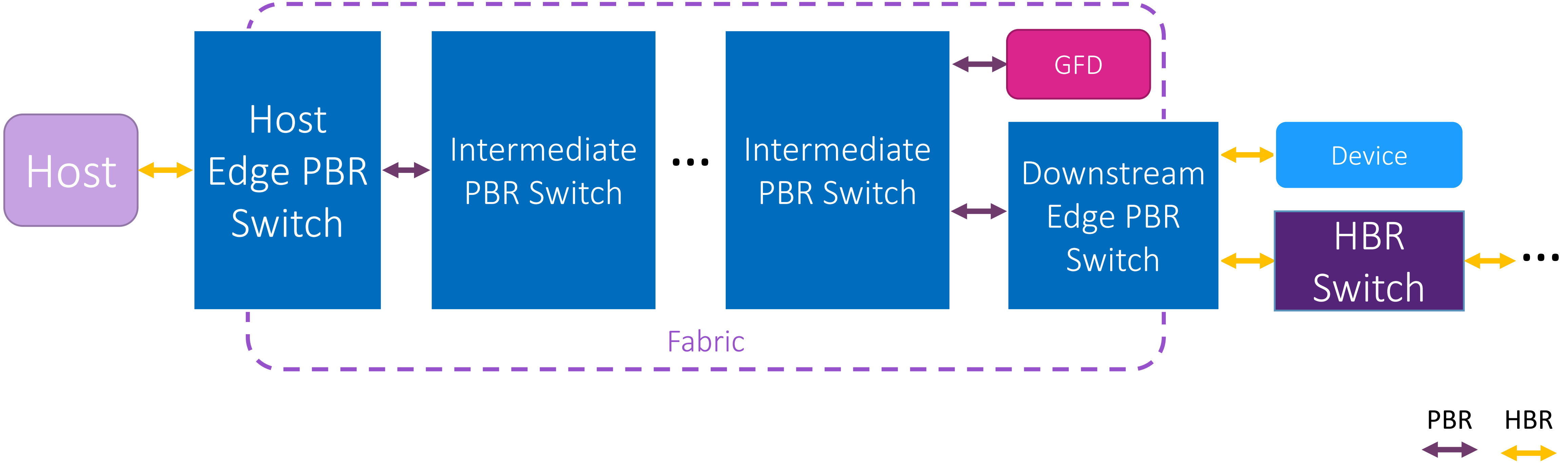
SNIA COMPUTE + MEMORY + STORAGE SUMMIT

Architectures, Solutions, and Community
VIRTUAL EVENT, APRIL 11-12, 2023

Routing Model

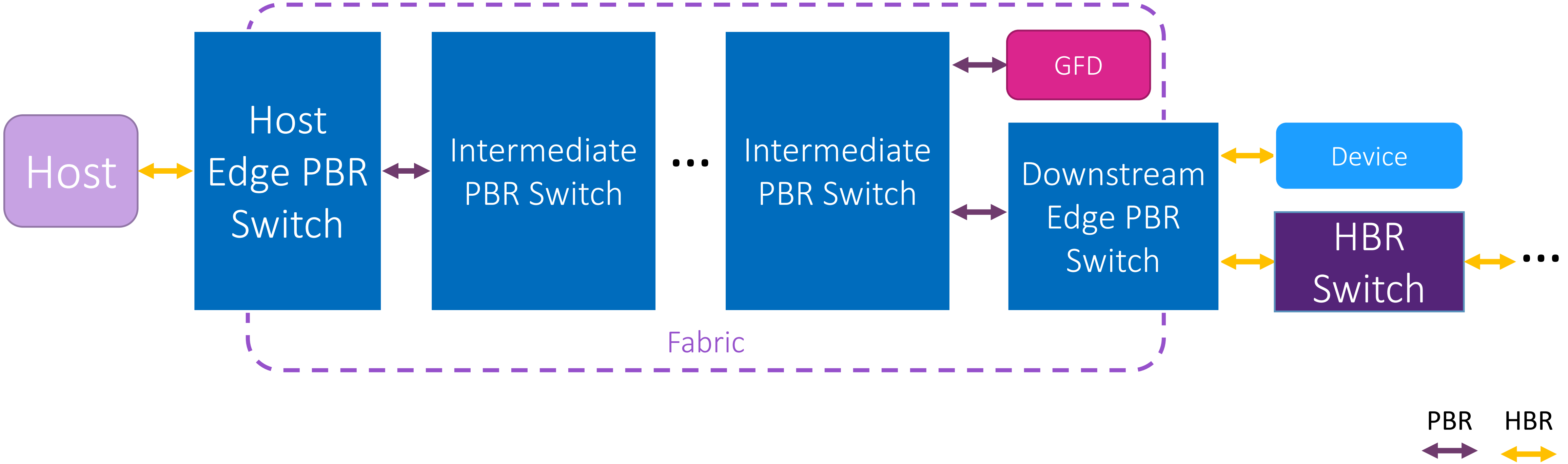
Routing Model

- Host requests begin in HBR format
- Host edge PBR switch converts to PBR format
- Any intermediate switches route by DPID



