cisco

Cisco live!

6-9 March 2018 · Melbourne, Australia

Nexus 9000 Architecture

Mike Herbert
Principal Engineer
CCIE 8479 Emeritus

BRKDCT-3640

Cisco live!

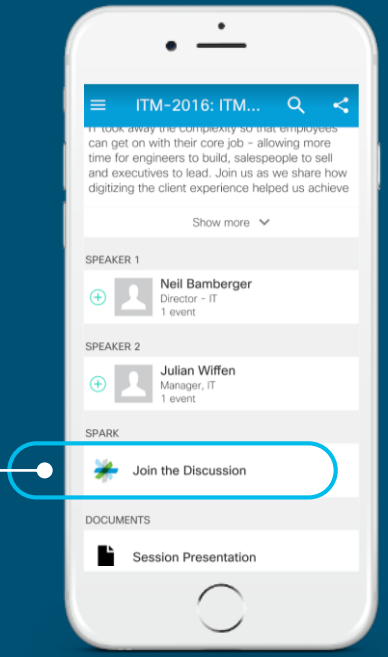
Cisco Spark

Questions?

Use Cisco Spark to communicate with the speaker after the session

How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install Spark or go directly to the space
4. Enter messages/questions in the space



What this Session Covers

- Latest generation of Nexus 9000 switches with Cloud Scale ASICs
- Nexus 9500 modular switches with Cloud Scale linecards
- Nexus 9300 Cloud Scale top-of-rack (TOR) switches
- System and hardware architecture, key forwarding functions, packet walks

Not covered:

- First generation Nexus 9000 ASIC/platform architectures
- Nexus 9500 merchant-silicon based architectures
- Other Nexus platforms
- *Catalyst* 9000 platform

Agenda

- Data Centre and Silicon Strategy
- Cloud Scale Architecture
  - Cloud Scale ASICs
  - Forwarding and Features
- Cloud Scale Switching Platforms
- Optics and What’s Next

Nexus 9000 Switching Portfolio

Key Elements of the ASAP Data Centre

Nexus 9500 X9400 / X9400-S

Merchant Broadcom XGS (Trident2+ / Tomahawk)

- Broadcom SOC solution
- Wide industry availability
- Published SDK

Nexus 9500 X9600-R / X9600-RX

Merchant Broadcom DNX (Jericho)

- Multi-chip architecture
- Large forwarding tables
- Deep packet buffer / VOQ
- Cell-based fabric

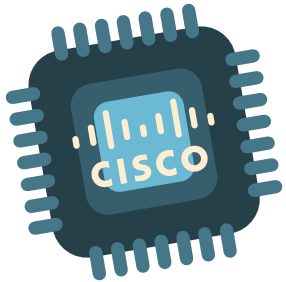
Nexus 9500 X9700-EX/FX
Nexus 9300-EX/FX/FX2

Cisco Cloud Scale (LSE / LS1800FX / S6400 / LS3600FX2)

- Cisco SOC solution
- Rich forwarding feature set
- Smart buffers
- Advanced telemetry
- Optimised scale, cost, power

Focus of this session

Why Custom Silicon?



Cisco competitive advantage – vehicle for differentiating innovations

- Application Centric Infrastructure (ACI) policy model + congestion-aware flowlet switching
- Flexible forwarding tiles
- Single-pass tunnel encapsulations
- In-built encryption technologies
  - MACSEC, CloudSec
- Intelligent Buffers – DBP / AFD / DPP
- Streaming telemetry:
  - Flow Table for Tetration Analytics
  - Flow table event notifications
  - Streaming Statistics Export (SSX)

Tight integration between hardware / software / marketing / sales / support

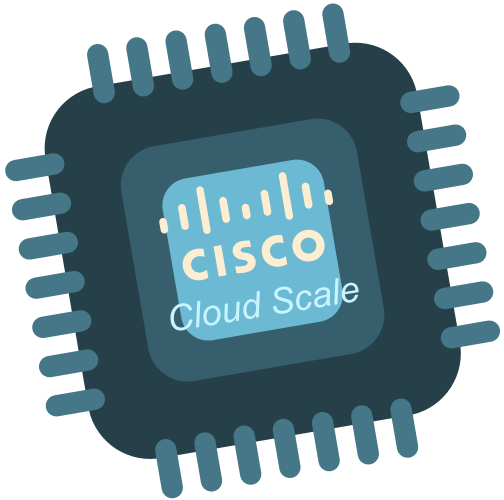
- Closely aligns hardware designs with software innovations, strategic product direction, competitive differentiators, serviceability

Agenda

- Data Centre and Silicon Strategy
- Cloud Scale Architecture
  - Cloud Scale ASICs
  - Forwarding and Features
- Cloud Scale Switching Platforms
- Optics and What's Next

Cisco Cloud Scale ASIC Family

- **Ultra-high port densities** → Reduces equipment footprint, enables device consolidation
- **Multi-speed 100M/1/10/25/40/50/100G** → Flexibility and future proofing
- **Rich forwarding feature-set** → ACI, Segment Routing, single-pass VXLAN routing
- **Flexible forwarding scale** → Single platform, multiple scaling alternatives
- **Intelligent buffering** → Shared egress buffer with dynamic, advanced traffic optimisation
- **In-built analytics and telemetry** → Real-time network visibility for capacity planning, security, and debugging



Cloud Scale Family Members

LSE

- 1.8T chip – 2 slices of 9 x 100G each
- X9700-EX modular linecards; 9300-EX TORs

LS1800FX

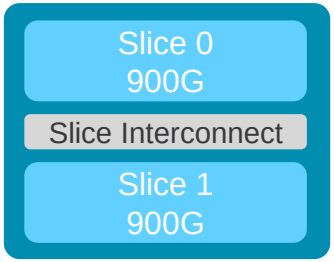
- 1.8T chip – 1 slice of 18 x 100G with MACSEC
- X9700-FX modular linecards; 9300-FX TORs

S6400

- 6.4T chip – 4 slices of 16 x 100G each
- 9364C TOR; E2 fabric modules

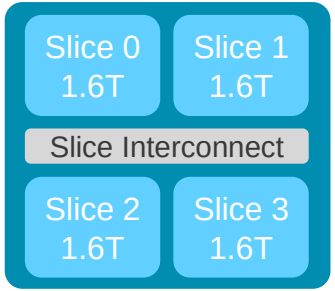
LS3600FX2

- 3.6T chip – 2 slices of 18 x 100G with MACSEC + CloudSec
- 9300-FX2 TORs

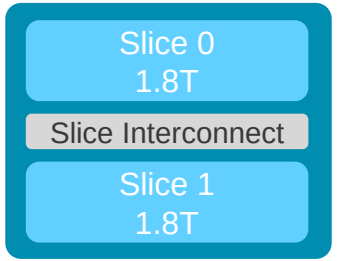


LSE – 18 x 100G

LS1800FX – 18 x 100G



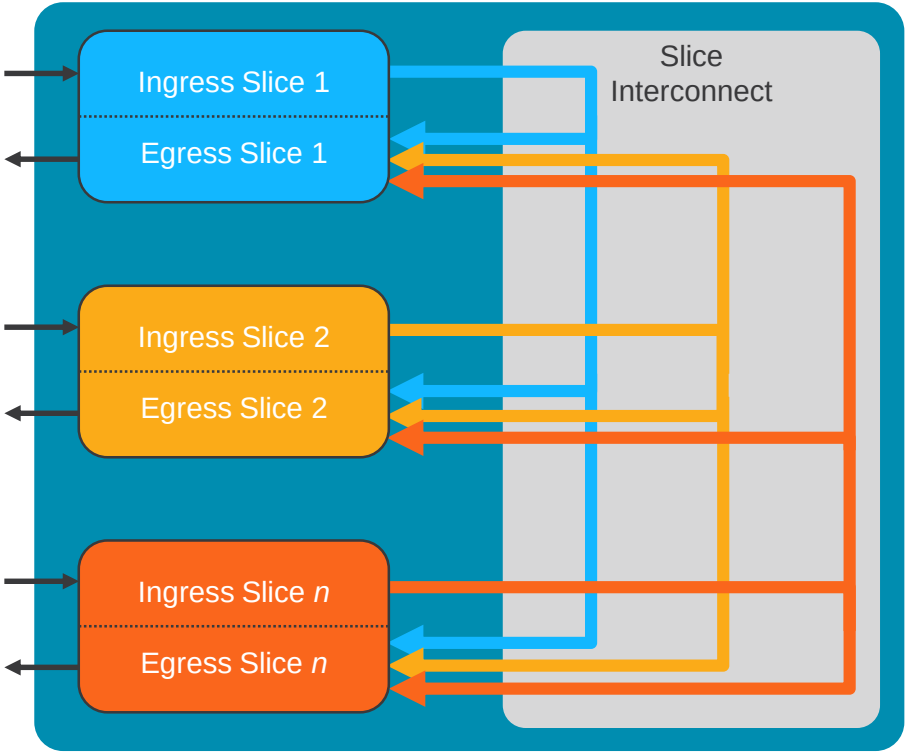
S6400 – 64 x 100G



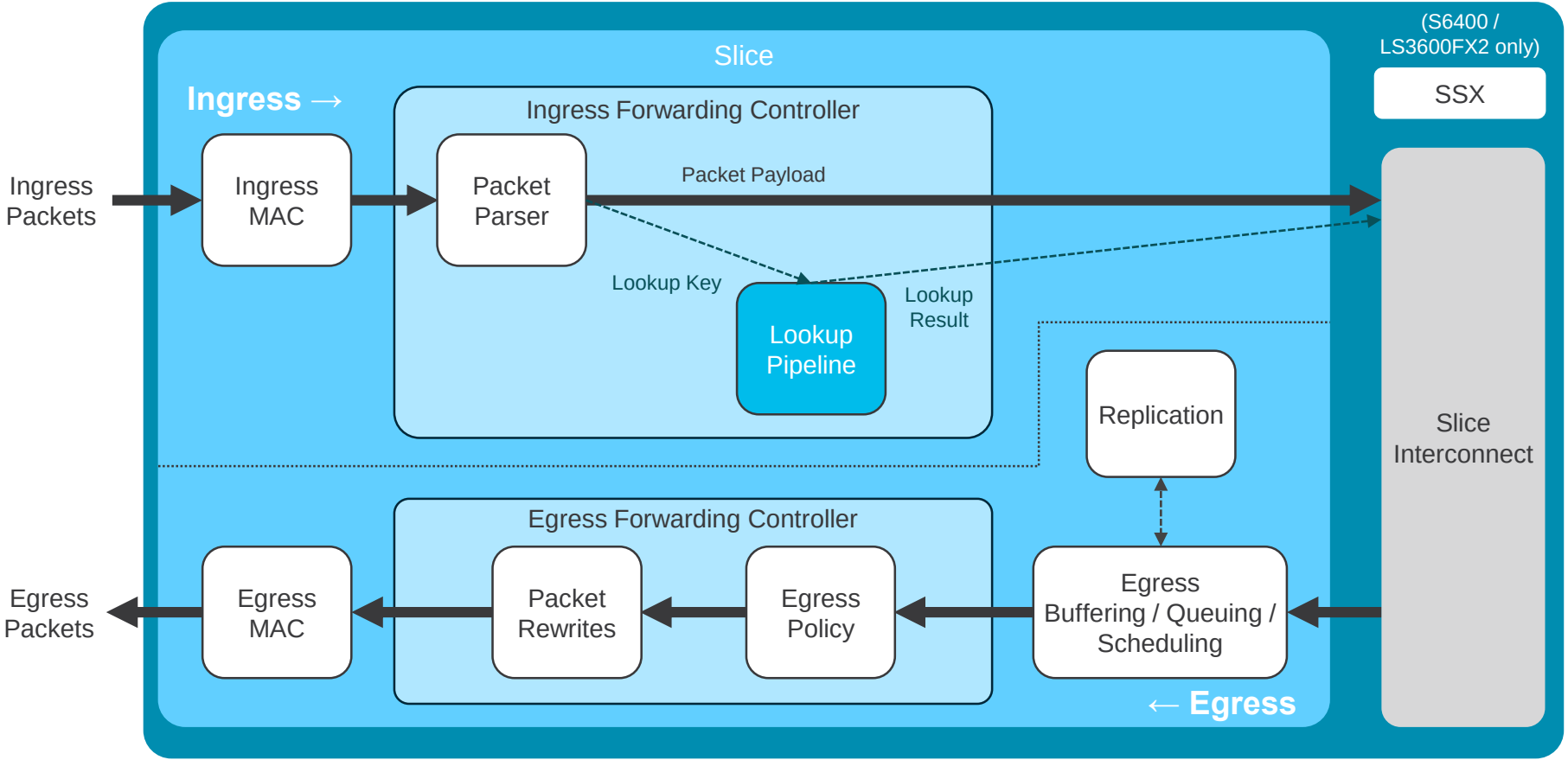
LS3600FX2 – 36 x 100G

What Is a “Slice”?

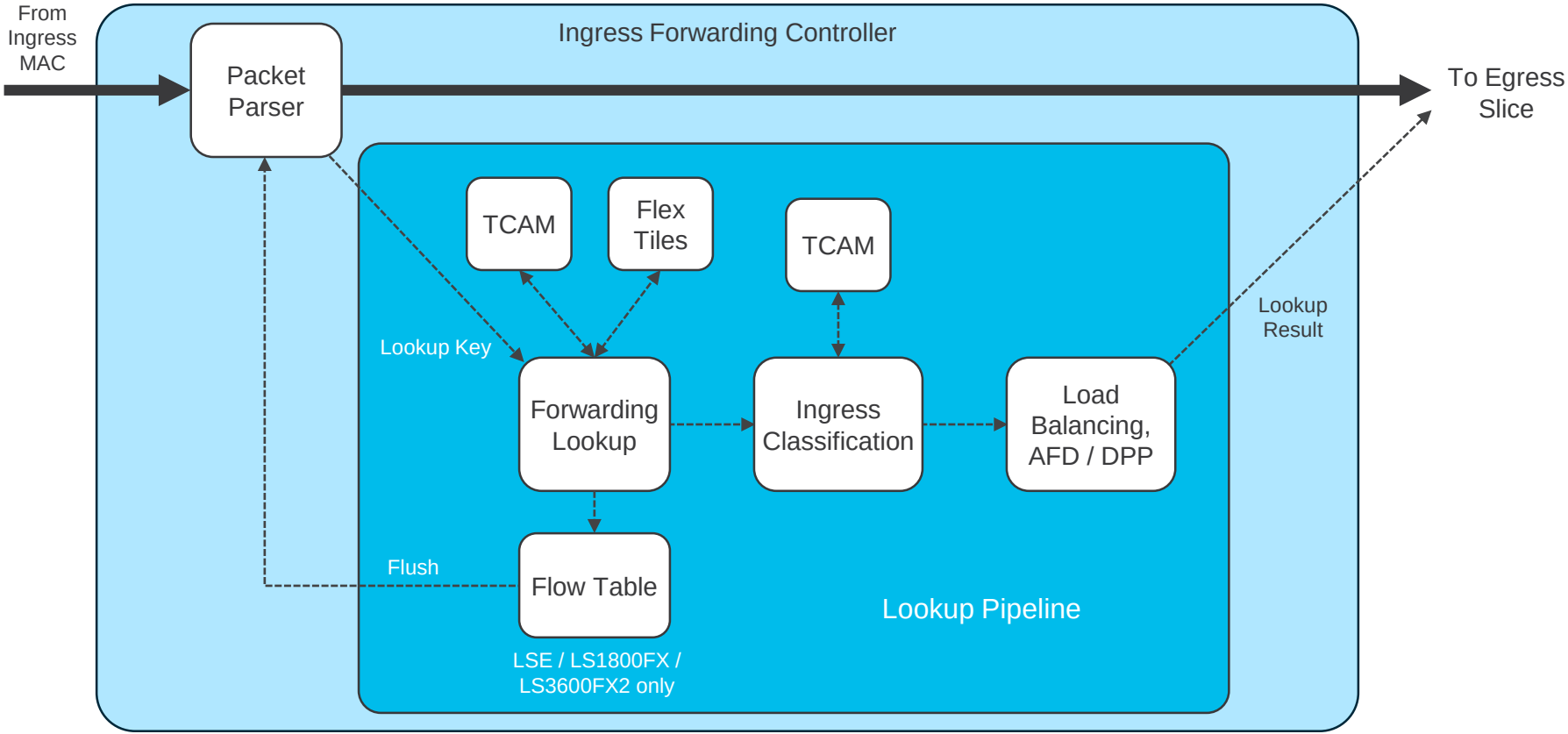
- Self-contained forwarding complex controlling subset of ports on single ASIC
- Separated into Ingress and Egress functions
- Ingress of each slice connected to egress of all slices
- Slice interconnect provides non-blocking any-to-any interconnection between slices



