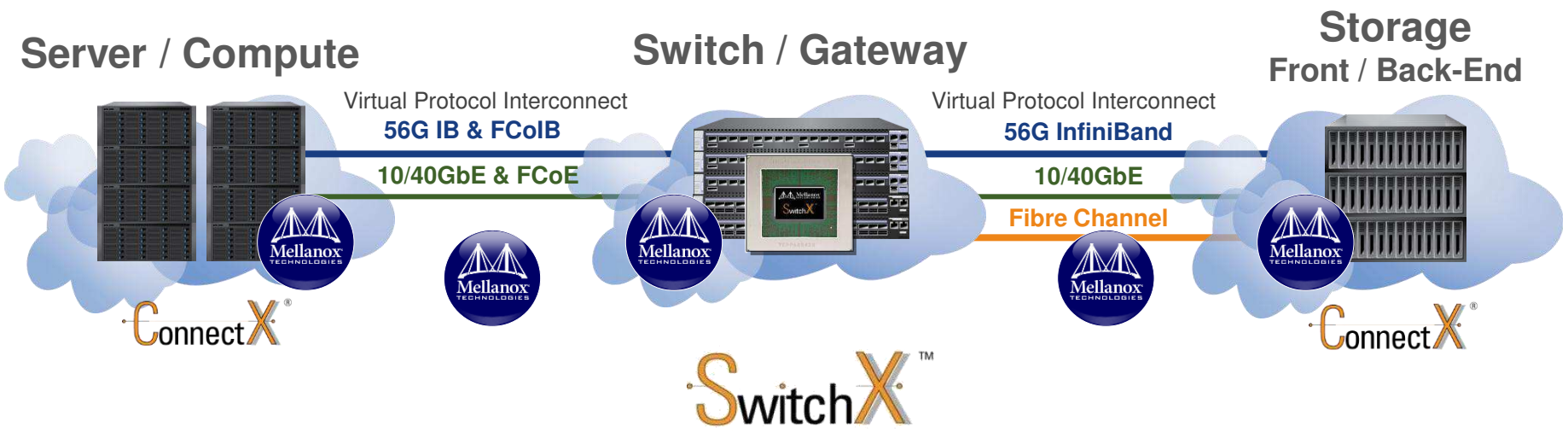
SwitchX

Virtual Protocol Interconnect (VPI)
Switch Architecture

SwitchX™ - Virtual Protocol Interconnect® Solutions



Mellanox End-to-End Virtual Protocol Interconnect Solution

SwitchX™

- **Fifth generation switching IC from Mellanox**
- **Virtual Protocol Interconnect (VPI) technology – 'One-Wire' fabric for InfiniBand – Ethernet – Fibre Channel traffic**
- **Provides Highest Capacity, Lowest Latency, Lowest Power consumption in the Industry**

SwitchX™ Performance and Configuration Flexibility

PERFORMANCE

- 4Tb/s
- 36 x 40/56G
- 200ns Latency
- 40 Watts @ 64 10GE
- 55 Watts @ 36 40GE

1U switch configuration options

- 36 Port FDR IB
- 36 Port 40GigE VPI IB/Ethernet
- 64 Port 10GigE VPI IB/Ethernet
- 12 Port 40GigE/48 Port 40GigE VPI

Blade switch configuration options

- 16 - 40GigE to servers
- 12 - 10GigE to LAN
- 8G FC to SAN/2 - 40GigE stacking ports

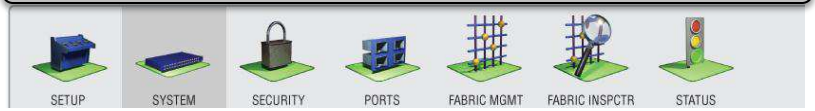
Modular switch chassis options

- Up to 648 56G IB ports
- Up to 648 40GigE ports

SwitchX™ VPI Switch

Unified Fabric Manager

Switch OS Layer



- 64 ports 10GbE
- 36 ports 40GbE
- 48 10GbE + 12 40GbE
- 36 ports IB up to 56Gb/s
- 8 VPI subnets

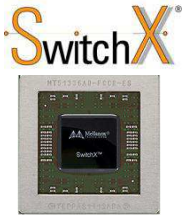
1U switches

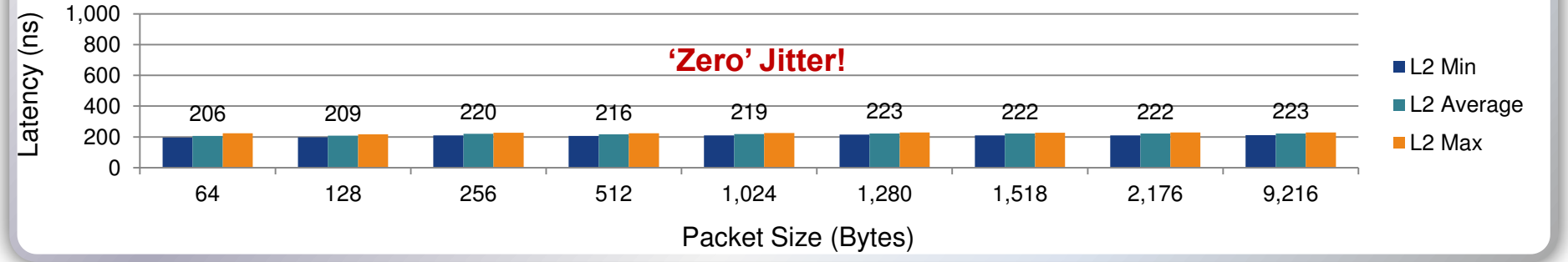
Blade switches

Modular switches

Ethernet Switching Performance

- Throughput (2.5X)
  - 2.88Tb/s throughput on a single chip, running Full Wire Speed at any packet size
- L2 UC/MC Latency for L2/L3 switches (2X)
  - 198-223ns for any packet size
- L3 Latency (2X)
  - 321-337ns for any packet size
- Power Efficiency (6X)
  - Sub 0.6Watt per 10GbE throughput with 100% load at Full Wire Speed