Routing Model

- GFDs support PBR flit mode
- All other EP types connected to Downstream edge switch
- Downstream edge converts to HBR



SNIA COMPUTE + MEMORY + STORAGE SUMMIT

Architectures, Solutions, and Community
VIRTUAL EVENT, APRIL 11-12, 2023

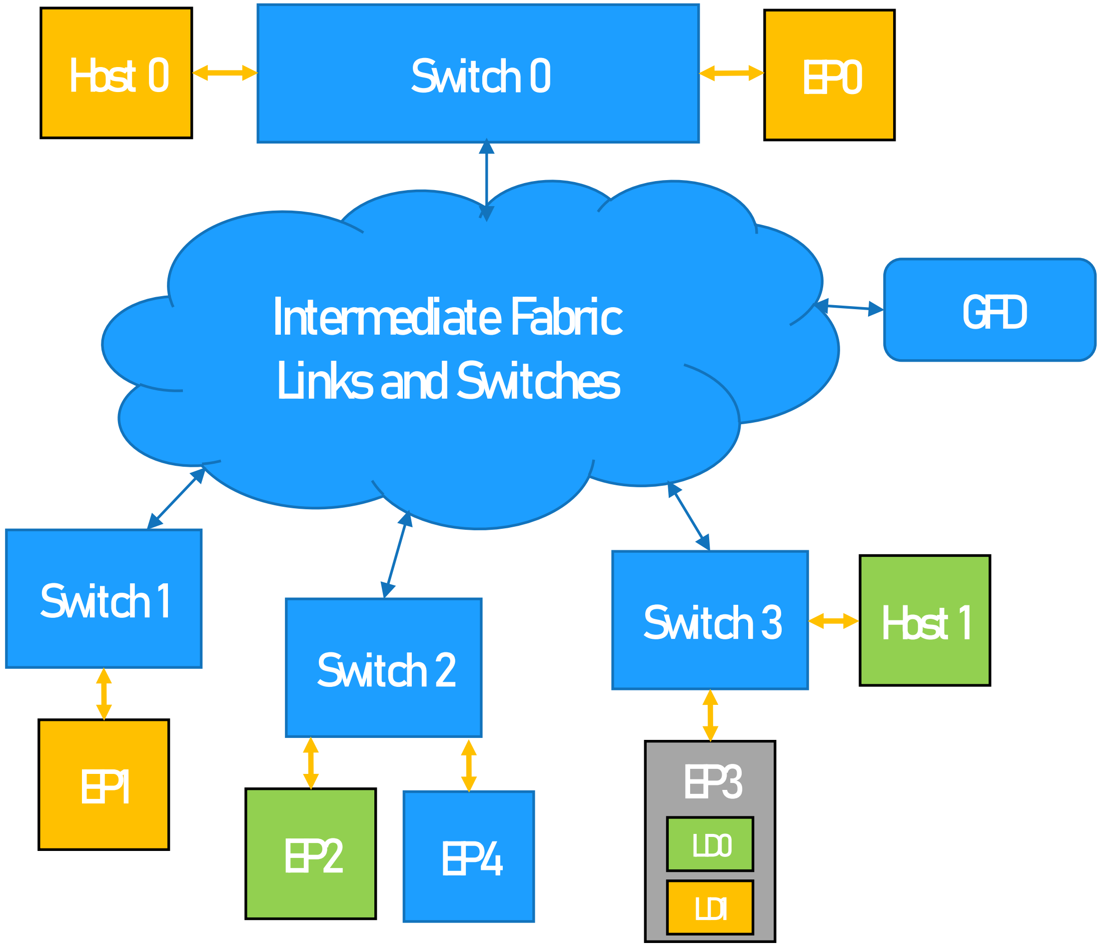
Fabric Management Architecture

Fabric Management Architecture

Fabric Manager (FM) is responsible for:

- Fabric discovery and initialization
- Composition (binding)
- Inter-switch link management
- GFD Management
- Unbound EPs

Host manages its edge link and bound EPs



SNIA COMPUTE + MEMORY + STORAGE SUMMIT

Architectures, Solutions, and Community
VIRTUAL EVENT, APRIL 11-12, 2023

Specification Roadmap

Specification Roadmap

Items to be defined in future specification release:

- CXL Fabric Management Specification
  - PBR Switch Management
  - GFD Management
- Host-to-Host communication
- Device-to-Device communication
- Cross-domain traffic

SNIA COMPUTE + MEMORY + STORAGE SUMMIT

Architectures, Solutions, and Community

VIRTUAL EVENT, APRIL 11-12, 2023

Please take a moment to rate this session.

Your feedback is important to us.