Slice Forwarding Path

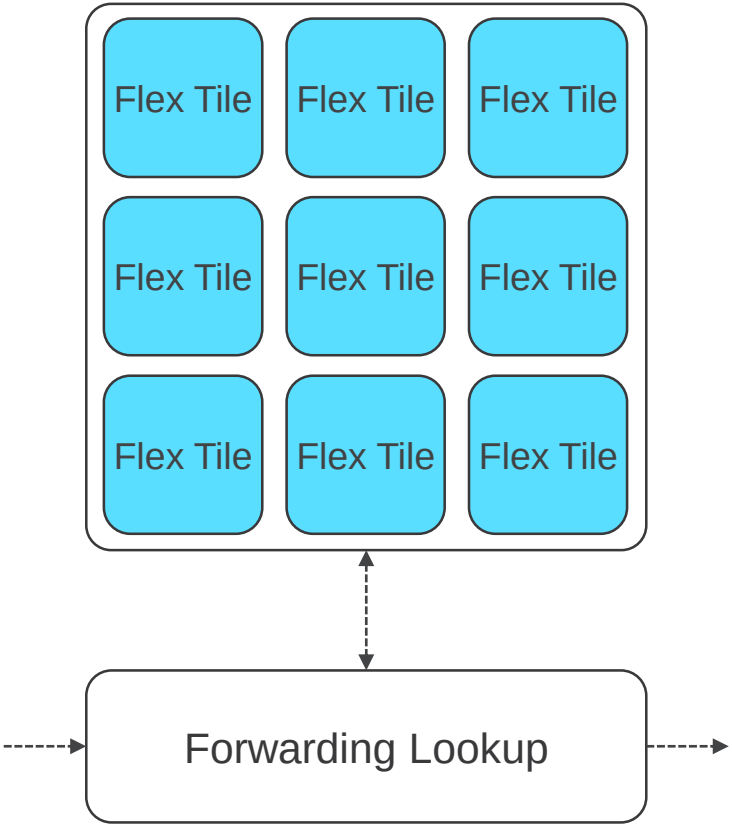


Ingress Lookup Pipeline



Flexible Forwarding Tiles

- Provide fungible pool of table entries for lookups
- Number of tiles and number of entries in each tile varies between ASICs
- Variety of functions, including:
  - IPv4/IPv6 unicast longest-prefix match (LPM)
  - IPv4/IPv6 unicast host-route table (HRT)
  - IPv4/IPv6 multicast (*,G) and (S,G)
  - MAC address/adjacency tables
  - ECMP tables
  - ACI policy



Flex Tile Routing Templates

- Configurable forwarding templates determine flex tile functions
  - “system routing template” syntax
- Templates as of NX-OS 7.0(3)I7(2):
  - Default
  - Dual-stack host scale*†
  - Internet peering*
  - LPM heavy
  - MPLS heavy*
  - Multicast heavy
  - Multicast NBM**
- Defined at system initialisation – reboot required to change profile

* Template does not support IP multicast

† Template not supported on modular Nexus 9500

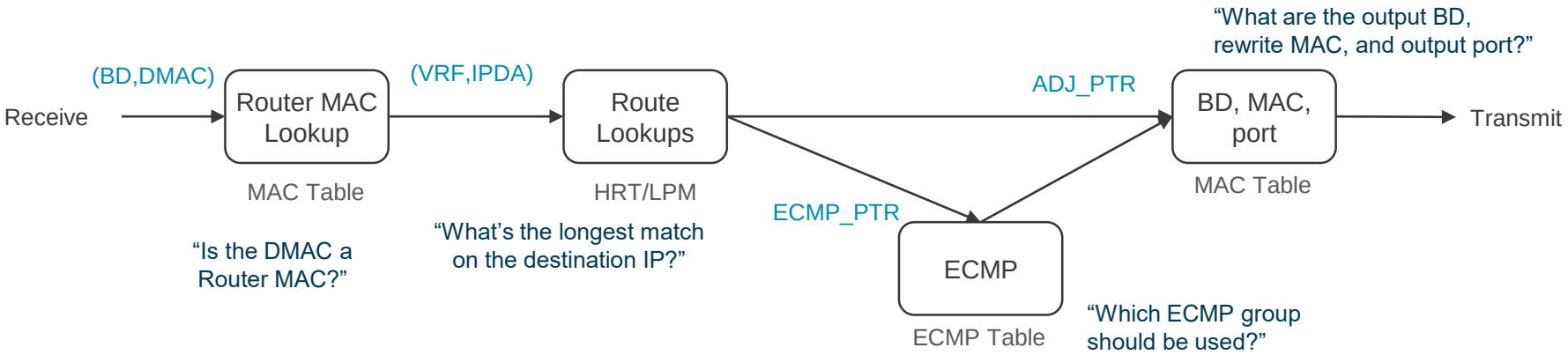
** Template not supported on TORs

Agenda

- Data Centre and Silicon Strategy
- Cloud Scale Architecture
  - Cloud Scale ASICs
  - Forwarding and Features
- Cloud Scale Switching Platforms
- Optics and What’s Next

IP Unicast Forwarding

- Router MAC match triggers L3 lookup
- Hardware performs exact-match on VRF and longest-match on IPDA
- Lookup result returns either adjacency pointer (index into MAC table), or ECMP pointer
- MAC table has output BD, rewrite MAC, and output port



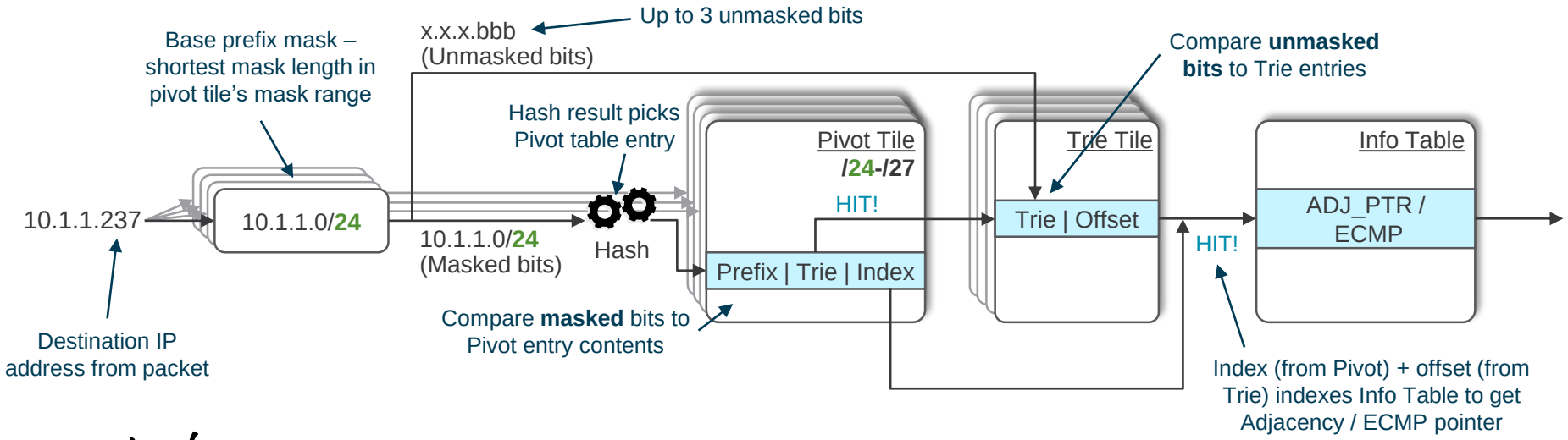
IP Tables

Several methods for storing IP prefixes in hardware:

- **HRT** – Hash table used for IPv4 /32 and IPv6 /128 host entries
  - Provisioned from flex tiles
- **LPM** – Traditional prefix/mask entries, or combination of “pivot” and “trie” tiles, used for other prefix lengths
  - Provisioned from flex tiles
- **TCAM** – Handles overflow/hash collisions
  - Traditional TCAM memory, front-ending flexible forwarding lookups

Pivot / Trie Tiles for Scaling LPM

- "**Pivot**" tiles are hash tables containing base prefixes – match "base mask" bits
- "**Trie**" tiles contain leaf entries for corresponding pivots – match up to 3 least-significant prefix bits
- Combination of pivot and trie lookups returns longest-match prefix entry and adjacency pointer



Trie Tile Lookup

Pivot tile mask range: ***/24-/27***

Don't care

IP from packet →	**10.1.1.237**
Binary →	**00001010.00000001.00000001.11101101**

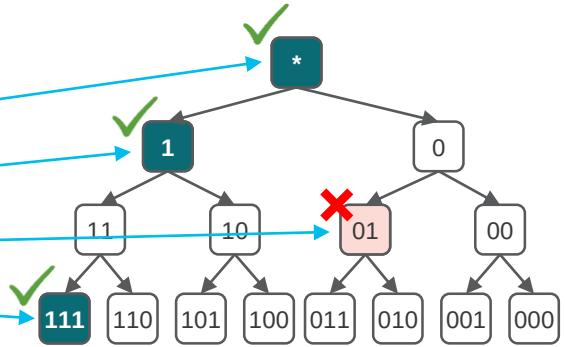
Matched using hash lookup in pivot table (24 bits)

Matched using trie lookup (3 bits)

Prefixes in FIB:

IP Prefix	Binary
10.1.1.0/24	**00001010.00000001.00000001.**********
10.1.1.128/25	**00001010.00000001.00000001.1*********
10.1.1.64/26	**00001010.00000001.00000001.01********
10.1.1.224/27	**00001010.00000001.00000001.111********

Trie lookup matches on these 3 bits

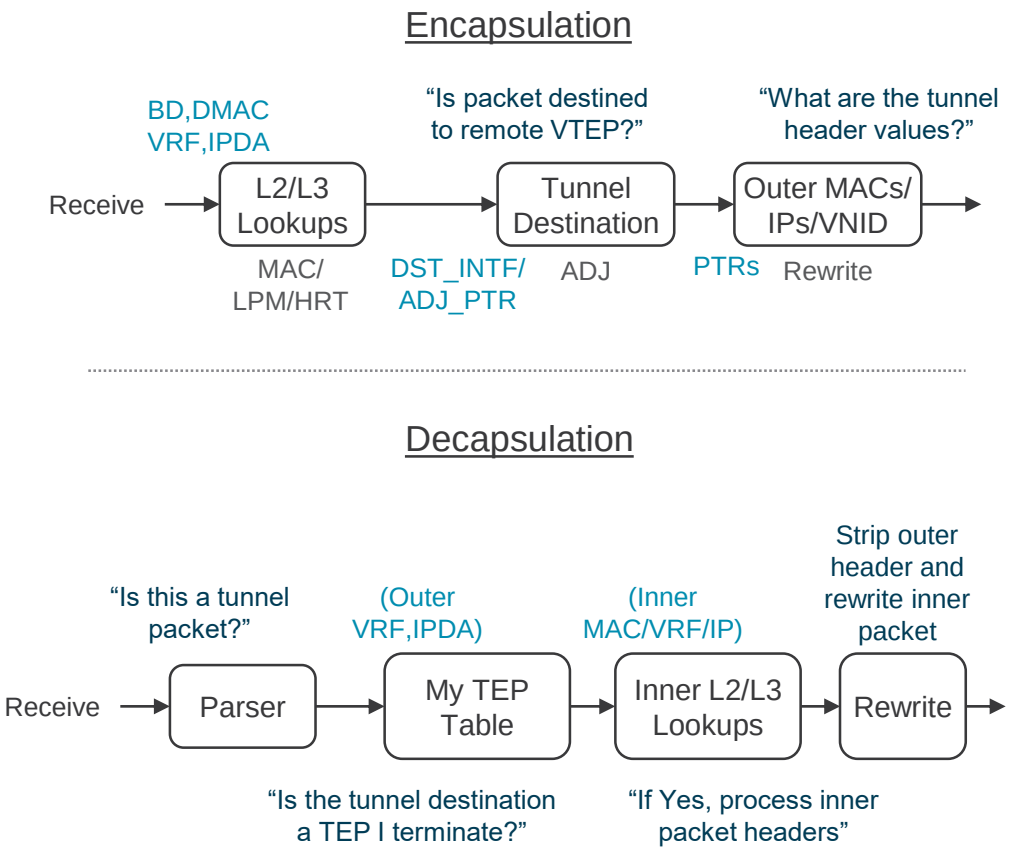


Trie bitmap: 1100000010000000

- Trie tiles contain leaf entries for corresponding pivots
- Up to 15 prefixes can be packed into one trie entry
- Much more efficient than consuming one table entry per prefix

VXLAN Forwarding

- VXLAN and other tunnel encapsulation/ decapsulation performed in single pass
- Encapsulation
  - L2/L3 lookup drives tunnel destination
  - Rewrite block drives outer header fields (tunnel MACs/IPs/VNID, etc.)
- Decapsulation
  - Packet parser determines whether and what type of tunnel packet
  - Forwarding pipeline determines whether tunnel is terminated locally, drives inner lookups



Load Sharing

Equal-Cost Multipath (ECMP)

- Static flow-based load-sharing
- Picks ECMP next-hop based on hash of packet fields and universal ID
  - Source / destination IPv4 / IPv6 address (L3)
  - Source / destination TCP / UDP ports (L4)
  - L3 + L4 (default)
  - GRE key field

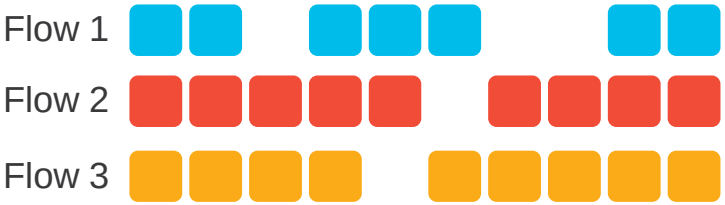
Dynamic Load-Balancing (DLB)

- Supported on leaf switches in ACI fabric
- Congestion aware, flow-based or flowlet-based – rebalances flows/flowlets based on path congestion

Flow Versus Flowlet

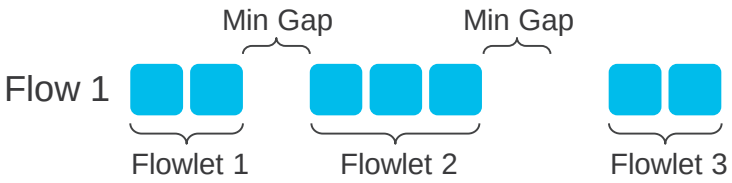
Flow

- 5-tuple of packet values
- All packets traverse same path
- Different flows may traverse different paths

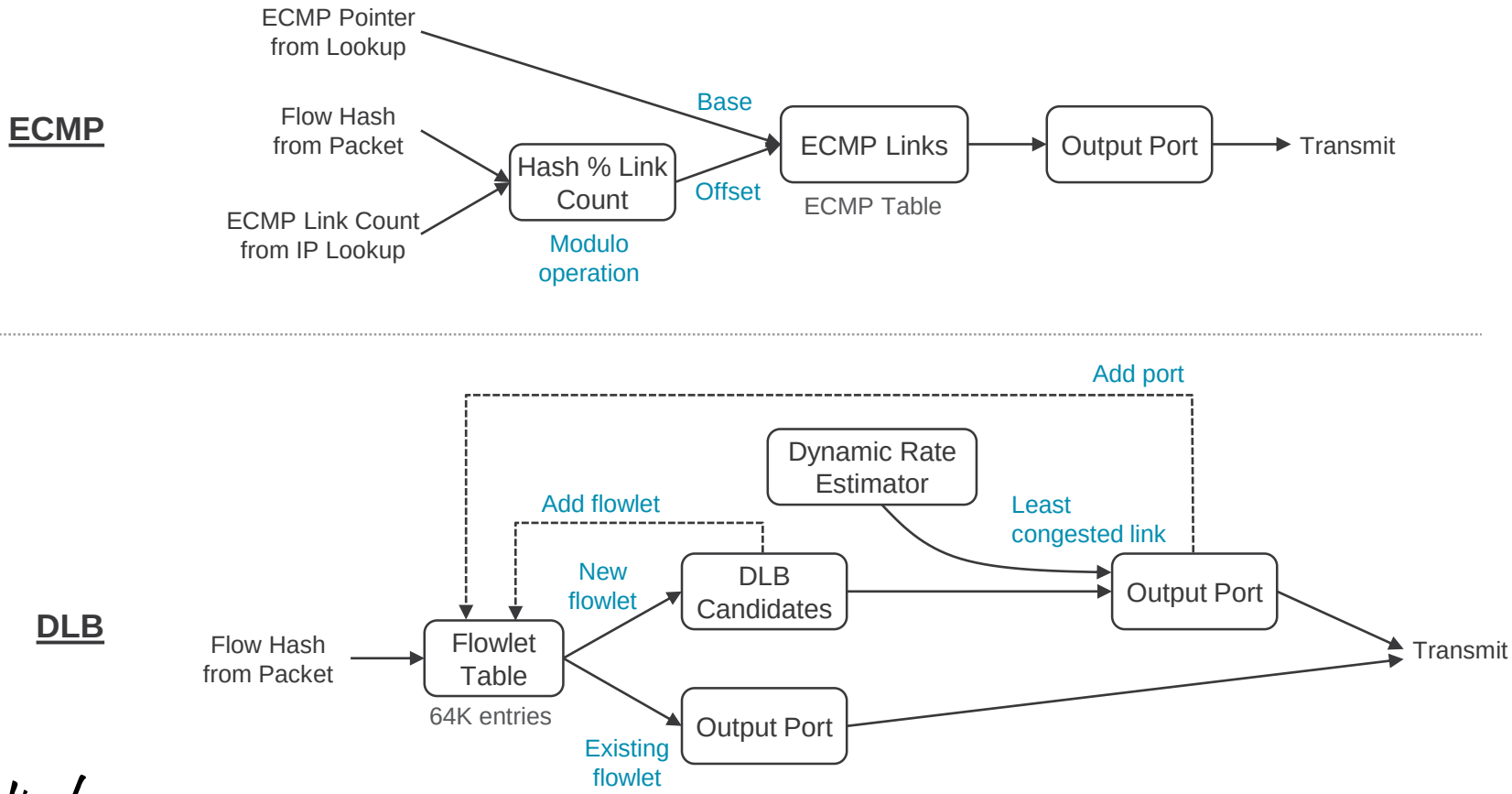


Flowlet

- Series of back-to-back packets of 5-tuple flow
- Gap of a minimum period between packets represents flowlet boundary
- Different flowlets may traverse different paths