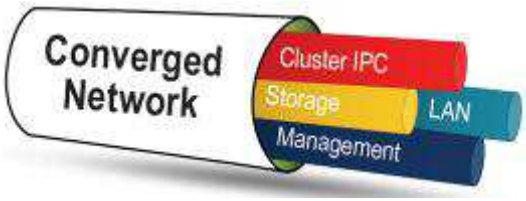


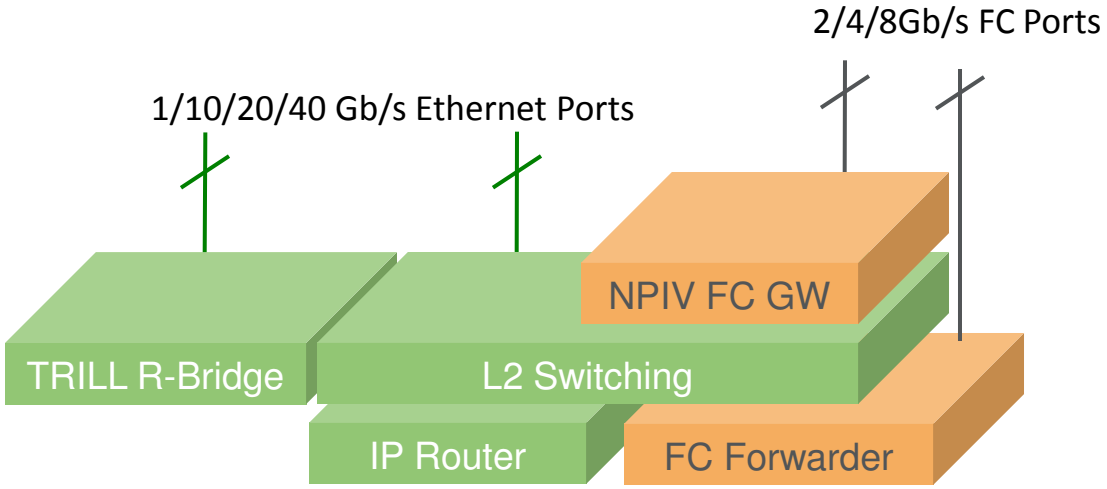


Feature Overview

Physical Layer	SFI XAUI/RXAUI/XLAUI KR/KR2/KR4 - 10/20/40/56 GigE IB SDR, DDR, QDR, FDR10, FDR - 10/20/40/56/100 Gb/s 2/4/8 Gb/s Fibre Channel
L2/L3 Protocols	Ethernet, Datacenter Bridging, TRILL , QCN, Fibre Channel Forwarder, IP Routing, IB Routing
Bridges	Ethernet to Fibre Channel Gateways (NPIV) IB-Ethernet and IB-FC Gateways
QoS	802.1p and DIFFSERV classification, marking Access Control Lists (iACL, eACL)
Management	Dual SGMII PCIe Gen2 x4, or PCIe x1 and Dual SGMII GPIOs, I2C, JTAG
Other Features	Energy Efficient Ethernet, Active Power Governor IEEE 1588 time stamping 40nm process, 1924 pins in 45x45mm FCBGA 1 mm pitch

Converged I/O Solutions

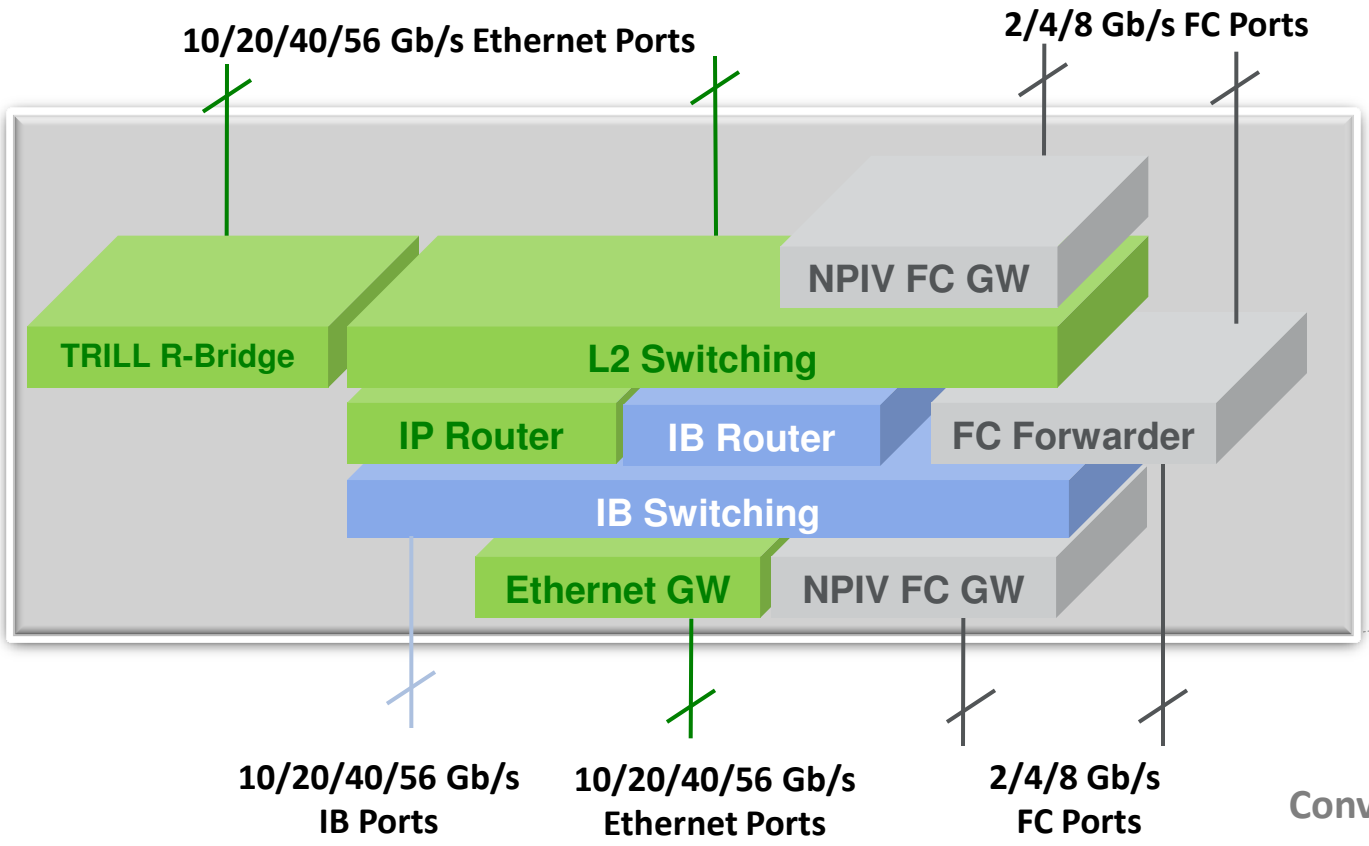




- 2.8/4 Tb/s lossless switching
- 64 10GigE, 36 40GigE
- Flexible mix of ports
  - i.e. 48 10GigE, 12 40GigE
- Multi-chip, high port count configurations
  - Efficient cluster scaling
  - Fat tree scaling
  - Adaptive routing

- NPIV, FCF based native FC ports
  - 2/4/8 Gb/s
  - N, VN, F, VF, E, VE port types
  - Soft and hard zoning
- Sample port configuration
  - 40 10GigE, 24 8Gb/s FC
  - 52 10GigE, 12 8Gb/s FC
  - 24 40GigE, 24 8Gb/s FC
  - 30 40GigE, 12 8Gb/s FC