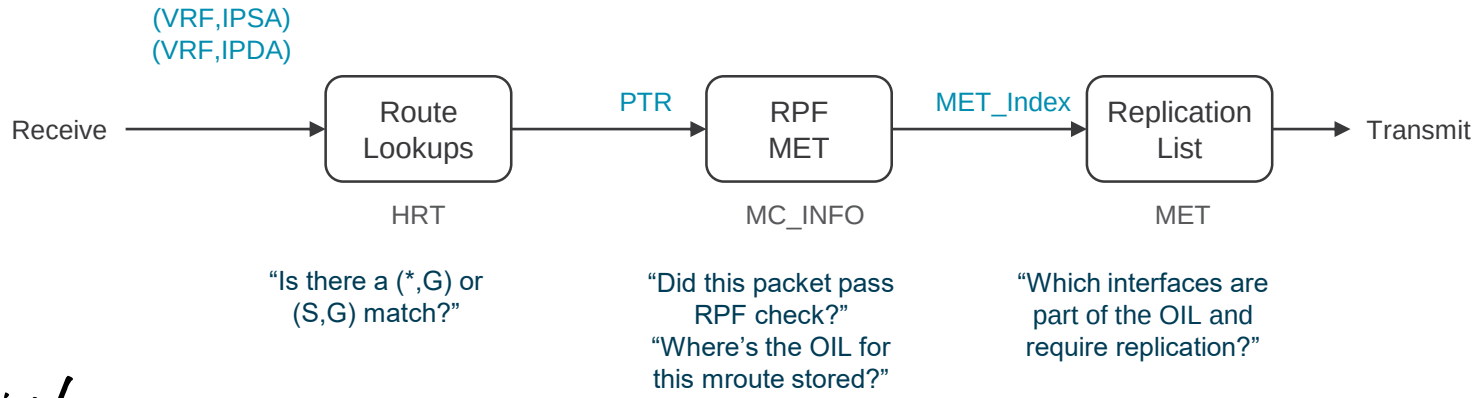


ECMP Versus DLB Load-Sharing



Multicast Forwarding

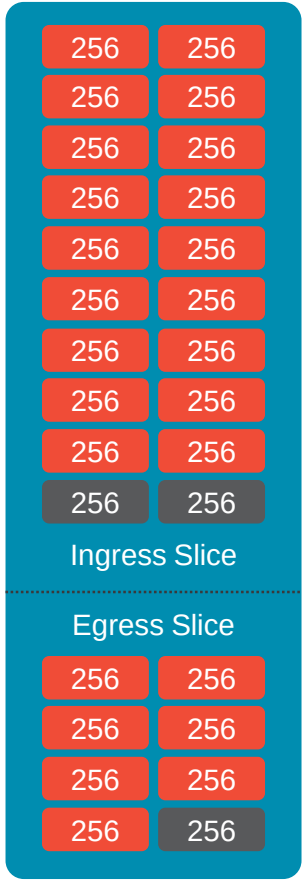
- Multicast source and group forwarding entries populated in HRT
- Additional, secondary table for multicast also provisioned (“MC_INFO”) from flex tiles
- MET table in egress slice holds output interface list (OIL)
- Replication is single copy, multiple reads



Classification TCAM

- Dedicated TCAM for packet classification
- Capacity varies depending on platform
- Leveraged by variety of features:
  - RACL / VACL / PACL
  - L2/L3 QOS
  - SPAN / SPAN ACL
  - NAT
  - COPP
  - Flow table filter (LS1800FX / LS3600FX2)

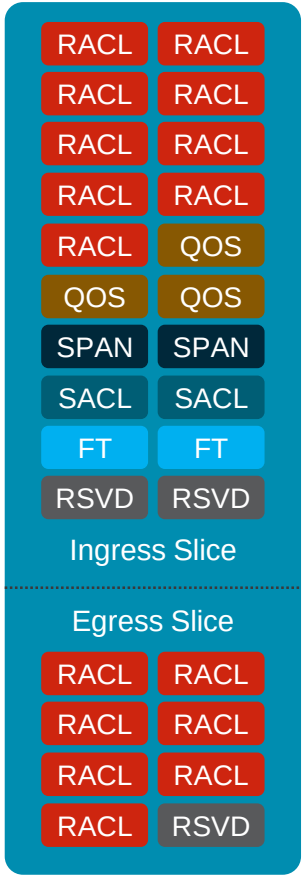
LSE
4K ingress ACEs /
2K egress ACEs



LS1800FX / S6400 / LS3600FX2
5K ingress ACEs /
2K egress ACEs

TCAM Region Resizing

- Default carving allocates 100% of TCAM and enables:
  - Ingress / Egress RACL
  - Ingress QOS
  - SPAN
  - SPAN ACLs
  - Flow table filter (LS1800FX / LS3600FX2 only)
  - Reserved regions
- Based on features required, user can resize TCAM regions to adjust scale
  - To increase size of a region, some other region must be sized smaller
- Region sizes defined at initialisation – changing allocation requires system reboot
  - Configure all regions to desired size (“hardware access-list tcam region”), save configuration, and reload



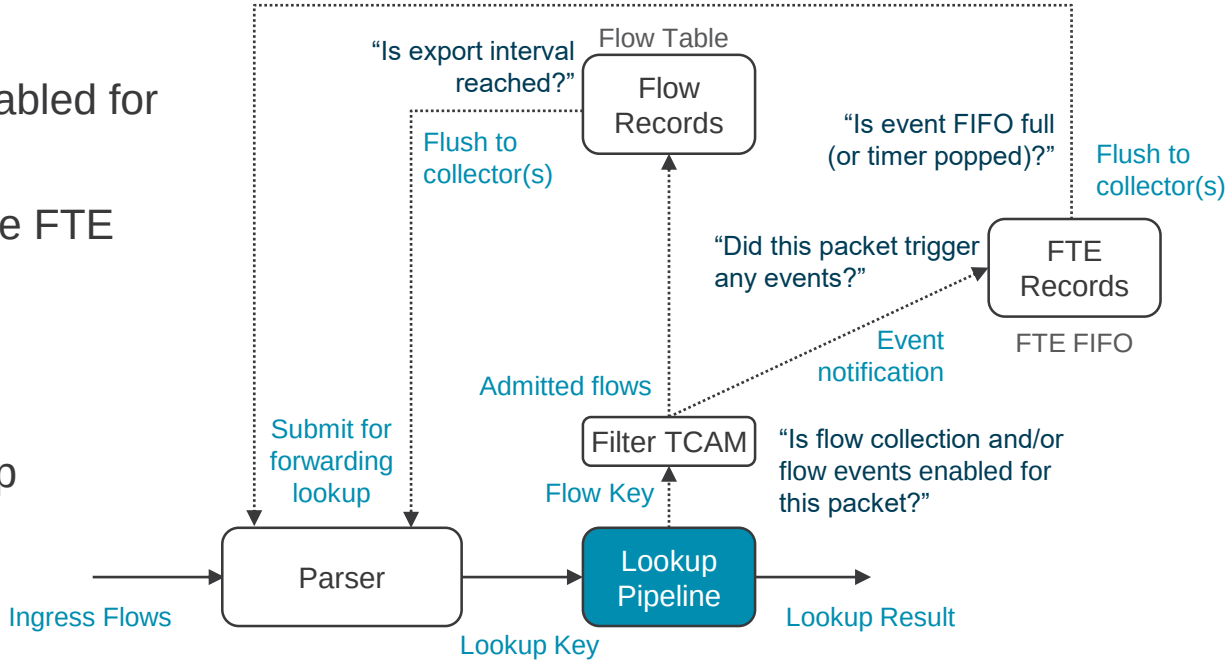
Flow Table / Flow Table Events

- LSE / LS1800FX / LS3600FX2 platforms support hardware flow table logic
- 32K flow table entries per slice + triggered event-based flow data capture
- Collects full flow information plus metadata for:
  - Tetration Analytics
  - Fabric Insights or third-party analytics platform
  - Netflow Data Export v9

Flow Table and Flow Table Events Logic

Flow table / FTE operation for telemetry:

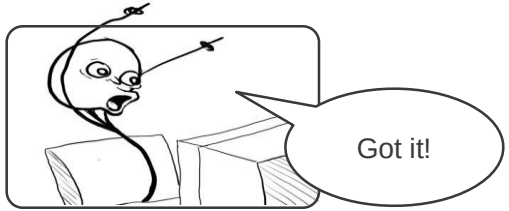
1. Determine if FT/FTE enabled for flow
2. Install FT record; capture FTE records if triggered
3. Flush FT / FTE records, encapsulate in IP/UDP
4. Submit packet for lookup

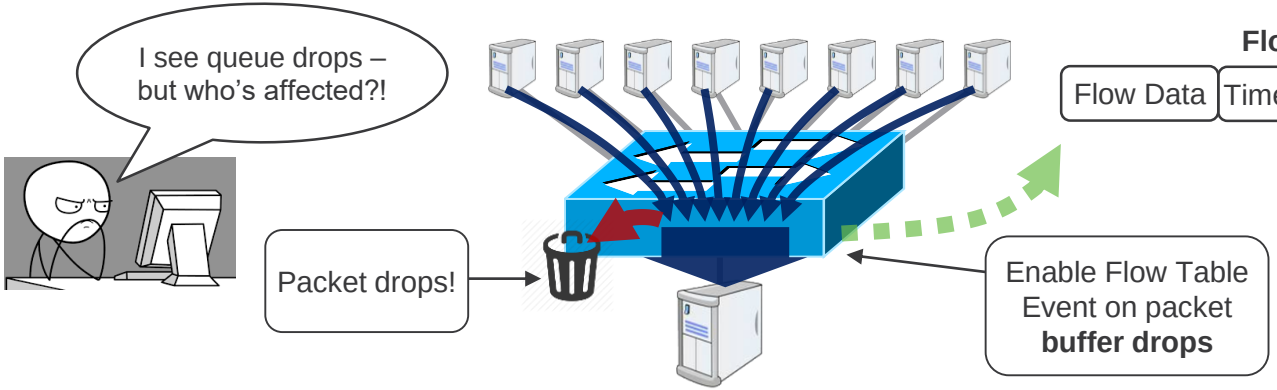


Flow Table Events

Event triggers:

Packet value match	Latency threshold
Buffer drop	Microburst threshold
ACL drop	Forwarding exception





Flow Table Event Record

Flow Data	Timestamp	Port	Queue	**Buffer Drop**

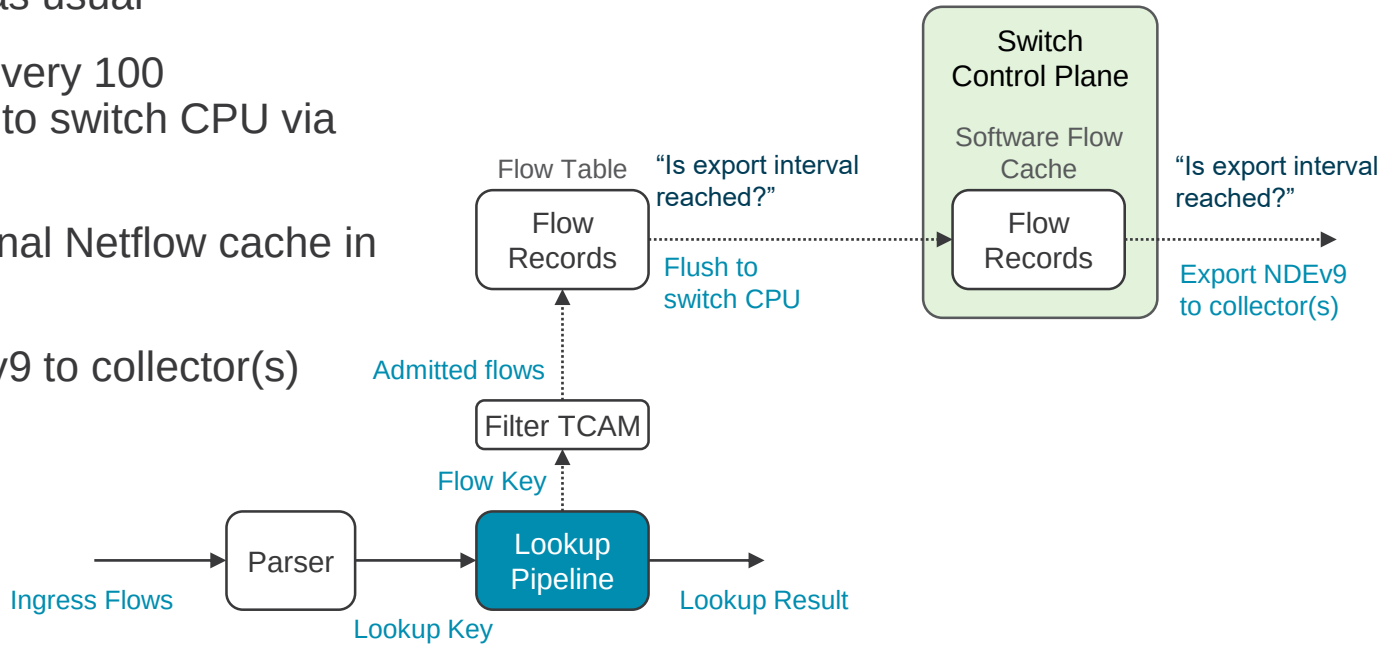
Full Netflow

Netflow v9 support:

- 9300-FX TORs: 7.0(3)I7(1)
- 9300-EX TORs: 7.0(3)I7(2)

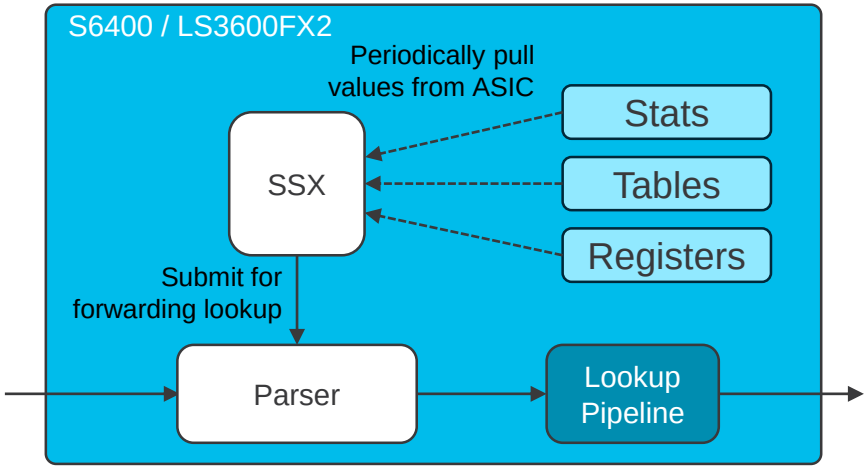
Flow table operation for full Netflow:

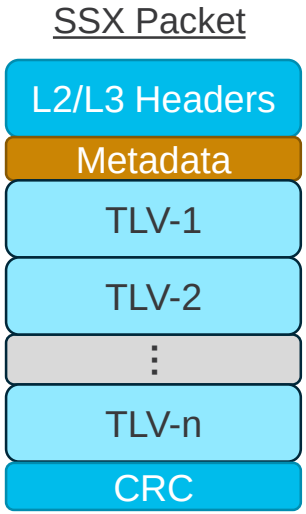
1. Install FT records as usual
2. Flush FT records every 100 milliseconds, send to switch CPU via forwarding pipeline
3. CPU builds traditional Netflow cache in software
4. CPU exports NDEv9 to collector(s) every 10 seconds



Streaming Statistics Export (SSX)

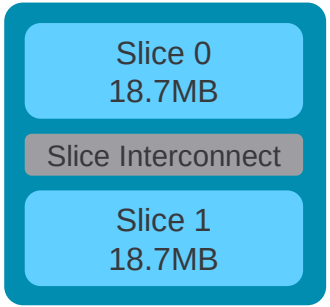
- Streams statistics and other ASIC-level data
- Direct export from ASIC – no switch CPU involvement
- User defines streaming parameters – which statistics, how often, and to which collector
- Hardware support in S6400 / LS3600FX2



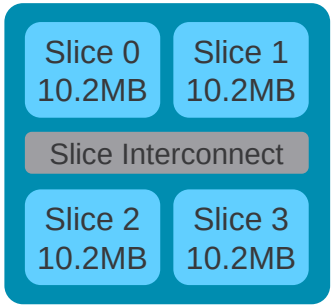


Buffering

- Cloud Scale platforms implement shared-memory egress buffered architecture
- Each ASIC slice has dedicated buffer – only ports on that slice can use that buffer
- Dynamic Buffer Protection adjusts max thresholds based on class and buffer occupancy
- Intelligent buffer options maximise buffer efficiency



LSE
18.7MB/slice
(37.4MB total)