Virtual Protocol Interconnect (VPI) IO Convergence



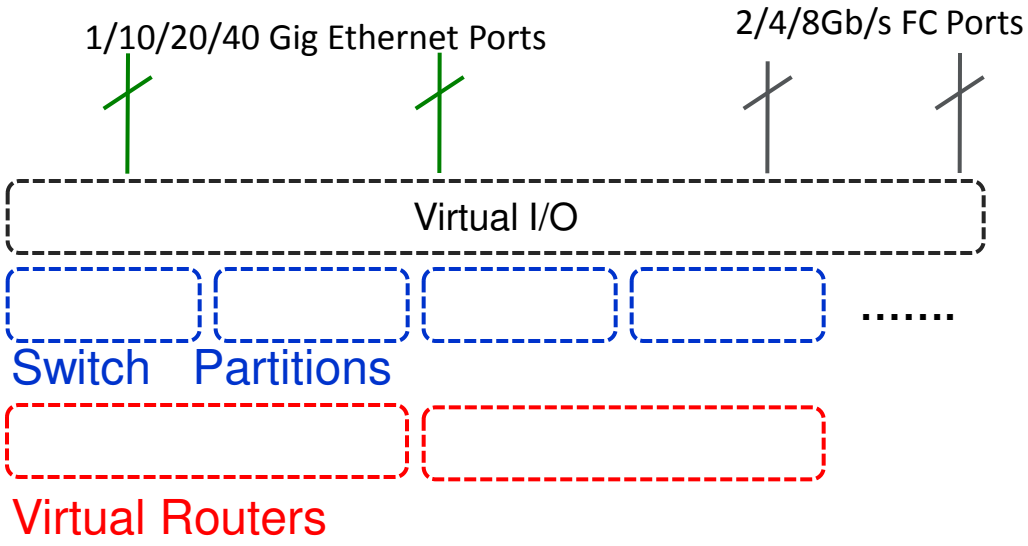
Converged I/O with Ethernet

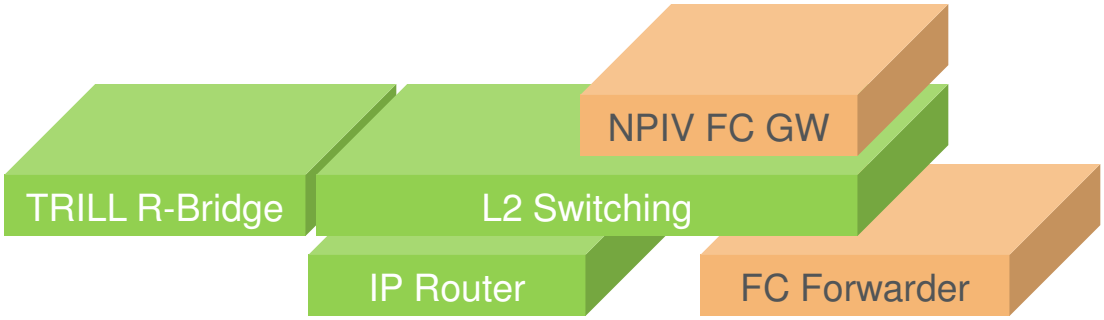
Repurposing

Converged I/O with InfiniBand

Switch Partitioning

- Up to 8 switch partitions can be activated
- Flexible number of Ethernet, FCoE, FC port assignments per switch partition
- Separate L2 data and control plane domains
- Multiple Virtual Routers
- Separate address space per VR
- Isolation and fault containment



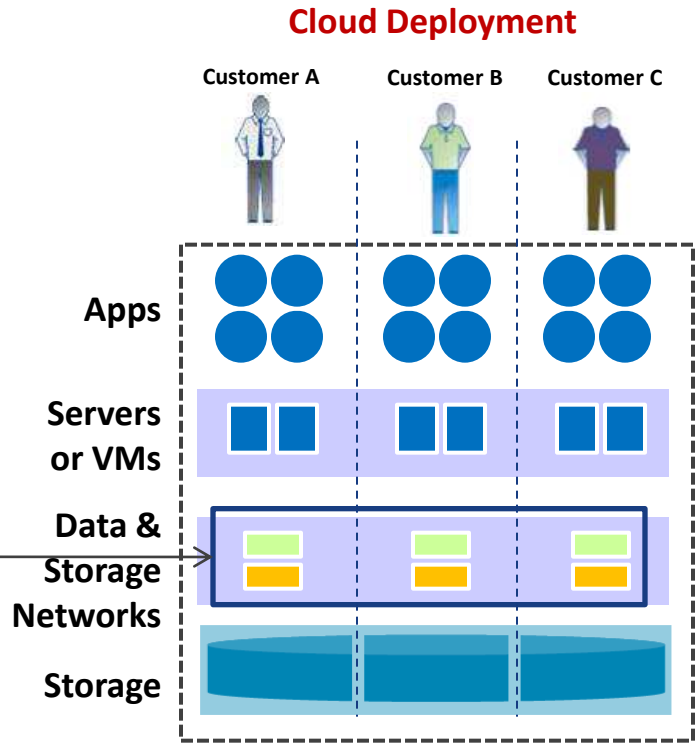


Control Plane Multitenancy

- Multiple switch partitions can be instantiated
  - Like virtual switches inside physical switch
  - Complements virtualized servers and storage
  - Control/data separation like separate switches
- Flexible # of ports & personalities
  - Per switch partition, e.g., IB, L2+ Eth, FC