S6400
10.2MB/slice
(40.8MB total)

LS1800FX
40.8MB/slice
(40.8MB total)

LS3600FX2
20MB/slice
(40MB total)

Intelligent Buffering

Innovative Buffer Management for Cloud Scale switches

- **Dynamic Buffer Protection** (DBP) – Controls buffer allocation for congested queues in shared-memory architecture
- **Approximate Fair Drop** (AFD) – Maintains buffer headroom per queue to maximise burst absorption
- **Dynamic Packet Prioritisation** (DPP) – Prioritises short-lived flows to expedite flow setup and completion

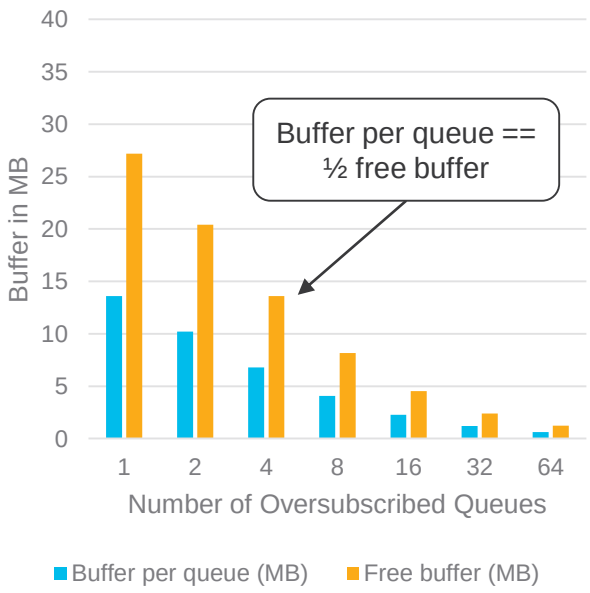
Miercom Report: Speeding Applications in Data Centre Networks
http://miercom.com/cisco-systems-speeding-applications-in-data-Centre-networks/

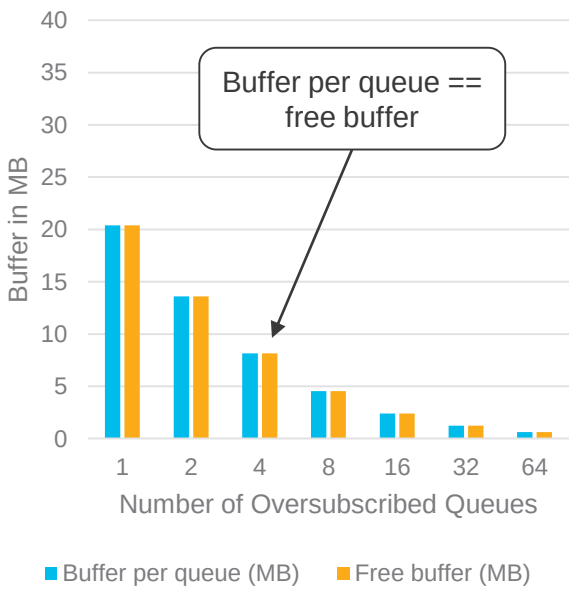
Dynamic Buffer Protection (DBP)

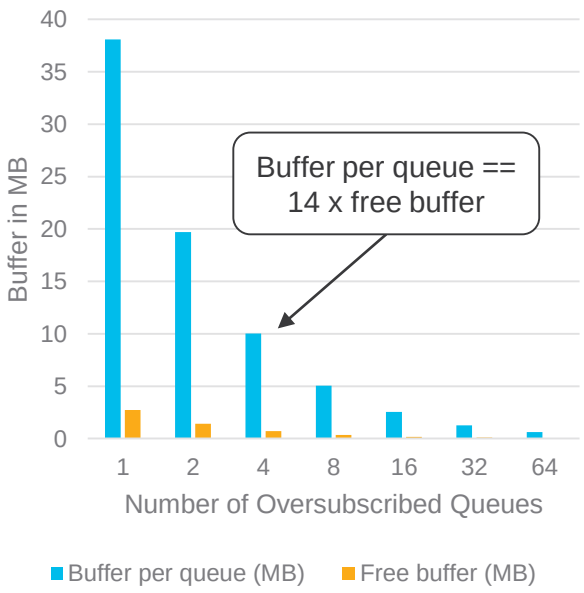
- Prevents any output queue from consuming more than its fair share of buffer in shared-memory architecture
- Defines dynamic max threshold for each queue
  - If queue length less than threshold, packet is admitted
  - Otherwise packet is discarded
- Threshold calculated by multiplying free memory by configurable **Alpha** (α) value (weight)
  - “queue-limit dynamic *alpha-value*” in queuing policy

Alpha Parameter Examples

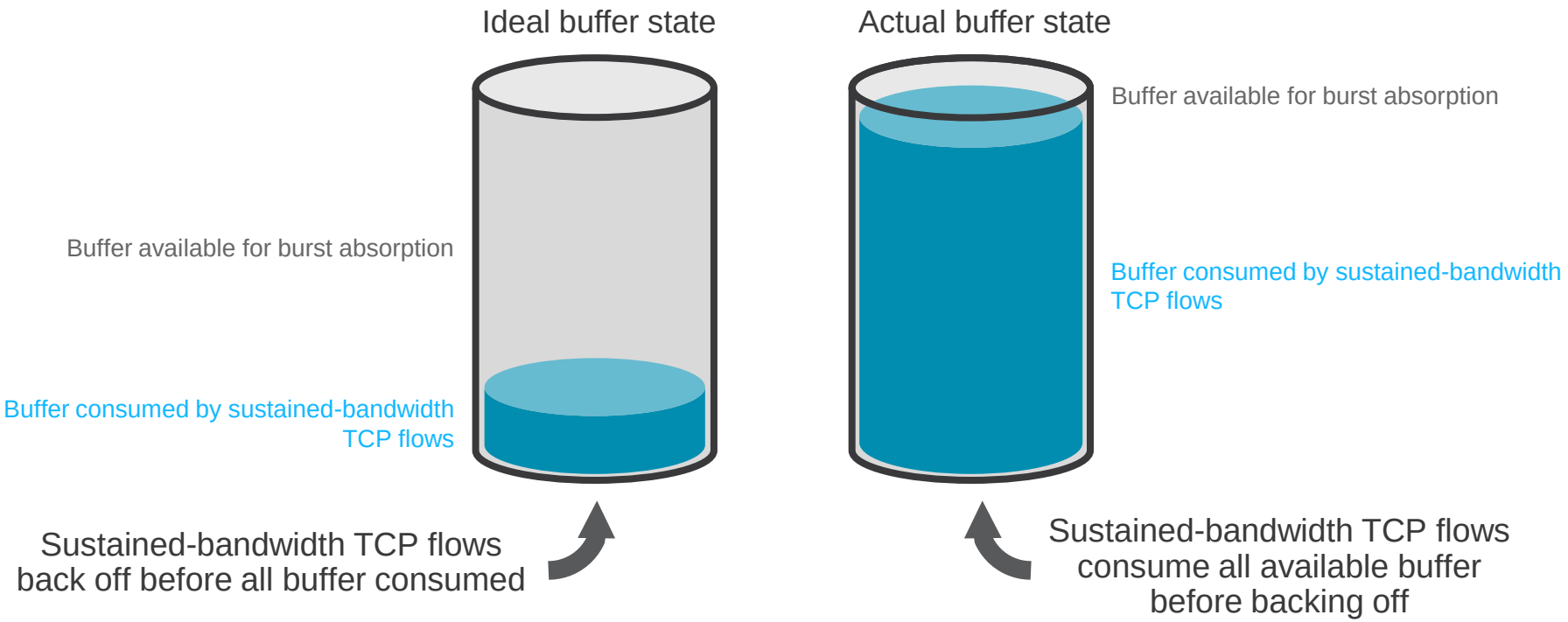
Default Alpha on
Cloud Scale switches







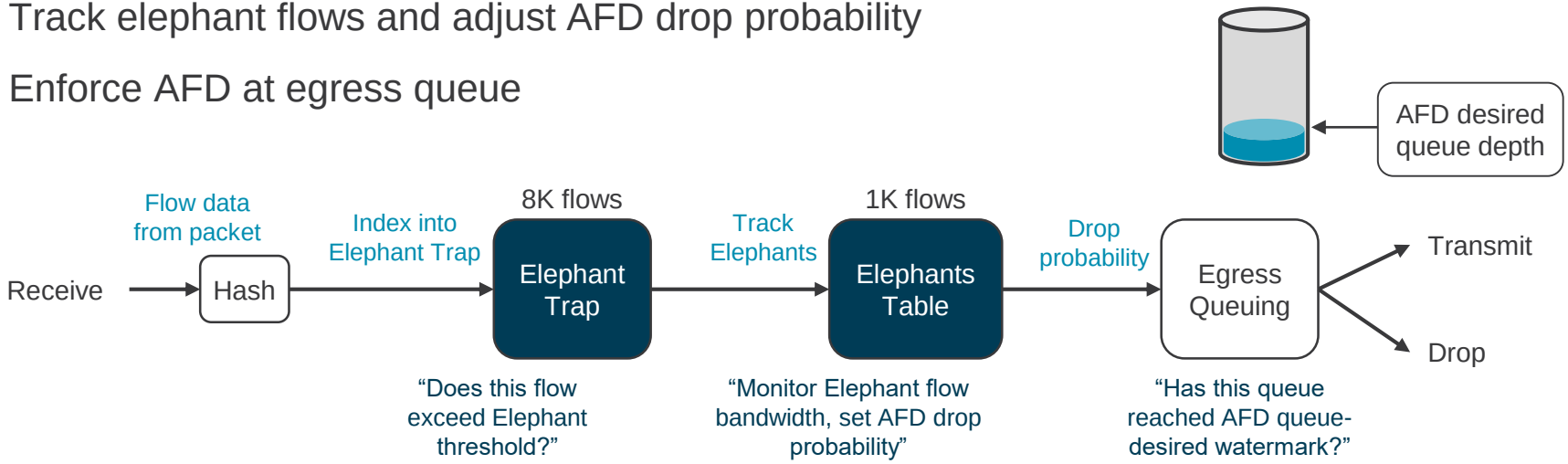
Buffering – Ideal Versus Reality



Approximate Fair Drop (AFD)

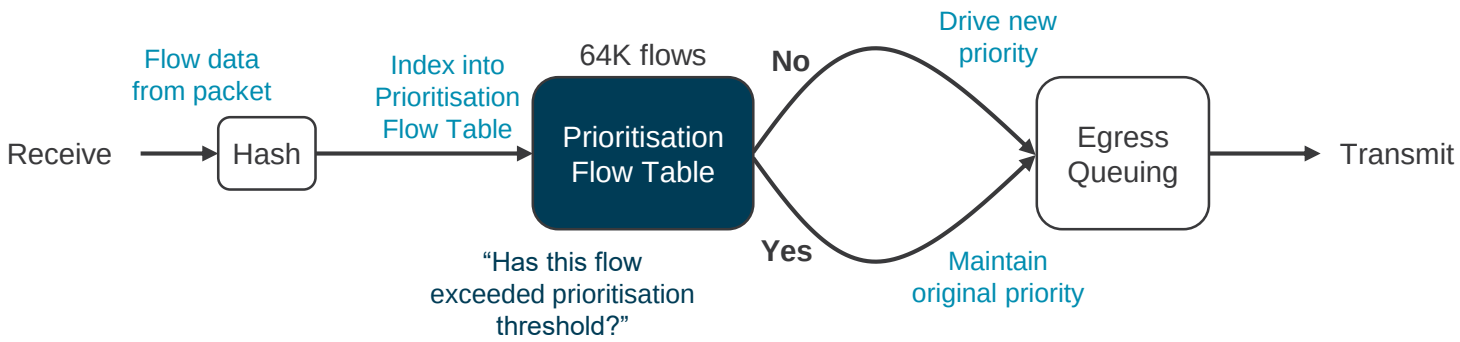
Maintain throughput while minimising buffer consumption by elephant flows – **keep buffer state as close to the ideal as possible**

1. Distinguish elephant flows from other flows
2. Track elephant flows and adjust AFD drop probability
3. Enforce AFD at egress queue

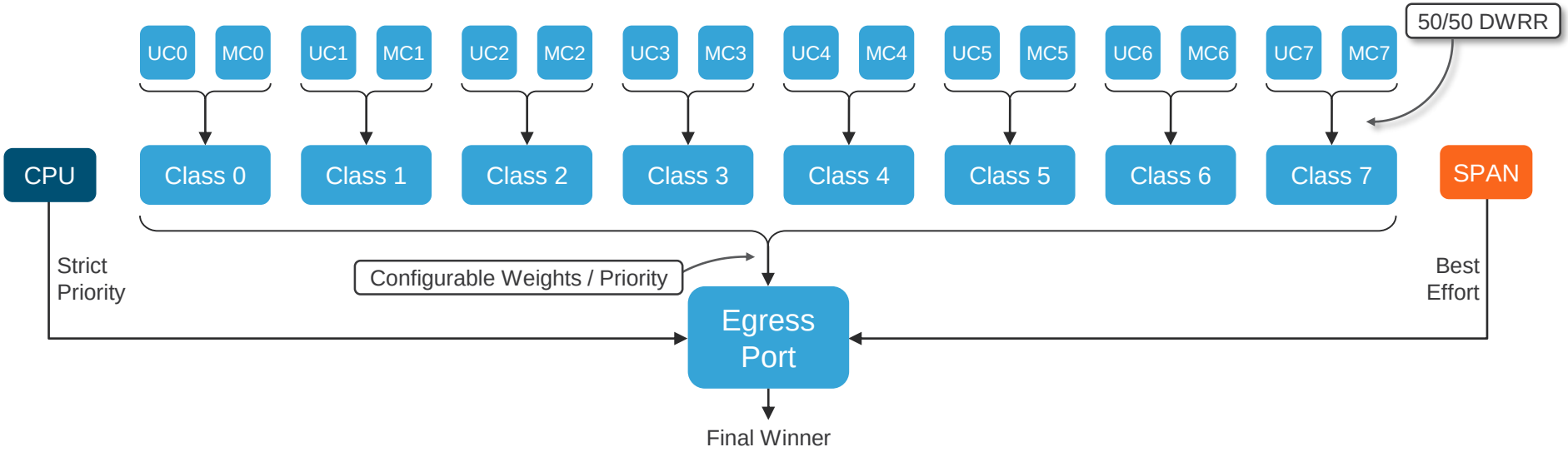


Dynamic Packet Prioritisation (DPP)

- Prioritise initial packets of new / short-lived flows
- Up to first 1K packets assigned to higher-priority qos-group



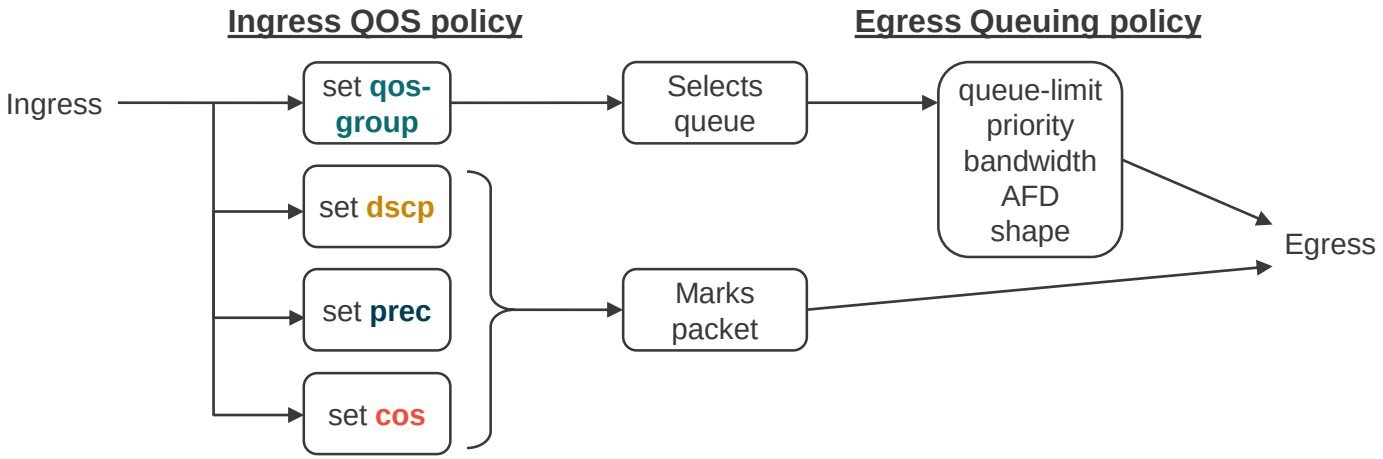
Queuing and Scheduling



- 8 user classes and 16 queues per output port (8 unicast, 8 multicast)
- QOS-group drives class; egress queuing policy defines class priority and weights
- Dedicated classes for CPU traffic and SPAN traffic

Ingress QOS / Egress Queuing Policies

- Default QOS behaviour:
  - All user data goes to q-default
  - **Trust** received QOS markings
- To select egress queue, use “set **qos-group**” in ingress QOS policy
- To set/change packet markings, use “set **cos** / **precedence** / **dscp**” in ingress QOS policy
- To change queuing behaviour, manipulate egress queuing policies



Agenda

- Data Centre and Silicon Strategy
- Cloud Scale Architecture
  - Cloud Scale ASICs
  - Forwarding and Features
- Cloud Scale Switching Platforms
- Optics and What's Next

Cloud Scale Platforms

Nexus 9300-EX and 9300-FX/FX2

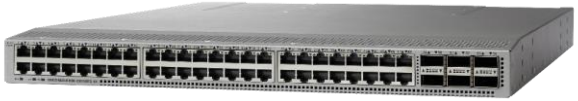
- Premier TOR platforms
- Full Cloud Scale functionality
- ACI leaf / standalone leaf or spine
- FX option with MACSEC using LS1800FX silicon
- FX2 option with key enhancements using LS3600FX2 silicon

Nexus 9500 X9700-EX and X9700-FX Modules

- Switching modules for Nexus 9500 modular chassis
- Full Cloud Scale functionality
- ACI spine / standalone aggregation or spine
- FX option with MACSEC using LS1800FX silicon

Nexus 9300-EX Cloud Scale TOR Switches

48-port 10/25G SFP28 + 6-port 100G QSFP28
N9K-C93180YC-EX – LSE-based
ACI: 1.3(1)
NX-OS: 7.0(3)I4(2)



48-port 1/10GBASE-T + 6-port 100G QSFP28
N9K-C93108TC-EX – LSE-based
ACI: 2.0(1)
NX-OS: 7.0(3)I4(2)

32-port 40G/50G/100G QSFP28
N9K-C93180LC-EX – LSE-based
ACI: 2.2(1)
NX-OS: 7.0(3)I6(1)

Key Features

Dual capability – **ACI and NX-OS mode**

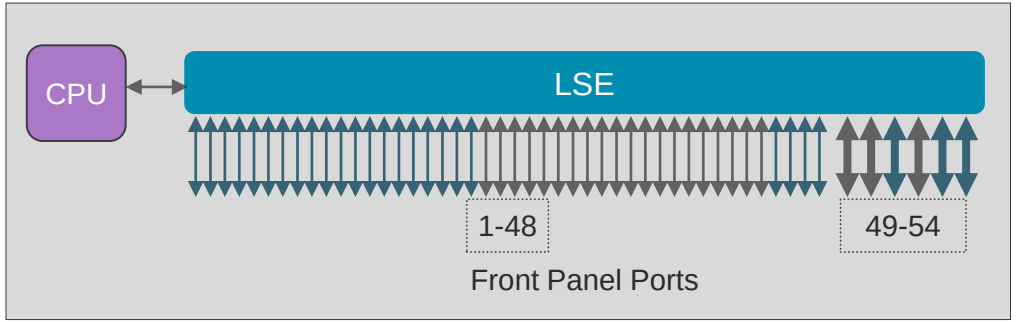
Flexible port configurations – 1/10/25/40/50/100G

Native 25G server access ports

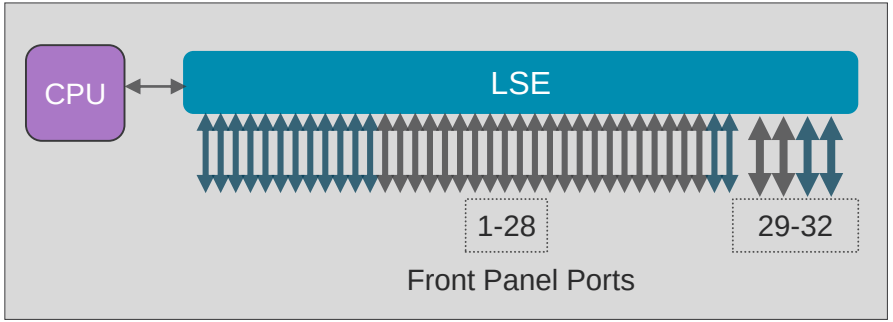
Flow Table / FTE for Tetration Analytics, Fabric Insights, Netflow

Smart buffer capability (AFD / DPP)

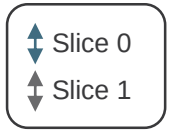
Nexus 9300-EX Switch Architectures



C93180YC-EX (10/25G + 100G) /
C93108TC-EX (10G + 100G)

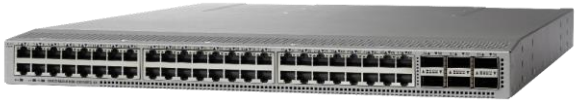


C93180LC-EX (40/50G + 100G)



Nexus 9300-FX Cloud Scale TOR Switches – Pervasive MACSEC

48-port 10/25G SFP28 + 6-port 100G QSFP28
N9K-C93180YC-FX – LS1800FX-based
ACI: 2.2(2e)
NX-OS: 7.0(3)I7(1)



48-port 1/10GBASE-T + 6-port 100G QSFP28
N9K-C93108TC-FX – LS1800FX-based
ACI: 2.2(2e)
NX-OS: 7.0(3)I7(1)

48-port 100M/1GBASE-T + 4-port 10G/25G + 2-port 100G QSFP28
N9K-C9348GC-FXP – LS1800FX-based
ACI: 3.0(1)
NX-OS: 7.0(3)I7(1)

Key Features

Dual capability – **ACI and NX-OS mode**

Flexible port configurations – 100M/1/10/25/40/50/100G

Line-rate 256-bit encryption on all ports

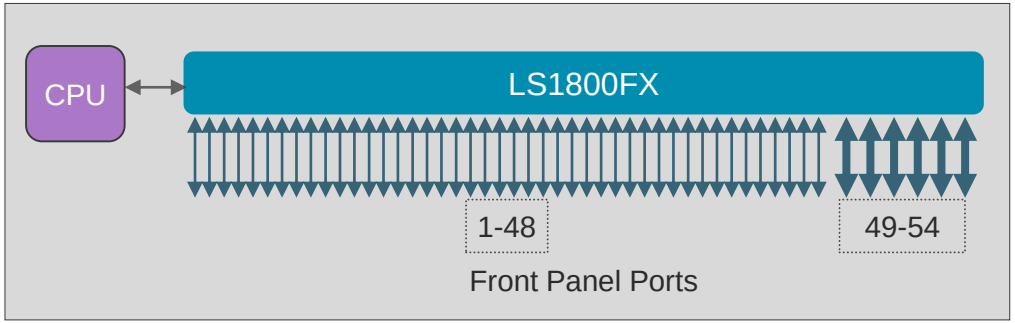
32G FC support on all SFP ports

25G distances beyond 3m (RS-FEC)

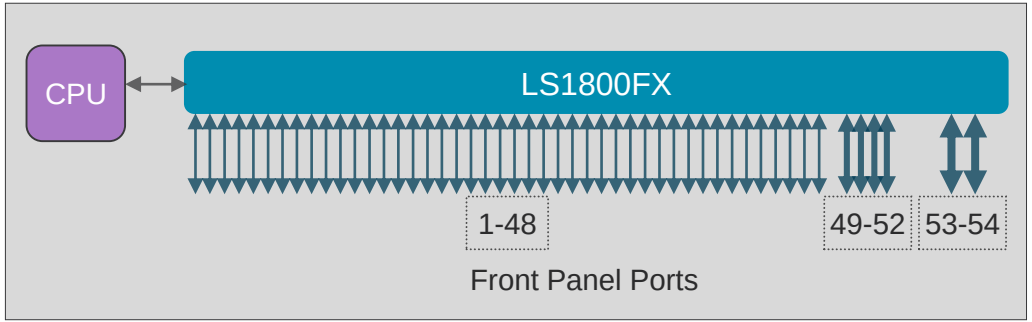
Flow Table / FTE for Tetration Analytics, Fabric Insights, Netflow

Smart buffer capability (AFD / DPP)

Nexus 9300-FX Switch Architectures



C93180YC-FX (10/25G + 100G) /
C93108TC-FX (10G + 100G)



C9348GC-FXP (100M/1G + 10/25G + 100G)

Nexus 9364C 100G Cloud Scale Switch

64-port 100G QSFP28 + 2-port 10G SFP+

N9K-C9364C – S6400-based
ACI: Roadmap
NX-OS: 7.0(3)I7(2)

Key Features

Dual capability – **ACI and NX-OS mode**

Compact, high-performance fixed ACI spine

100G/50G/40G/10G (single port mode – no breakout)

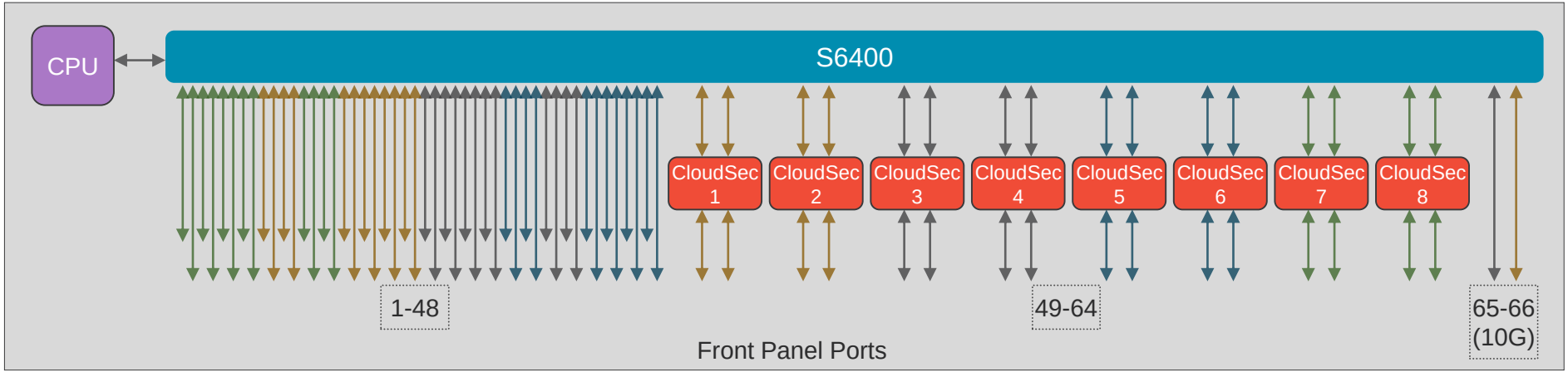
2 x 100M/1G/10G SFP+ ports

MACSEC/CloudSec on 16 ports

Streaming Statistics Export (SSX)

Smart buffer capability (AFD / DPP)

Nexus 9364C Switch Architecture



C9364C (100G + 10G)

Nexus 9300-FX2 Cloud Scale TOR Switches

36-port 100G QSFP28
N9K-C9336C-FX2 – LS3600FX2-based
ACI/NX-OS: Roadmap

48-port 10/25G SFP28 +
12-port 100G QSFP28
N9K-C93240YC-FX2 – LS3600FX2-based
NX-OS: Roadmap

Key Features

Dual capability – **ACI and NX-OS mode**

Versatile standalone 100G switch

Compact, high-performance fixed ACI spine

100G/50G/40G/10G with breakout capability

Flow Table / FTE for Tetration Analytics, Fabric Insights, Netflow

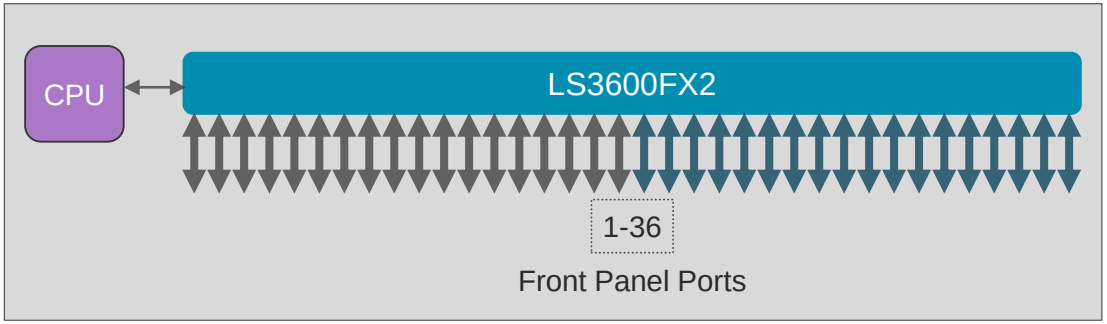
Streaming Statistics Export (SSX)

MACSEC/CloudSec on all ports

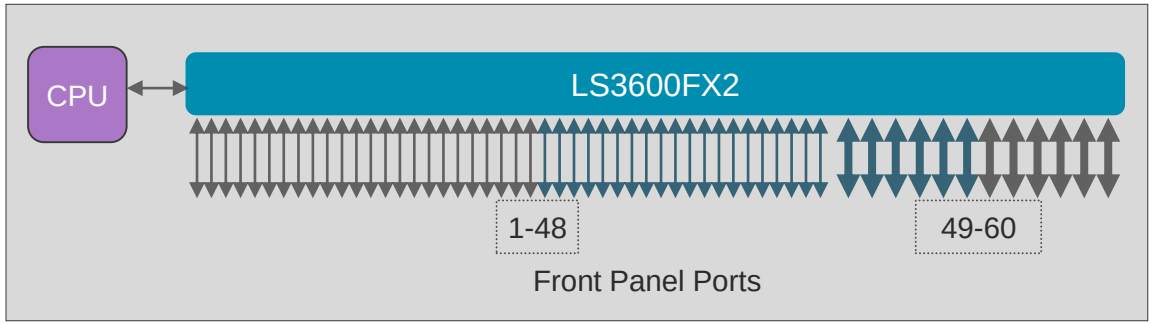
VXLAN ESI multi-homing

Smart buffer capability (AFD / DPP)

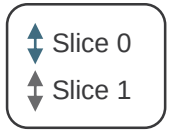
Nexus 9300-FX2 Switch Architecture



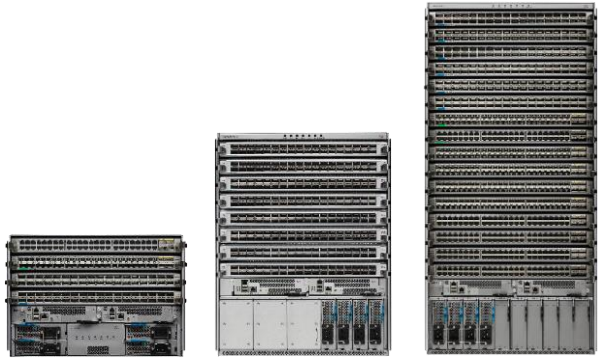
C9336C-FX2 (100G)



C93240YC-FX2 (10/25G + 100G)



Nexus 9500 Modular Cloud Scale Switches



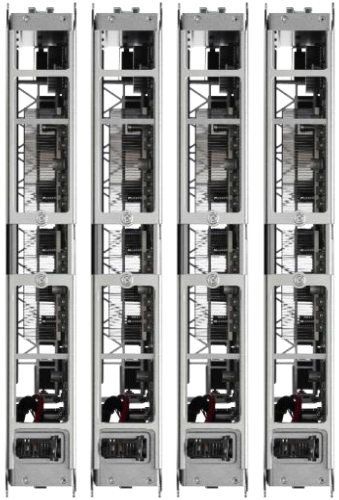
Nexus 9504 Nexus 9508 Nexus 9516

Common Equipment

+

EX / FX Series Line Cards

+



E-Series Fabric Modules

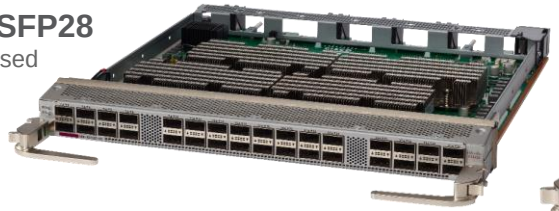
X9700-EX 100G Cloud Scale Modules

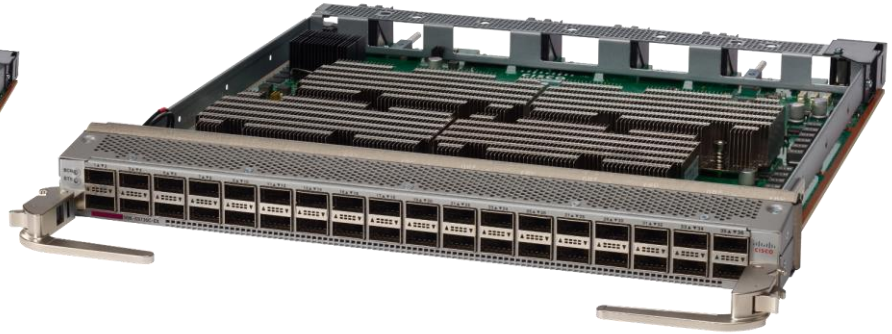
N9K-X9732C-EX / N9K-X9736C-EX

Line-rate performance up to 3.2Tbps capacity

Advanced features –
- Smart buffer capability (AFD / DPP)
- Flexible forwarding tables
- VXLAN routing

32-port 100G QSFP28
X9732C-EX – LSE-based
ACI: 1.3(1)
NX-OS: 7.0(3)I4(2)