SwitchX™

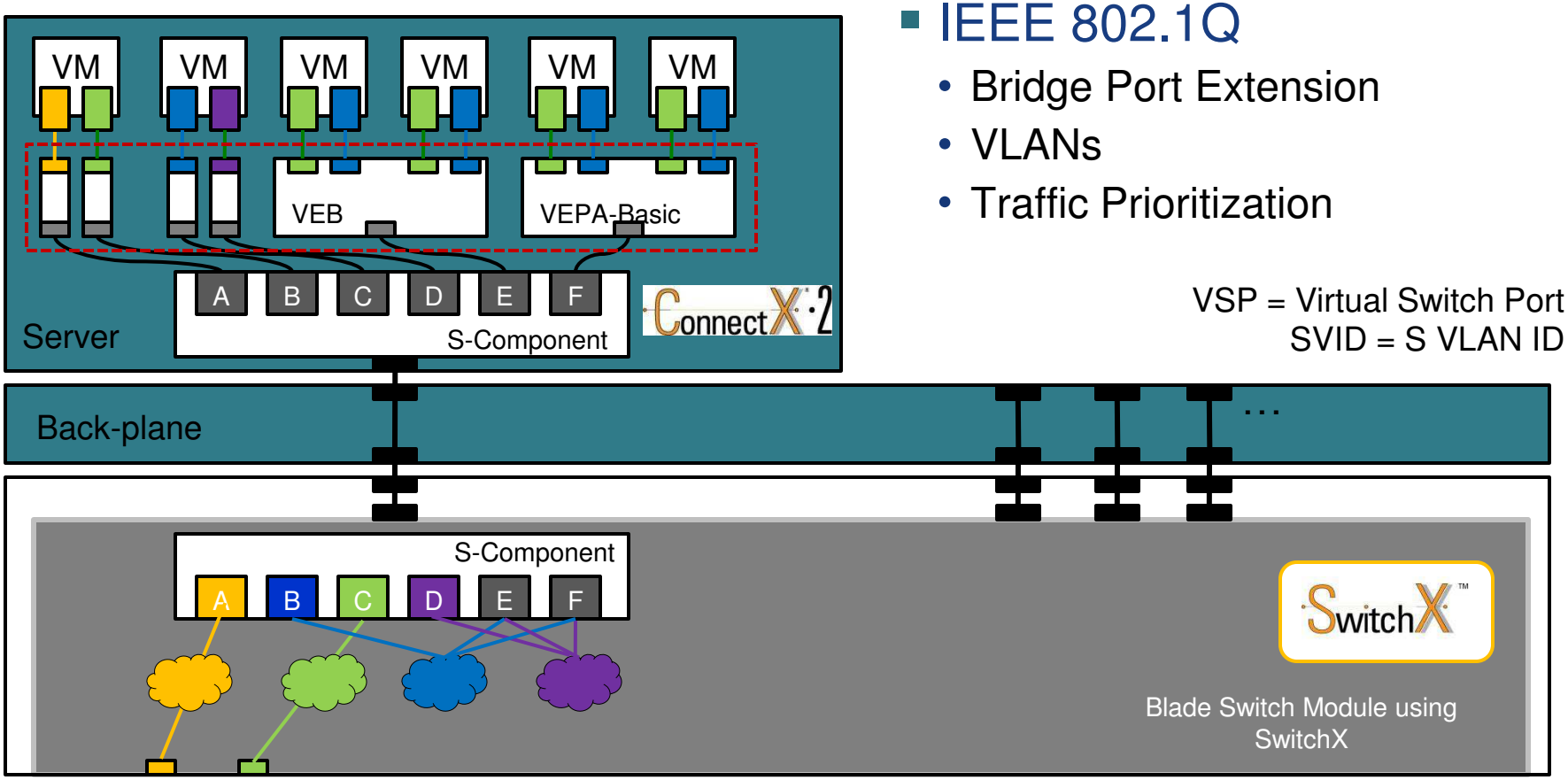
With Switch Partitions



Supports evolving cloud & multi-tenancy architectures

Switch Port Virtualization

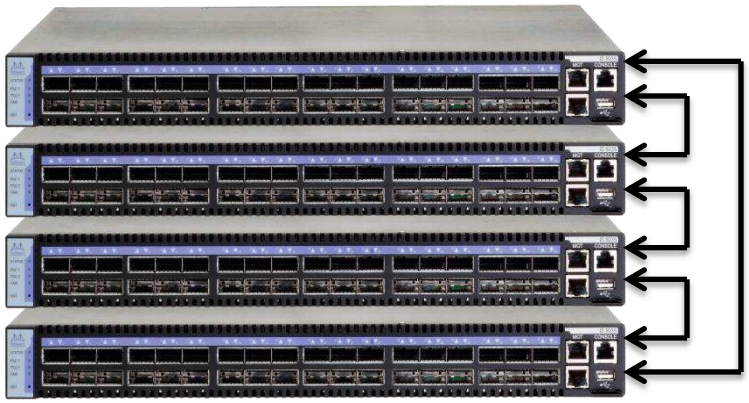
- Flexible VSP Allocation
  - 16 VSPs on 18 ports
  - 8 VSPs on 36 ports
  - 4 VSPs on 64 ports
- Hairpin Mode per VSP
- Switch Partition per VSP
- SVID Allocation
- IEEE 802.1Q
  - Bridge Port Extension
  - VLANs
  - Traffic Prioritization



VSP = Virtual Switch Port
SVID = S VLAN ID

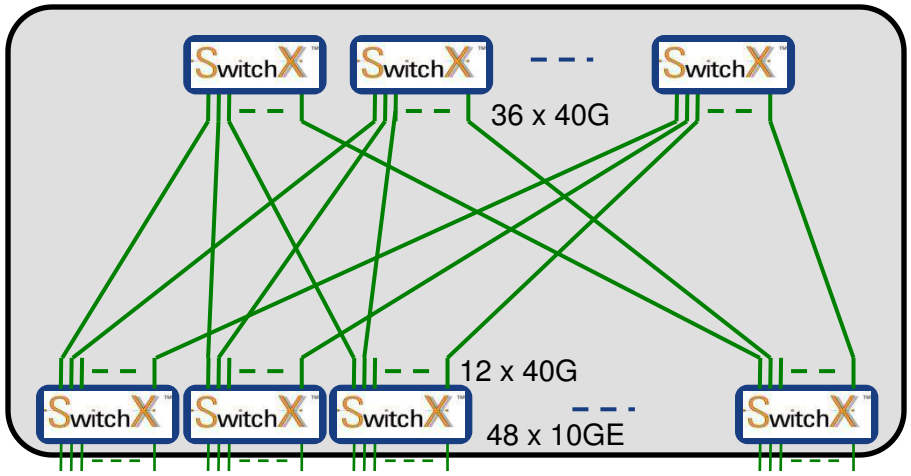
SwitchX Multi-chip Stacking

- Chain and ring topologies
- Any port can be a stacking port
- Single point of management across stacking units (SU)
  - Efficient Inband configuration over management datagrams (eMAD)
- System resiliency
  - Any SU can take charge of the system
  - Alternate paths dynamically used when stacking link down
- Cross system features
  - Link aggregation – ports across SUs in same LAG group
  - ACL – same policy to ports across SUs
    - e.g. VLAN ACL
  - Unified tables are populated on all SUs
    - e.g. L2 filtering DB, L3 routing tables