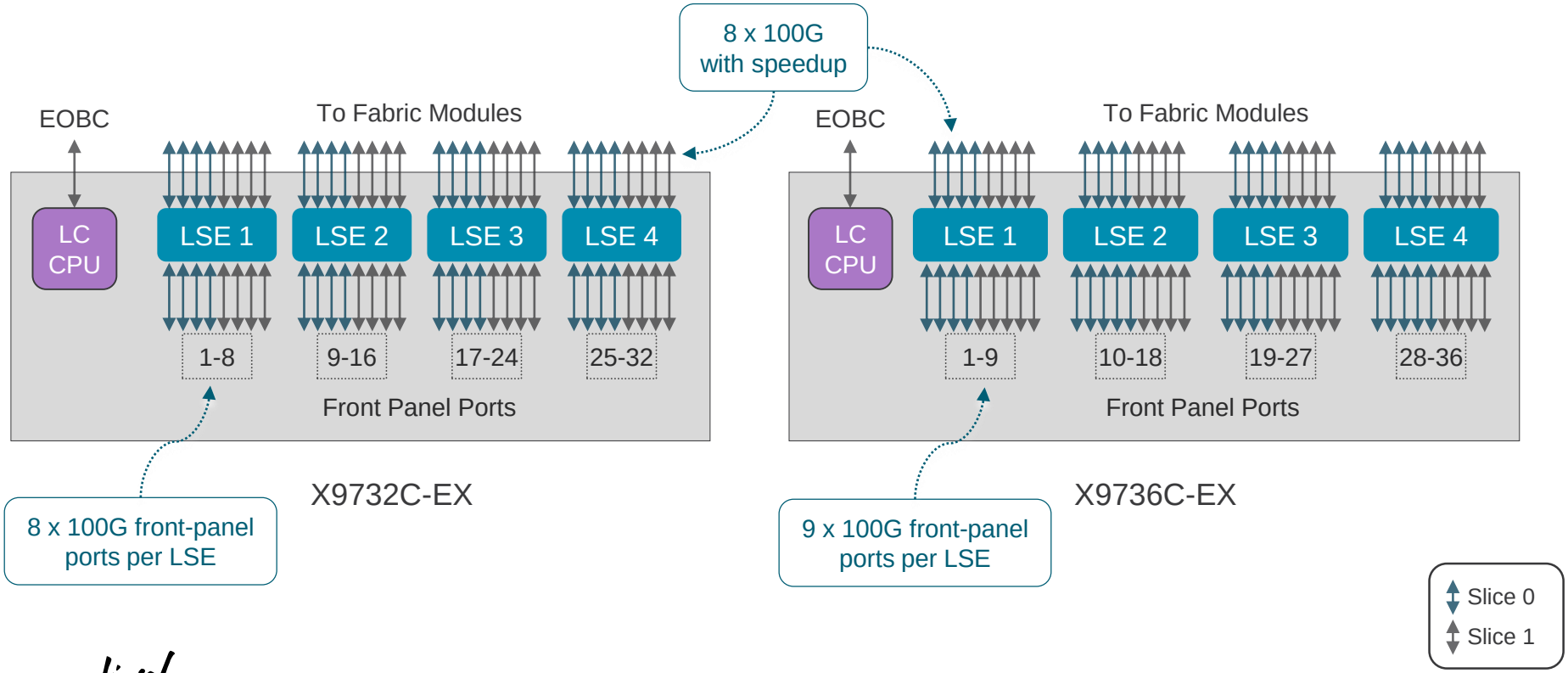


32 / 36 x QSFP28-based 100G ports
- Pin-compatible with 40G QSFP+
- Flexible speed ports – 1 / 10 / 25 / 40 / 50 / 100G capability

Supported in ACI and NX-OS standalone mode

36-port 100G QSFP28
X9736C-EX – LSE-based
ACI: Roadmap
NX-OS: 7.0(3)I6(1)

N9K-X9732C-EX / N9K-X9736C-EX Architecture

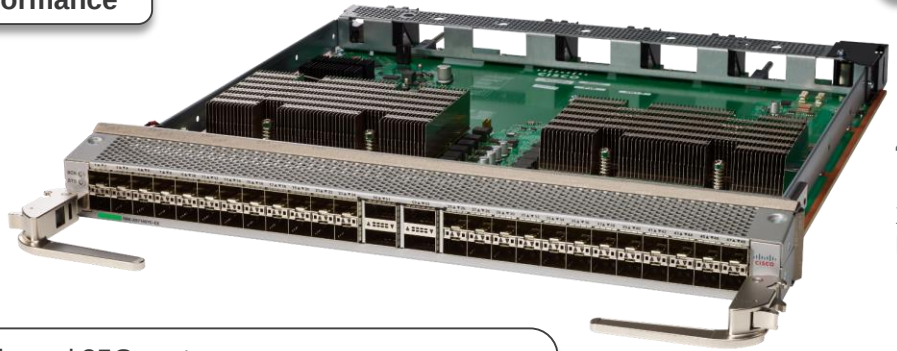


X9700-EX 10/25G + 100G Cloud Scale Module

N9K-X97160YC-EX

Advanced features –
- Smart buffer capability (AFD / DPP)
- Flexible forwarding tables
- VXLAN routing

1.6Tbps capacity with line-rate performance



48p 10/25G SFP+ and 4p 100G QSFP28

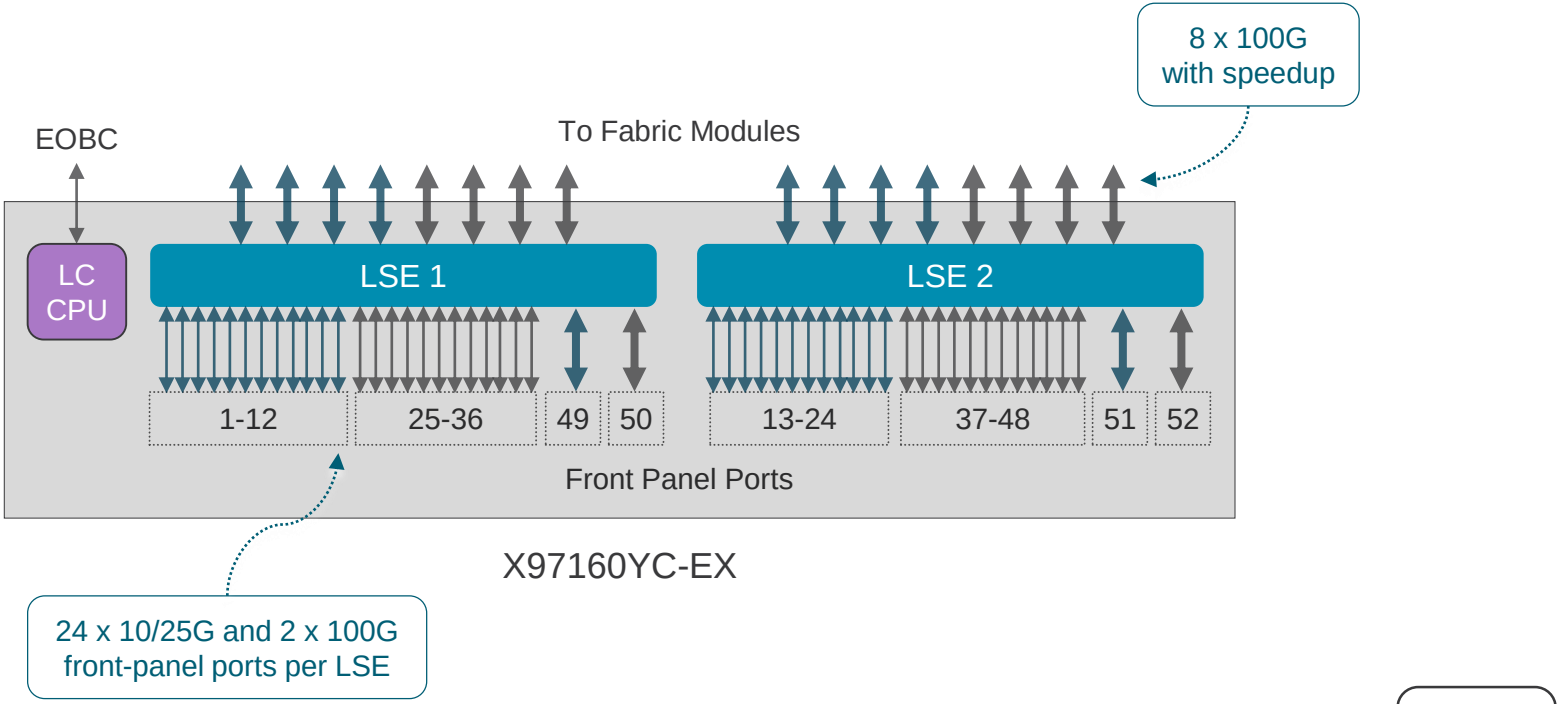
X97160YC-EX – LSE-based
NX-OS: 7.0(3)I5(2)

48 x SFP28-based 25G ports
- Pin-compatible with 1G SFP and 10G SFP+
- Flexible speed ports – 1 / 10 / 25G capability

4 x QSFP28-based 100G ports
- Pin-compatible with 40G QSFP+
- Flexible speed ports – 1 / 10 / 25 / 40 / 50 / 100G capability

Supported in NX-OS standalone mode only

N9K-X97160YC-EX Architecture

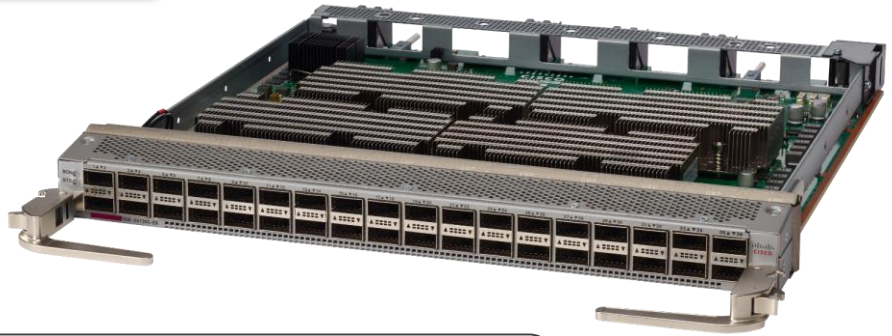


X9700-FX 100G Cloud Scale Module

N9K-X9736C-FX

Advanced features –
- **Line-rate MACSEC on all ports**
- **CloudSec encryption (8 ports)**
- Smart buffer capability (AFD / DPP)
- Flexible forwarding tables
- VXLAN routing

3.2Tbps capacity line-rate performance at 170-byte frames

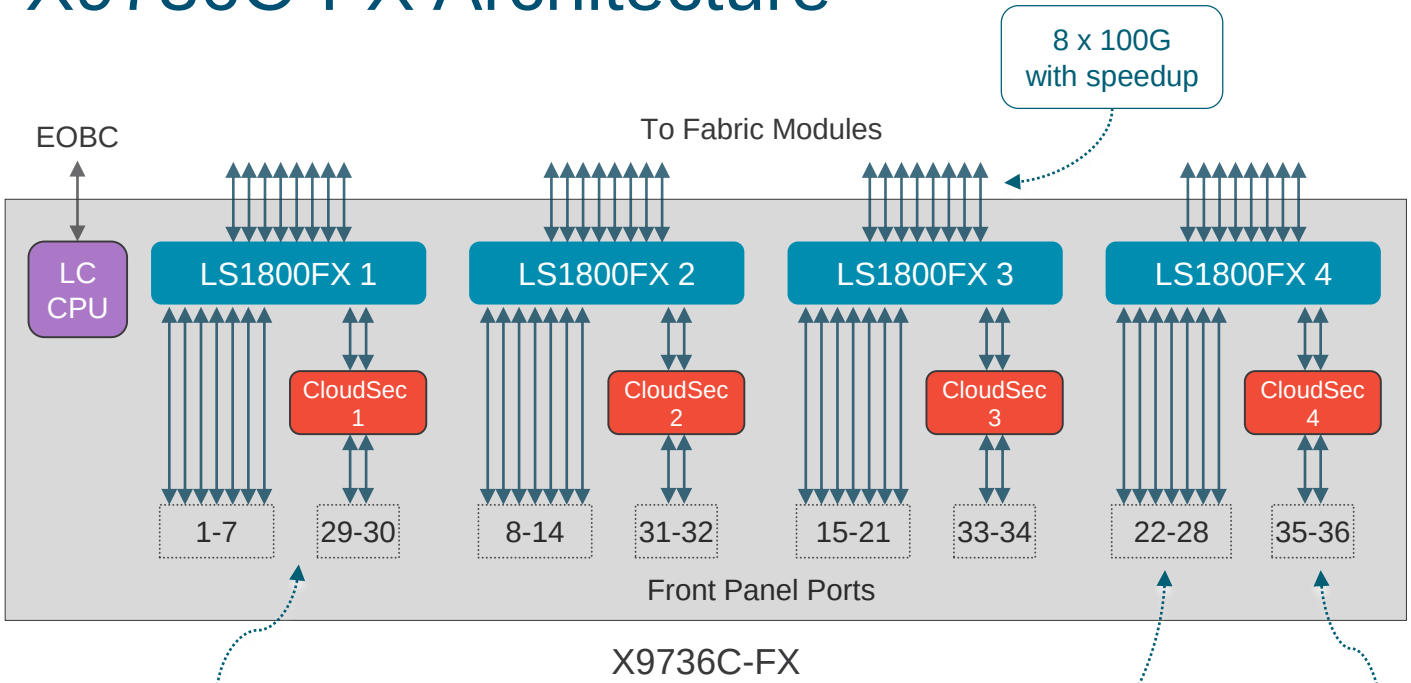


36p 100G QSFP28
X9736-FX – LS1800FX-based
ACI: Roadmap
NX-OS: Roadmap

36 x QSFP28-based 100G ports
- Pin-compatible with 40G QSFP+
- Flexible speed ports – 1 / 10 / 25 / 40 / 50 / 100G capability

Supported in ACI and NX-OS standalone mode

N9K-X9736C-FX Architecture



9 x 100G front-panel ports per LS1800FX

7 x 100G MACSEC capable ports per LS1800FX

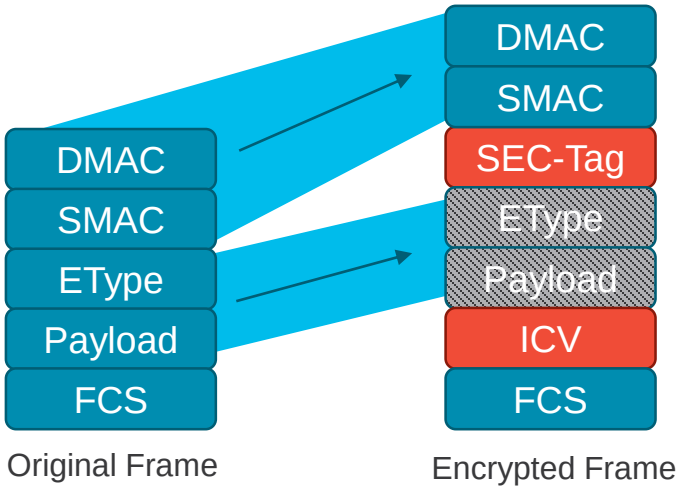
2 x 100G MACSEC / CloudSec capable ports per LS1800FX

Slice 0

MACSEC Hardware Encryption

- Provides link-level hop-by-hop encryption
- IEEE 802.1AE 128-bit and 256-bit AES encryption with MKA Key Exchange
- Native hardware support available on:
  - All ports on X9736C-FX linecard
  - All ports on Nexus 93180YC-FX / 93108TC-FX switches
  - 16 x 100G ports on Nexus 9364C switch
  - All ports on Nexus 9336C-FX2 / N9K-C93240YC-FX2 switches

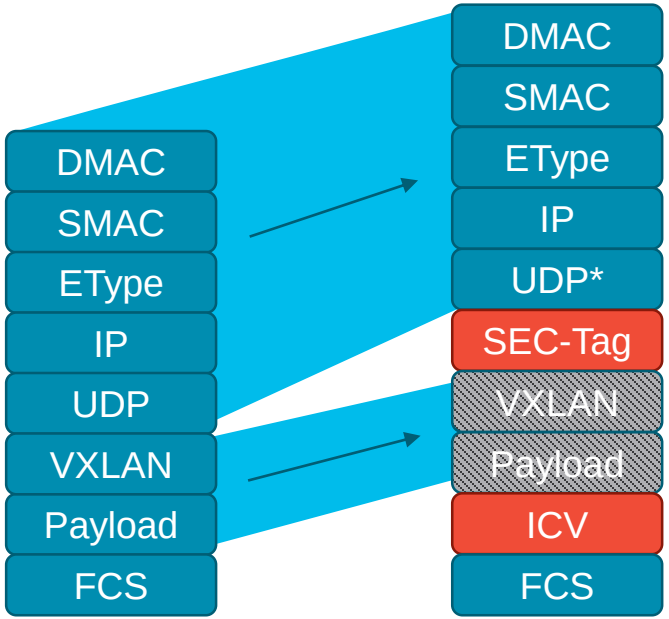
MACSEC Frame Format



CloudSec Hardware Encryption

- Provides VTEP-to-VTEP encryption
- Encrypts VXLAN header and payload for transport over arbitrary IP network
- Hardware support available on:
  - 8 x 100G ports on X9736C-FX linecard
  - 16 x 100G ports on Nexus 9364C
  - All ports on 9300-FX2 TORs
- No support on other TOR switches

CloudSec Frame Format

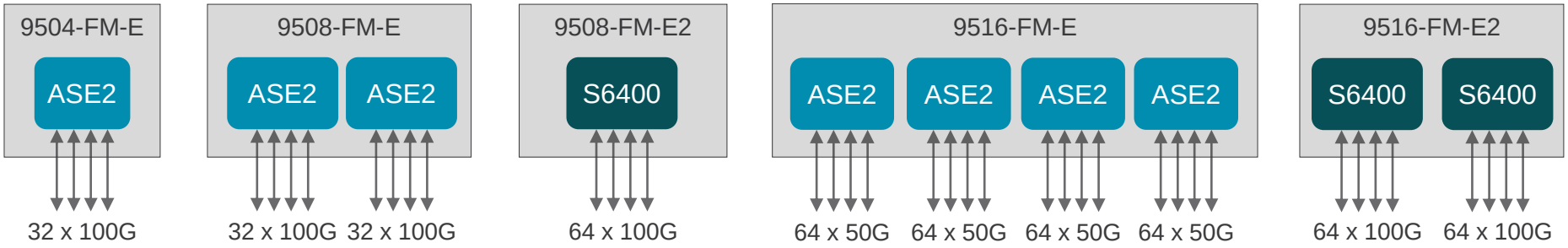


Original VXLAN Packet

Encrypted Packet

* CloudSec UDP dest port

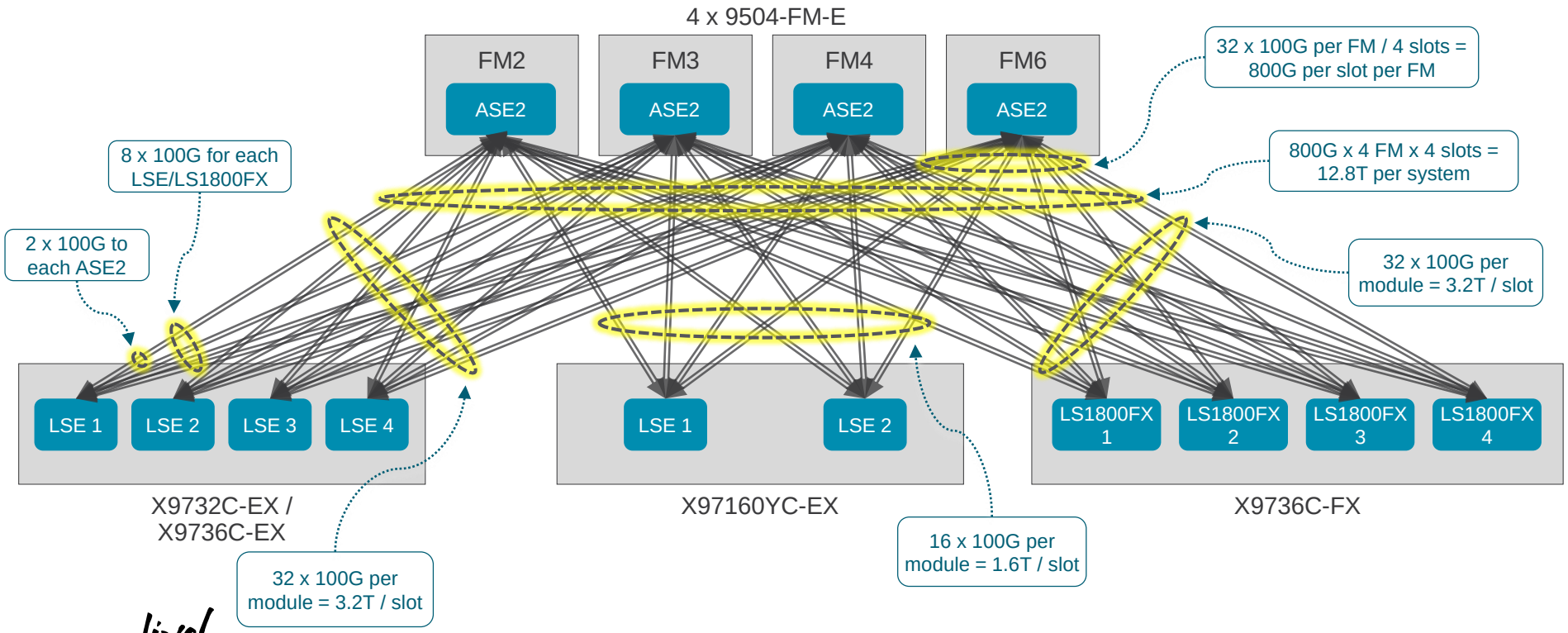
Cloud Scale Fabric Modules – FM-E and FM-E2



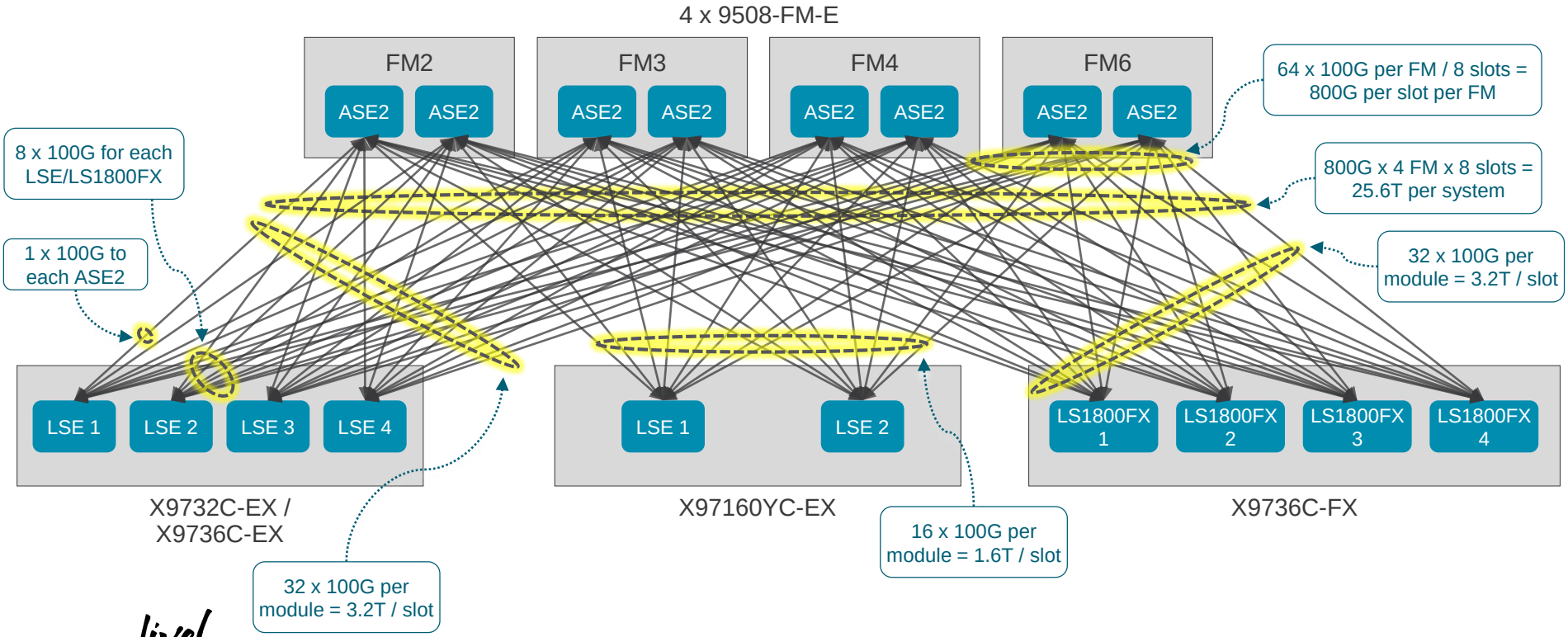
- Cloud Scale linecards require Cloud Scale fabric modules
- Provide up to 3.2Tbps capacity per IO module slot with 4 FMs
- **Note:** Cloud Scale FMs support X9700-EX and X9700-FX modules **only**

- N9K-C9504-FM-E
  ACI: 1.3(1)
  NX-OS: 7.0(3)I4(2)
- N9K-C9508-FM-E
  ACI: 1.3(1)
  NX-OS: 7.0(3)I4(2)
- N9K-C9508-FM-E2
  ACI/NX-OS: Roadmap
- N9K-C9516-FM-E
  ACI: Roadmap
  NX-OS: 7.0(3)I5(2)
- N9K-C9516-FM-E2
  ACI/NX-OS: Roadmap

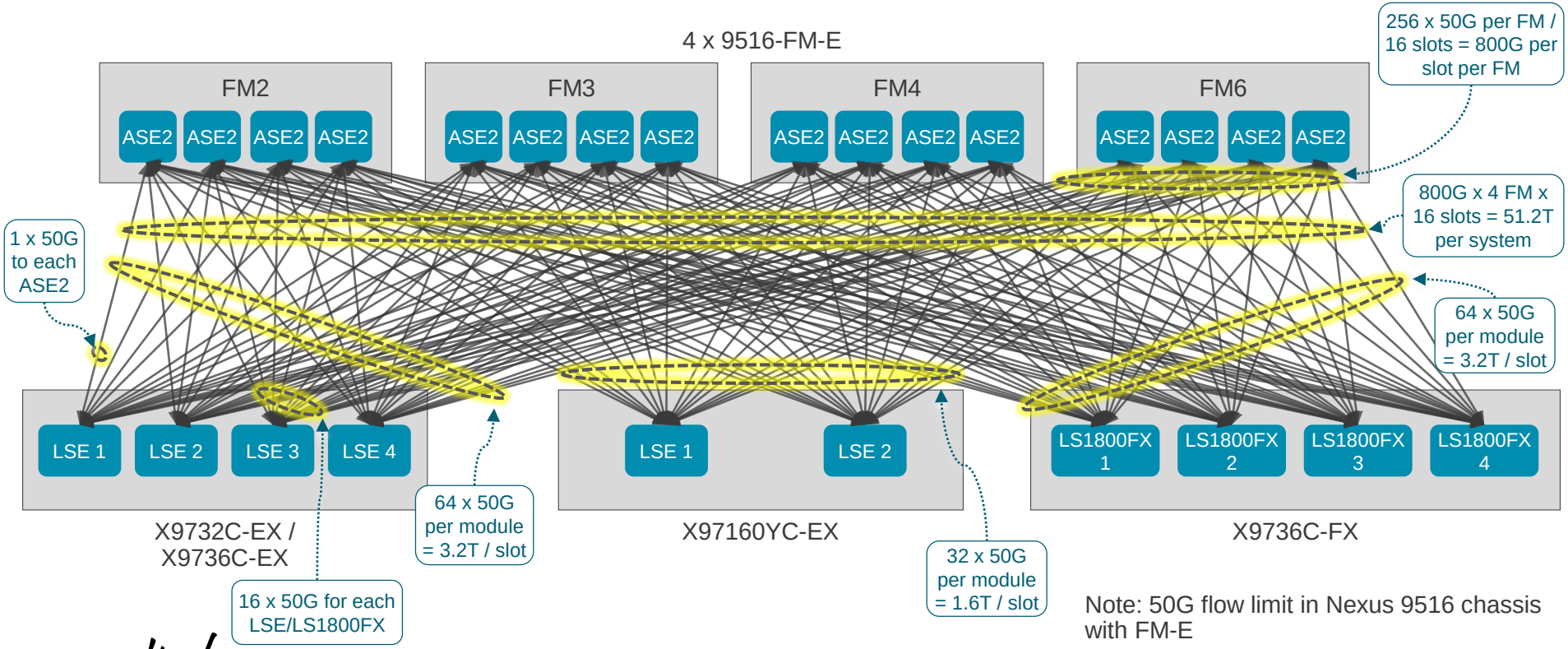
Cloud Scale Fabric Connectivity – Nexus 9504



Cloud Scale Fabric Connectivity – Nexus 9508

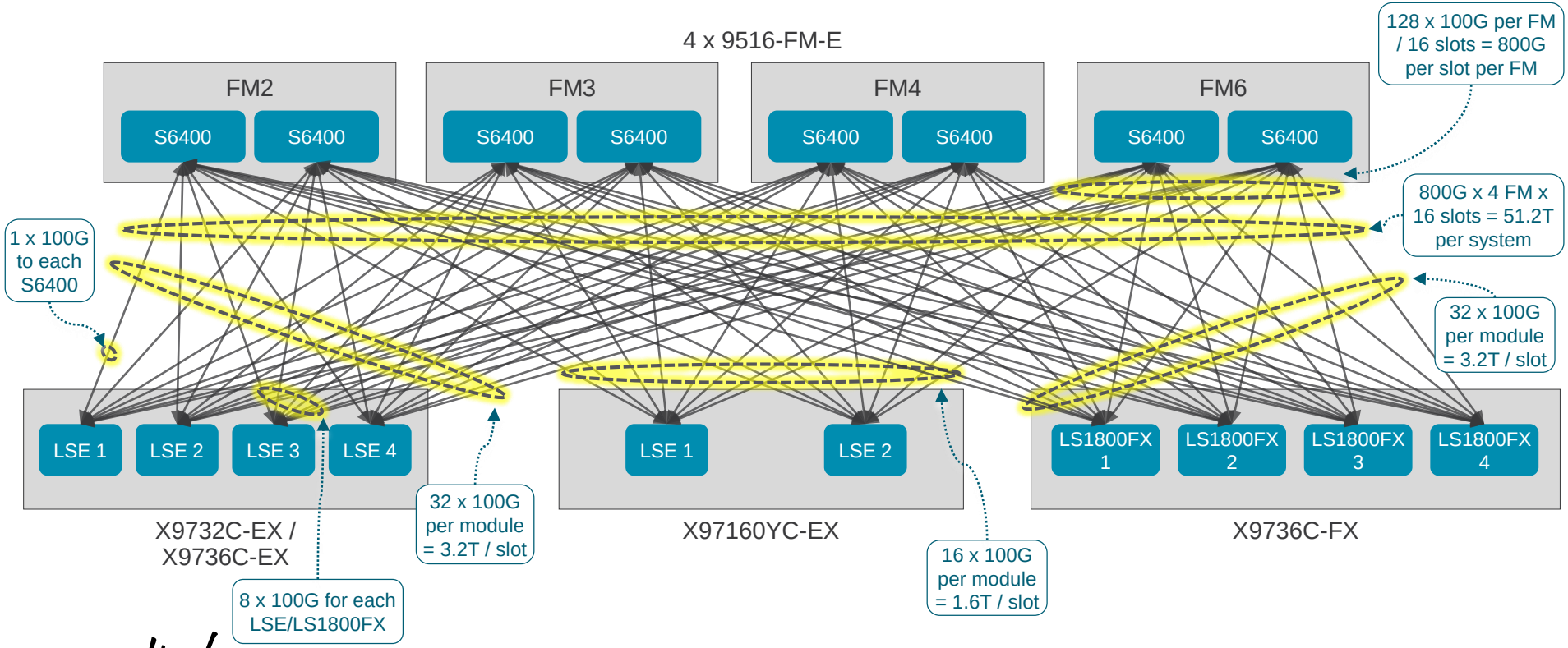


Cloud Scale Fabric Connectivity – Nexus 9516 FM-E



Note: 50G flow limit in Nexus 9516 chassis with FM-E

Cloud Scale Fabric Connectivity – Nexus 9516 FM-E2



Agenda

- Data Centre and Silicon Strategy
- Cloud Scale Architecture
  - Cloud Scale ASICs
  - Forwarding and Features
- Cloud Scale Switching Platforms
- Optics and What’s Next

25/50G Ethernet Standards

Successful plug fest UNH IOL 25/50G
Aug 1-4 2016

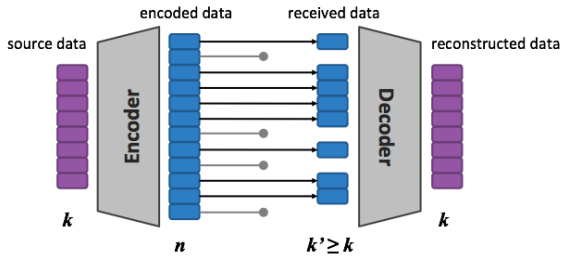
	Consortium	IEEE	Cisco TMG Cables*
Distance	Passive: 1,2,3 meter	Passive: 1,2,3,5 meter Optics: SR	AOC cables: 1,2,3,5,7,10M (Shipping Jan CY17')
Deployment	Within Rack	Across Rack	Within/Across Rack
Supporting Platform	N9200, N9300-EX N3200, X9700-EX	Roadmap N9300-FX X9700-FX, X97160YC-EX	N9200, N9300-EX, N3200, X9700-EX, N9300-FX
Forward Error Correction	3m needs FC FEC	3m needs FC FEC >3m need RS FEC	Can work with either FC FEC or RS FEC
NIC (Verified)	Mellanox		NIC needs to support the same FEC mode as the switch
NIC (Ongoing Testing)	Qlogic, BRCM, Intel		

*IEEE or Consortium does not spec for AOC 1,2,3,5, 7,10 meter.

What about 25G?