Multi-chip Configuration - Topologies

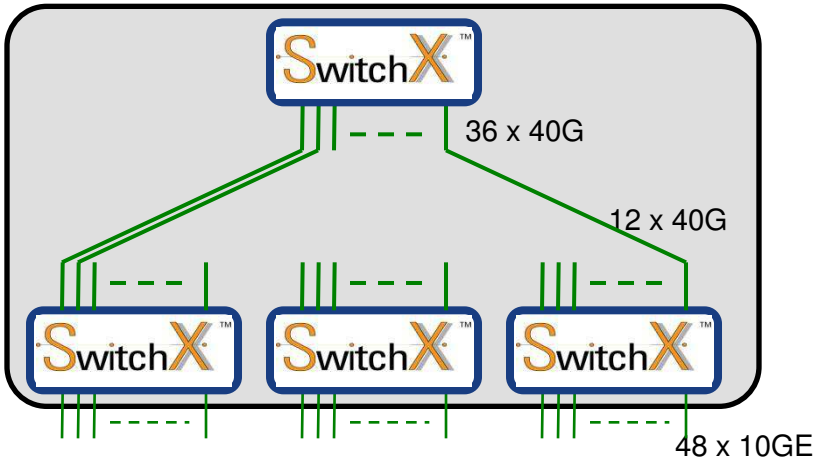
2 Layers FAT-TREE



Up to 1728 10GE Ports

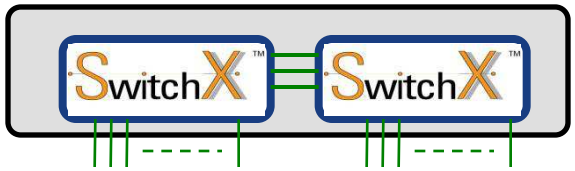
2 Layers FAT-TREE

-Single Spine Chip



Up to 144 10GE Ports

Back to Back



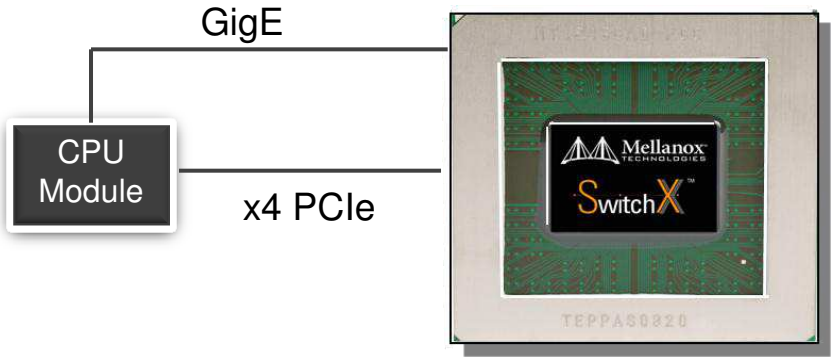
Up to 96 10GE Ports

- Non Blocking
- L2 and L2 Multicast forwarding
- Link Aggregation across fabric
- Port Mirroring across fabric
- Seamless class of service support
- Preserving VLAN membership

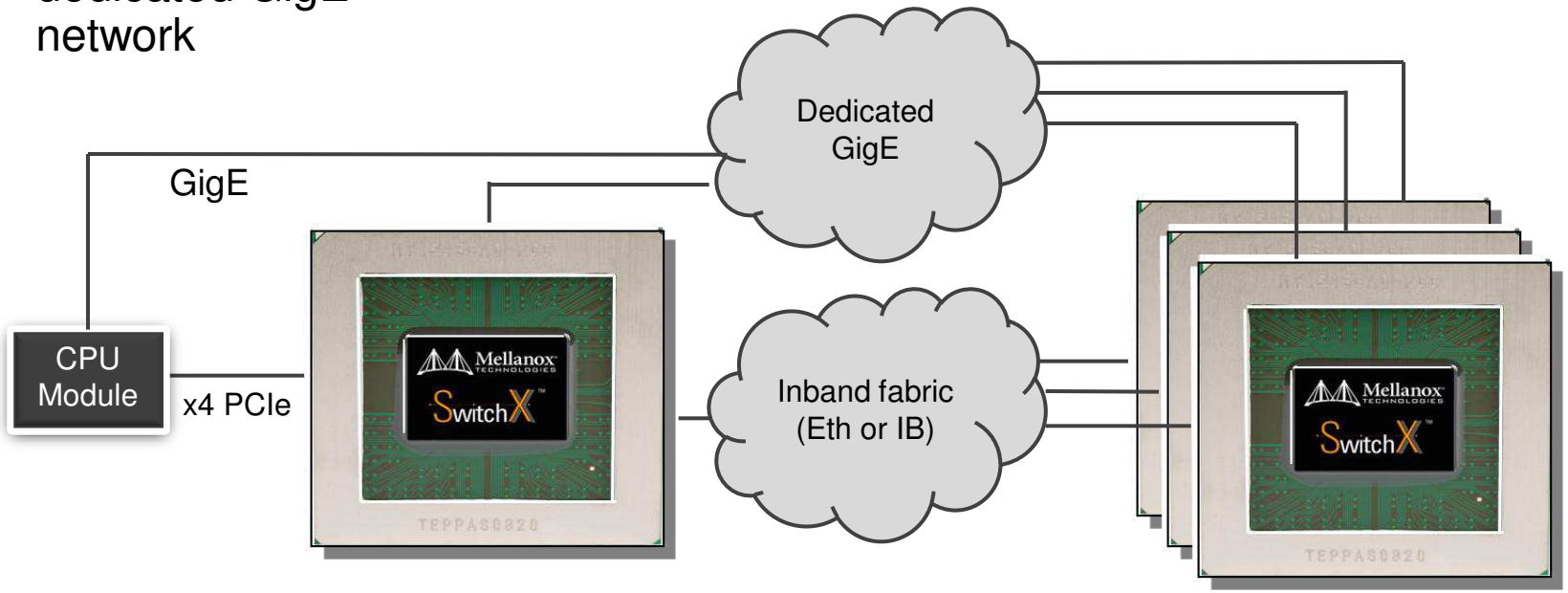
This slide does not present all possible configurations – but rather most reasonable multi-chip configuration topologies

Flexible Switch Management Interface

- Single chip
  - x4 PCIe or dedicated GigE network

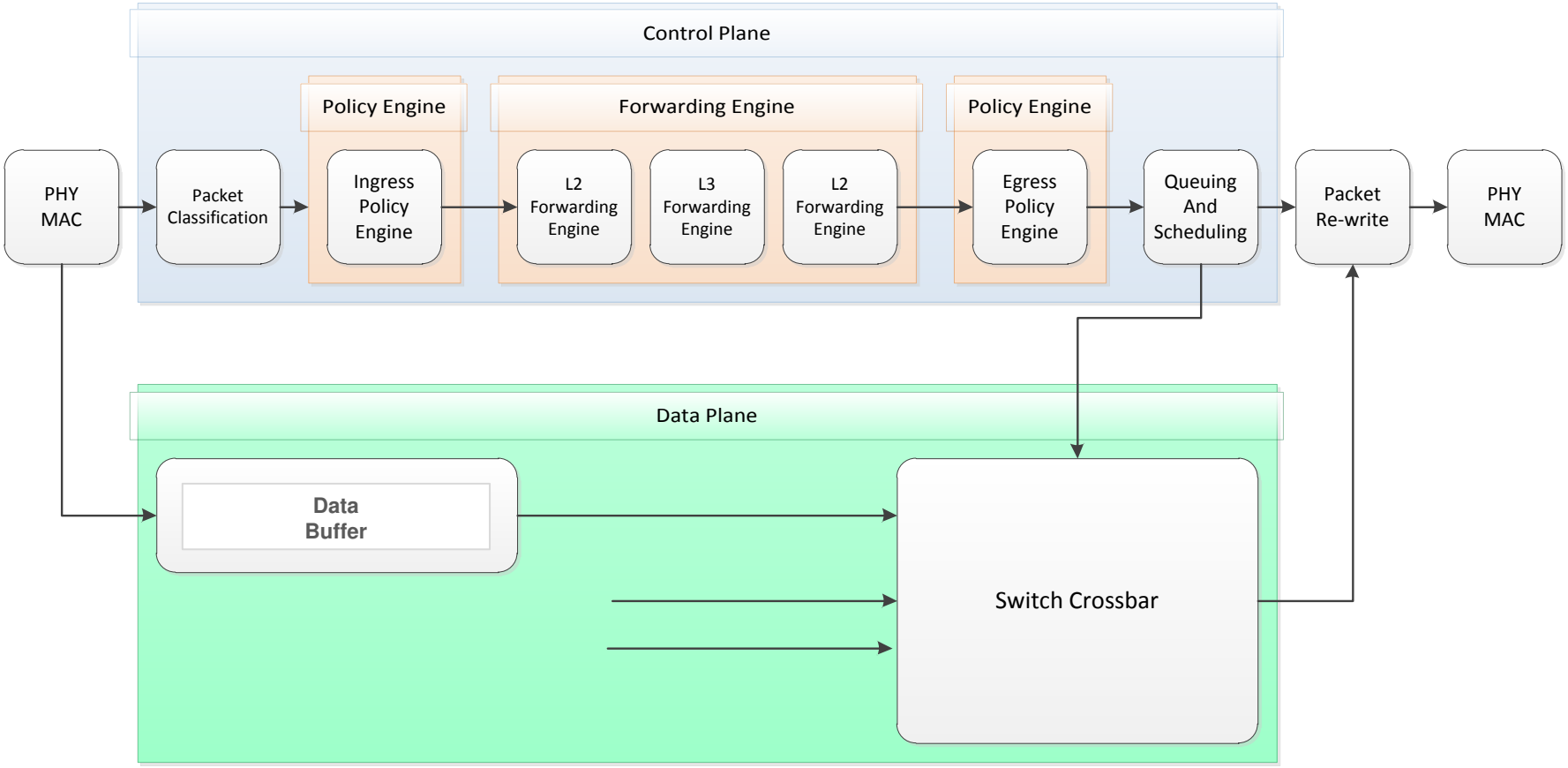


- Multi chip
  - x4 PCIe to inband fabric (Eth or IB) or dedicated GigE network

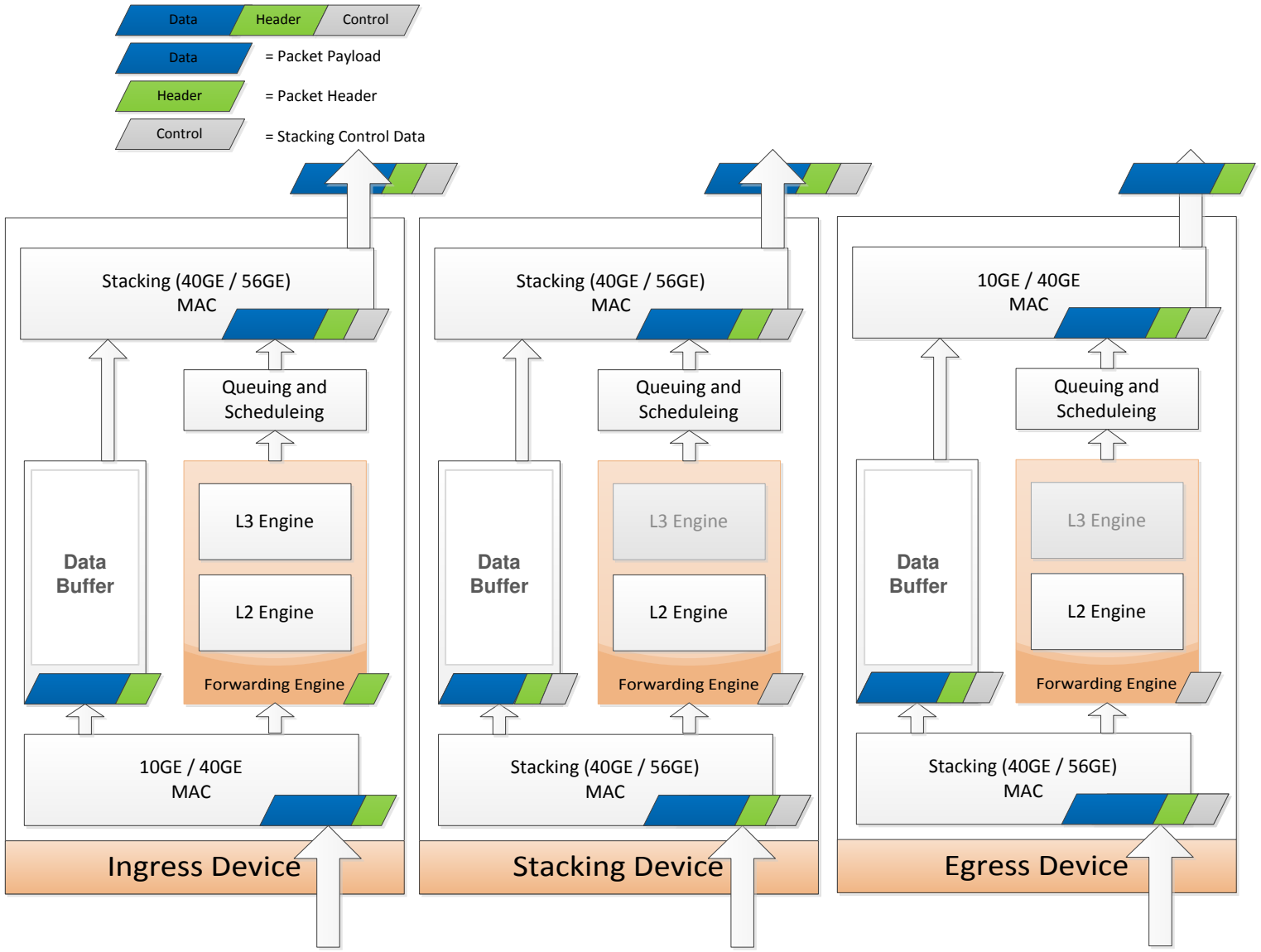


SwitchX Packet Flow Overview

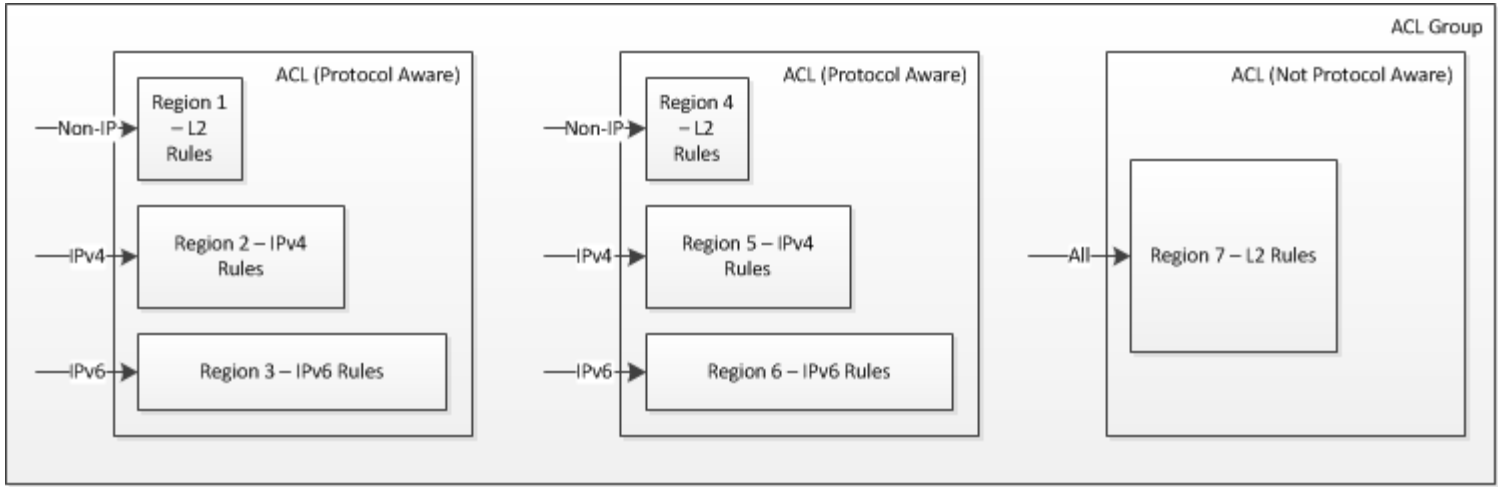
- Fully Pipelined Implementation
- Low-latency cut-through switching support
- Wire speed forwarding