FEC (Forward Error Correction)

- FEC greatly reduce uncorrected errors across the media and help to extend the usable reach of those media
- FEC introduces latency penalty and depending on the distance FEC could be disabled to optimise the latency (~250 nsec)
- 25G standard support 3 modes of FEC to support different twinax cable reach
  - Clause 74 Fire code FEC: FC FEC
  - Clause 108 Reed-Solomon FEC: RS FEC
- Passive cable 1 and 2 meter does not require FEC
- Passive cable 3 meter requires FC FEC
- Passive cable more than 3 meter or 100m MMF SR optics requires RS FEC
- RS FEC introduce more latency than FC FEC



Raw BER*	BER after FEC*
5.7E-7	1.97E-29

* Example of FEC improvement of realised BER with 56G PAM4 encoding

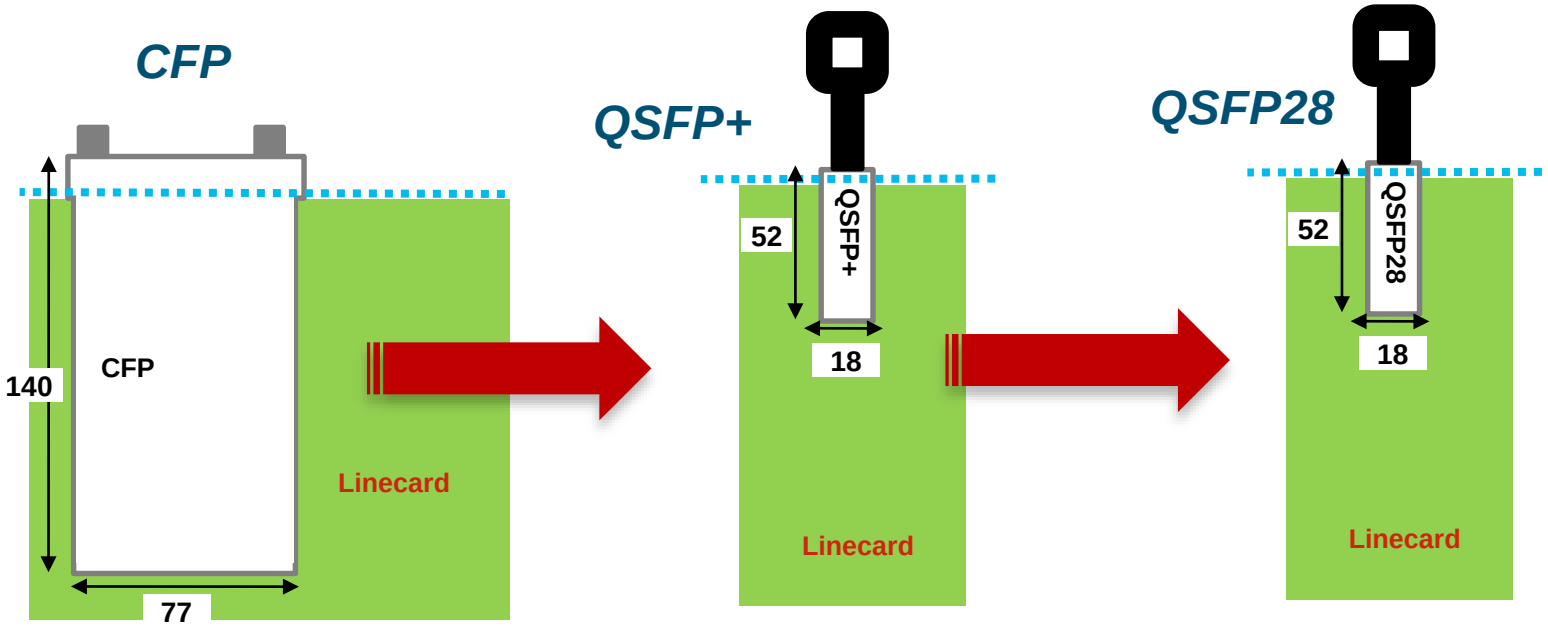
25G / 10G Backward Compatibility

- 25G Ethernet passive cable support both 10G and 25G speed
- 10G and 40G Ethernet passive cable are not designed to run at 25G Ethernet single lane

Optics		Platform
Passive Cables	1/2/3/5 meter	Nexus 92160YC-X
Active Cables	1/2/3 meter *	Nexus 92160YC-X
Breakout Cables	1/2/3 meter	Nexus 9232C Nexus 9236C Nexus 92160YCX

* Active cable greater than 3 meter requires FEC RS which is not supported on Nexus 92160YCX

Next Generation Packages for 40/100G
QSFP+ & QSFP28



QSFP28

1/2 the power & 1/5 the size of CPAK

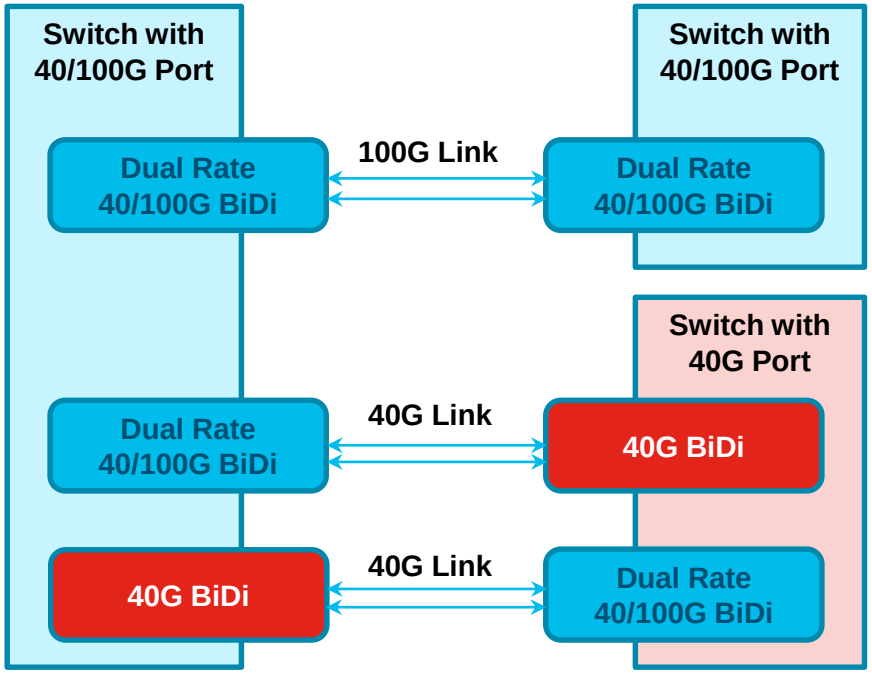
44% the size of CFP4

	QSFP+	QSFP28
Power (W)	3.5	~3.5
Electrical	4x10G	4x25G

Cost of Optics and Fibre

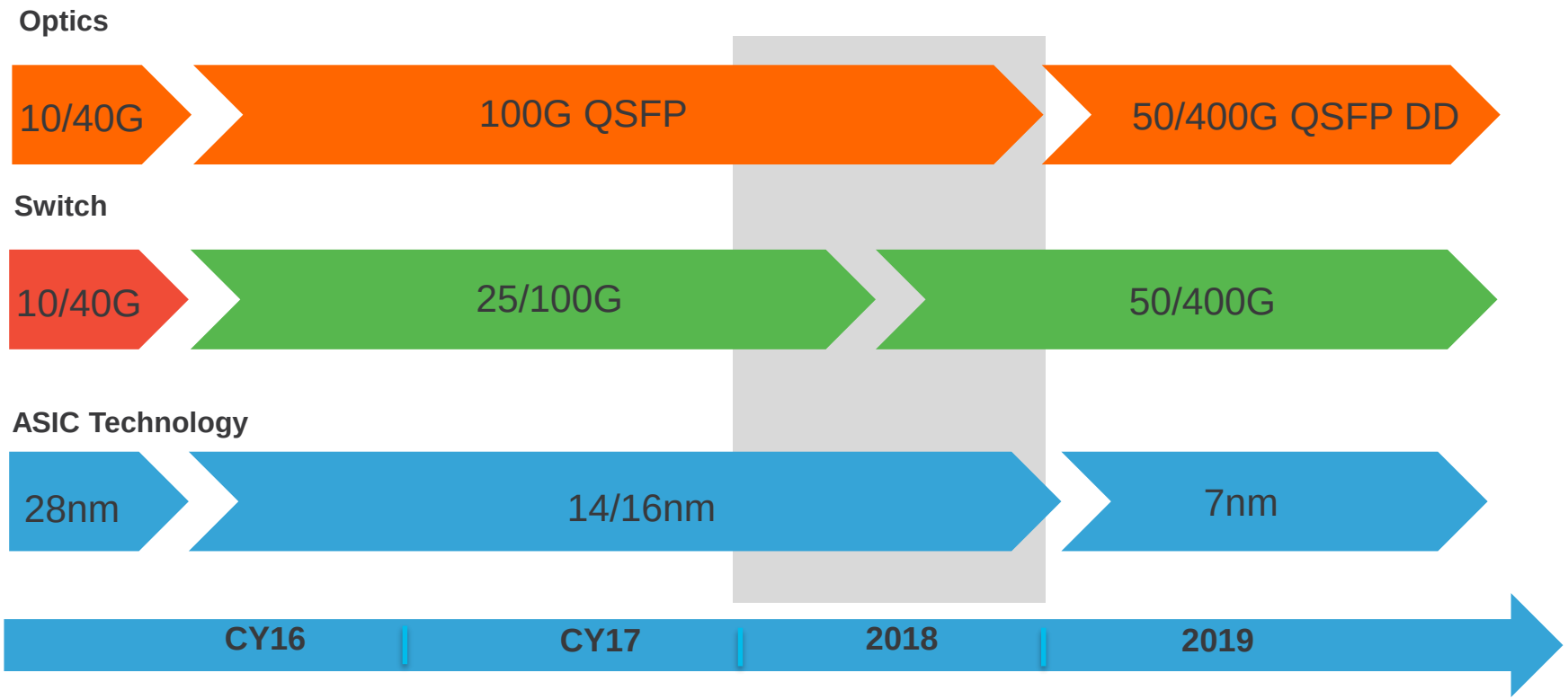
Cisco 40/100G Bidi

- Standard QSFP
- Leverage existing MMF infrastructure to support 100G
- 40G or 100G Dual Rate Optics
- Price Parity with 100G SR



What's Next?

50/400G



What's Next?

50/400G

Pulse **A**mplitude **M**odulation
4 Indicates the number of valid signal levels

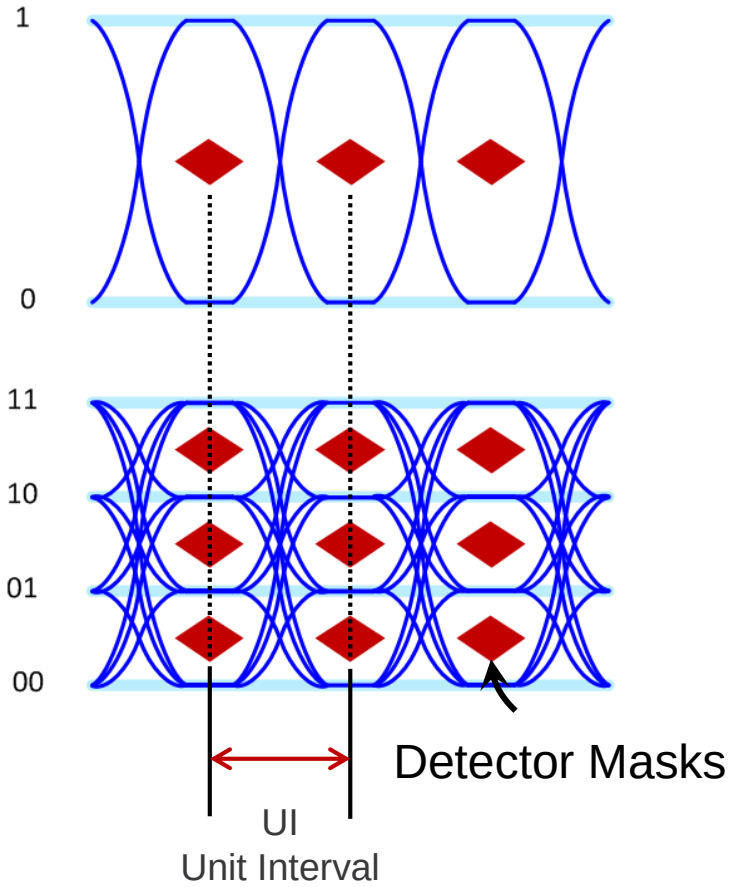
- NRZ is the same as PAM2
- PAM3 is used in 100Base-T
- PAM5 is used in 1000Base-T
- PAM16 is used in 10GBase-T

Higher order modulation with PAM has been used for decades to achieve higher bit rates

Ideal Differential Eye Diagrams

NRZ:
2 Levels
1 Bit per UI

PAM4:
4 Levels
2 Bits per UI



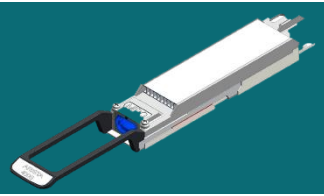
400G Transceivers – Competing Form Factors

QSFP-DD

- ✓ Up to 36 ports in 1RU
- ✓ Same front-panel dimension as QSFP28
- ✓ Compatible with QSFP & QSFP28

VS

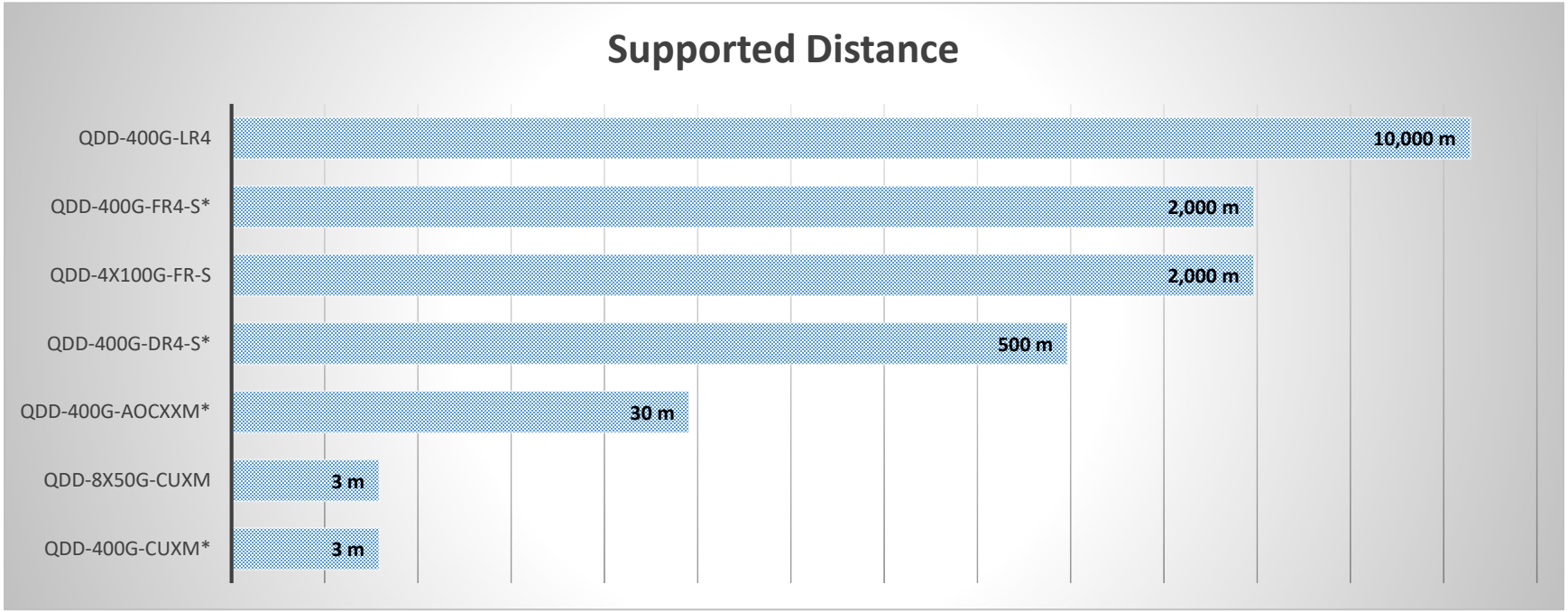
OSFP



- – Up to 32 ports in 1RU
- – Larger than QSFP28
- – Incompatible with QSFP & QSFP28

- ✓ Supports 25G, 50G & 100G SERDES
- ✓ Supports all media (Fibre & Copper)
- ✓ Supports all reaches (3m – 100km)
- ✓ Meets Thermal & Signal Integrity Requirements

Cisco QSFP-DD 400G Optical Module Product Portfolio



* Committed Programs. Other modules and cables are in planning stage

Agenda

- Data Centre and Silicon Strategy
- Cloud Scale Architecture
  - Cloud Scale ASICs
  - Forwarding and Features
- Cloud Scale Switching Platforms
- Packet Walks
- Key Takeaways

Nexus 9000 – Market Momentum

14,500+
Nexus 9K Customers Globally

4500+
ACI Customers

65+
Ecosystem Partners

ECOSYSTEM PARTNERS

Key Takeaways

- You should now have a thorough understanding of the Nexus 9000 Cloud Scale switching platform architecture
- Feature-rich, innovative switching platform addresses virtually every deployment scenario
- Nexus 9000 Cloud Scale platform forms foundation of the ASAP Data Centre

Q & A

Complete Your Online Session Evaluation

- Give us your feedback and receive a **Cisco Live 2018 Cap** by completing the overall event evaluation and 5 session evaluations.
- All evaluations can be completed via the Cisco Live Mobile App.

Don’t forget: Cisco Live sessions will be available for viewing on demand after the event at www.CiscoLive.com/Global.

Thank you

cisco

You're

Cisco live!