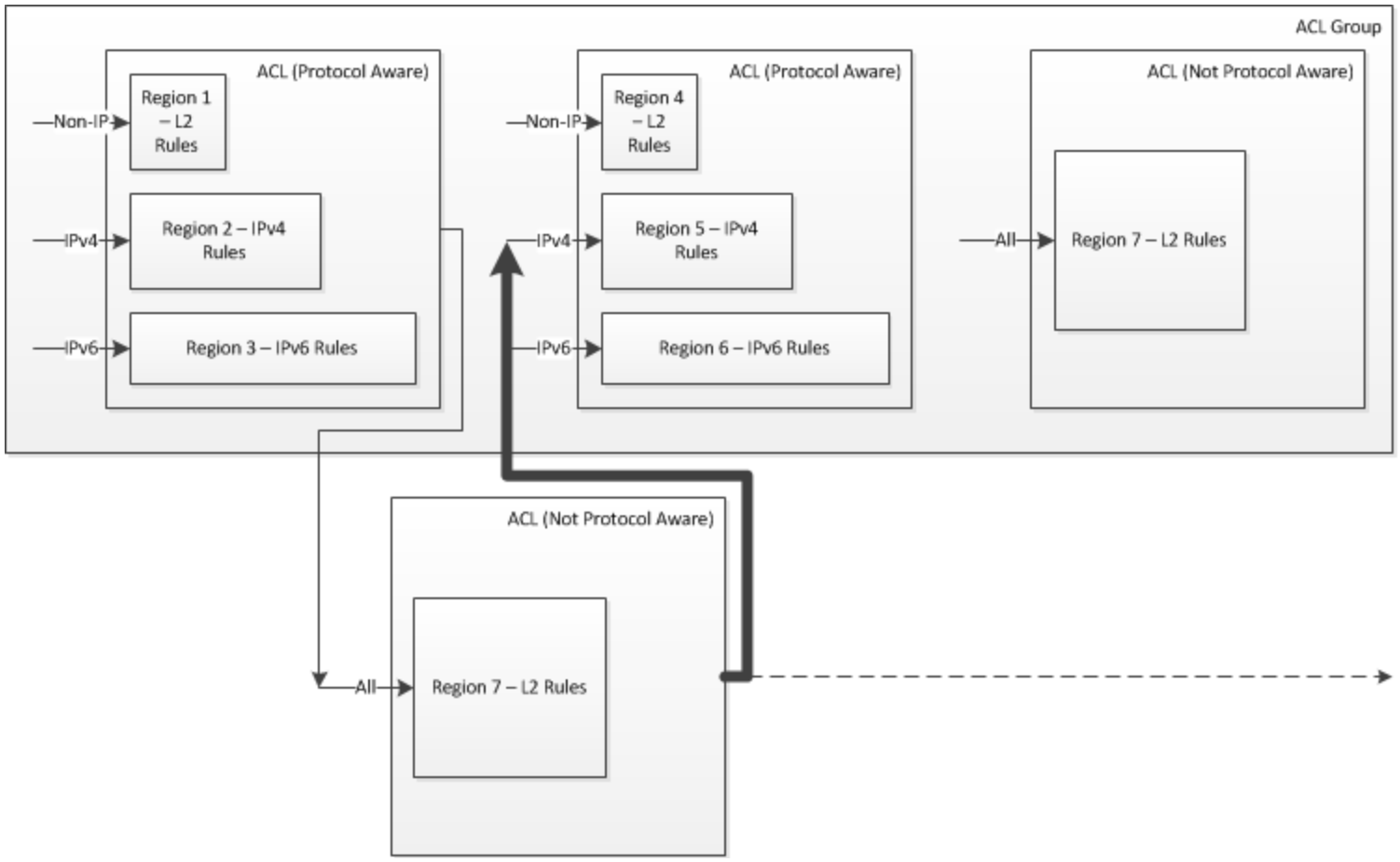
Forwarding Engine – Packet Flows



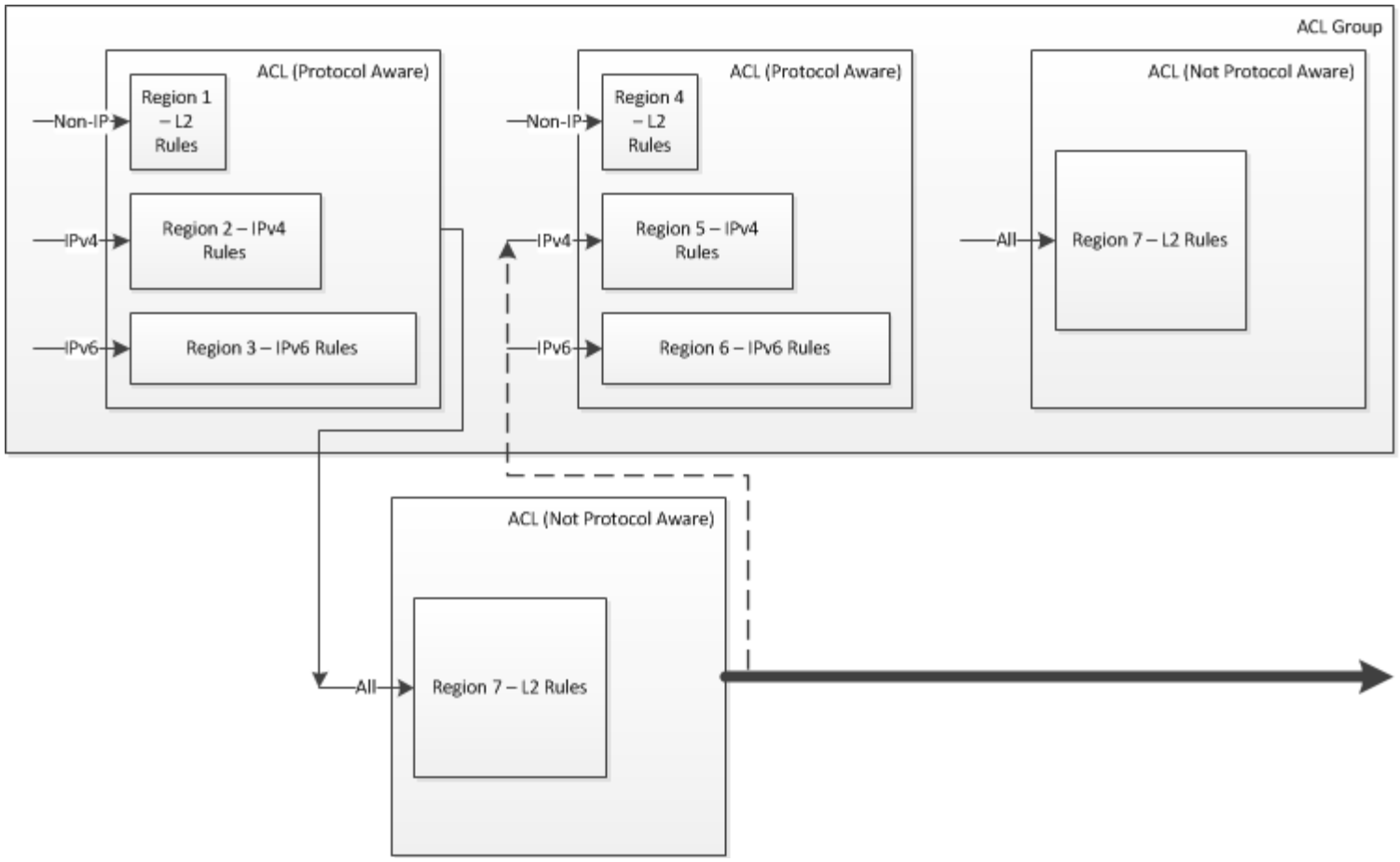
ACLs Architecture Overview



Rule Binding - Nesting

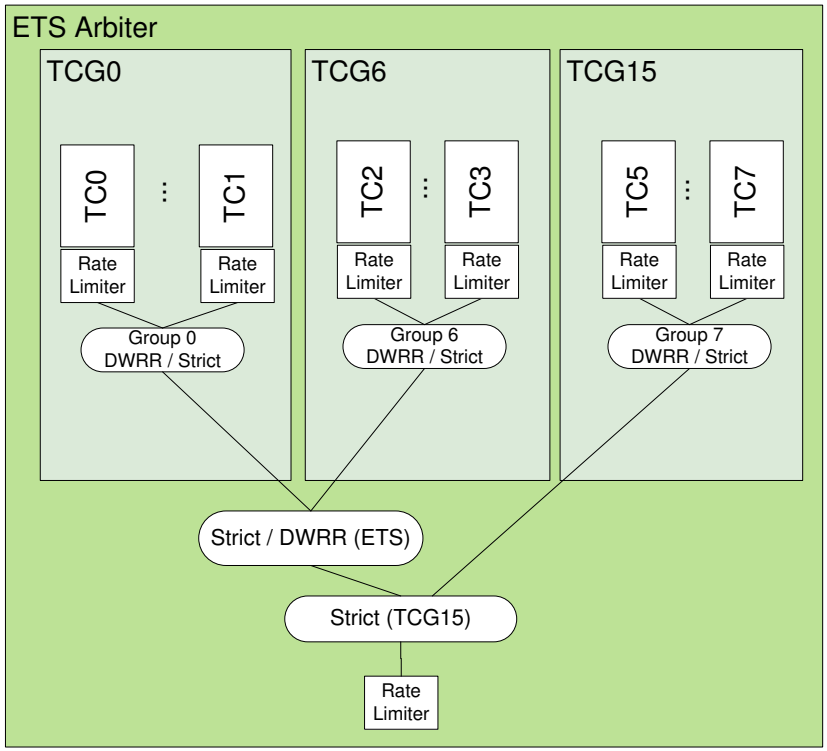


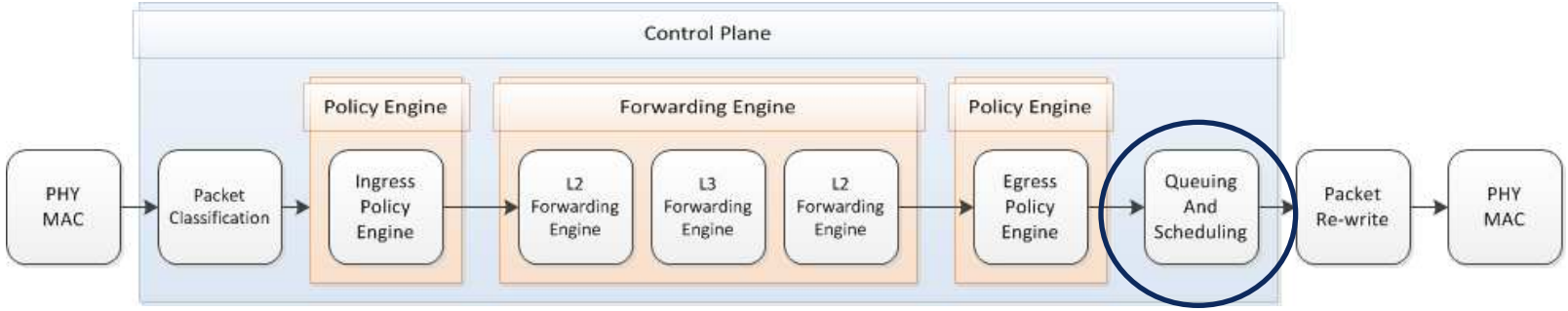
Rule Binding - Break



Queuing and Scheduling

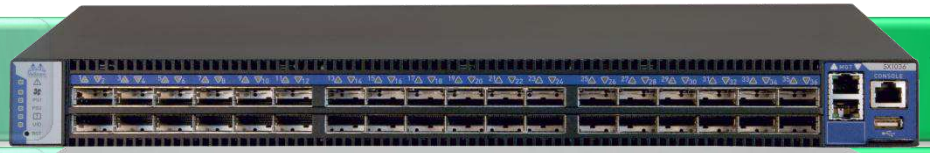
- 8 Traffic Classes
- ETS Scheduling
- Mirroring/Replication
- UC/MC Flows





SwitchX TOR Portfolio

Ideal as a ToR/Core using efficient 40GbE links between 1st and 2nd tiers



SX1036

ToR with 960G of BW equally split between 10G downlinks and 40G uplinks

SX1024

Ideal as a ToR connected to a 3rd party core switch with no 40GigE links

SX1016

- Capacity
  - 36 40GbE ports
  - 64 10GbE ports
  - 48x10GbE+12x40GbE combo
  - Various other port schemes via breakout cables
- Key Features
  - L2/L3 stack
  - VPI
  - 56GbE
  - End to end solution
- Latency
  - 220ns latency 40GbE
    - 330ns L3 latency
  - 270ns latency 10GbE
    - 430ns L3 latency
- Throughput
  - 2.88Tb/s of non-blocking throughput
- Power
  - Under 1W per 10GbE interface
  - 2.3W per 40GbE interface
  - 0.6W per 10GbE of throughput

Mellanox SX6536

- 648 x QSFP 40GE* ports
- 1152 x SFP+ 10GE* ports
- 51.84Tb/s throughput
- 9.6 Watt/40GE port
- Latency: 700ns inter line, 230ns same line
- World's first Cut-Through modular Ethernet switch
- N+N PS Redundancy
- L2/L3 SW Stack
- Same Chassis is used for IB FDR (56Gbps)
- Smaller Chassis (324p, 216p, 108p)
  - Same leafs, spines, management boards
  - Same architecture

Thanks