

# CRS Architecture

Yue Feng (CSE)



# Agenda

- CRS1/3 overview
- Multichassis
- CRS-X overview
- IOS-XR software



# What Makes the CRS Different?

- What does Carrier Class architecture mean?

Reliability

Scalability

Predictability

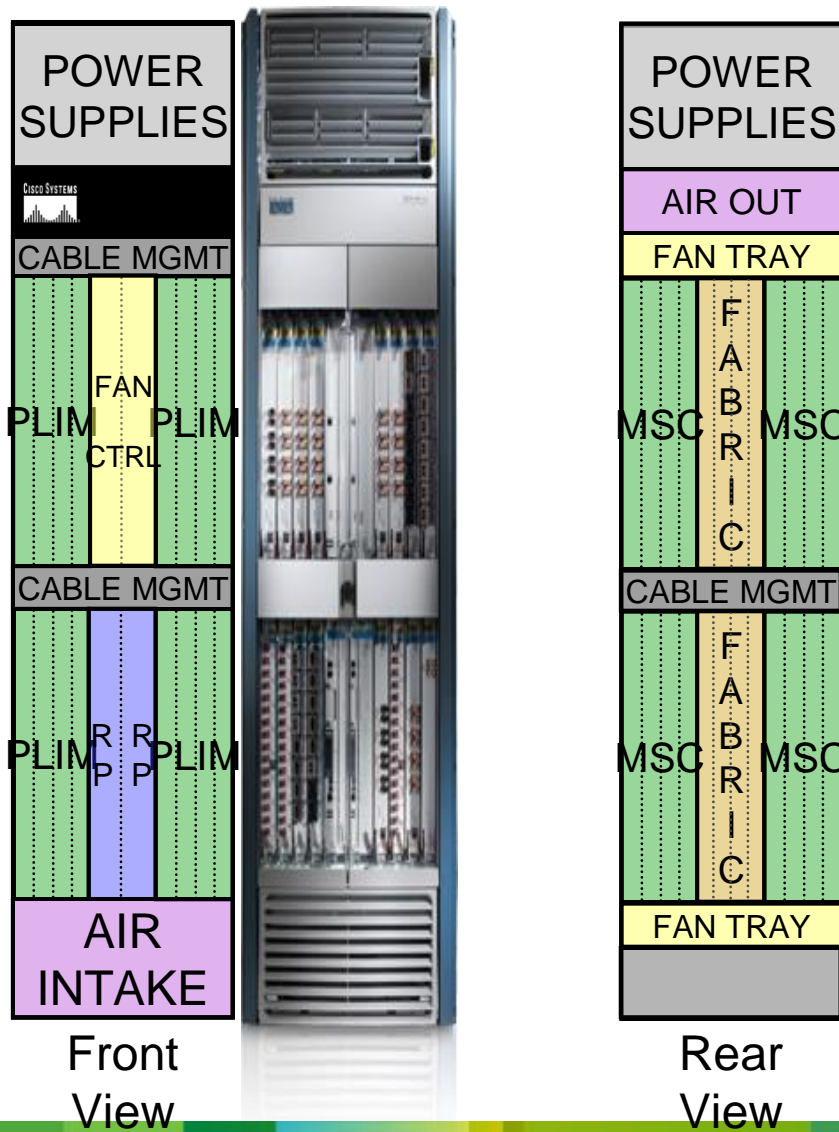
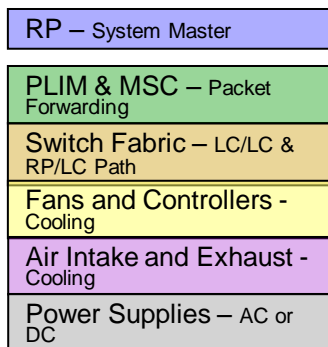
- Where do you need it?
- How does a router deliver?



# CRS Portfolio: Various form-factors

## 16-slot Chassis

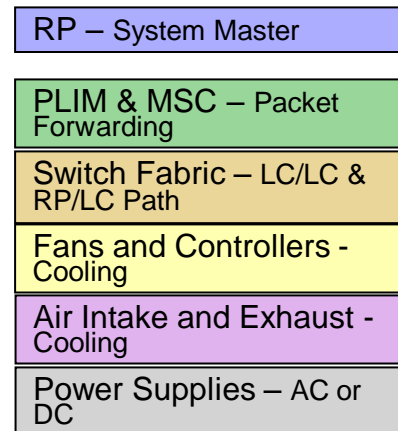
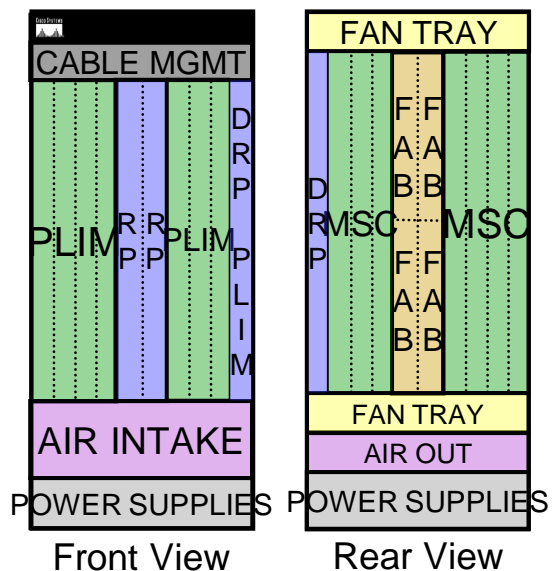
- PLIM (interfaces)
- Line Cards (“controller”, ex: MSC)
- Route Processors (2)
- Fabric Cards (8)
- Fan Controllers / Fan Trays (2)
- Power Shelves / Power Modules
- Rack by itself
- 60 W x 104.2 D x 213.36 H (cm)
- 725kg



# CRS Portfolio: Various form-factors

## 8-slot Chassis

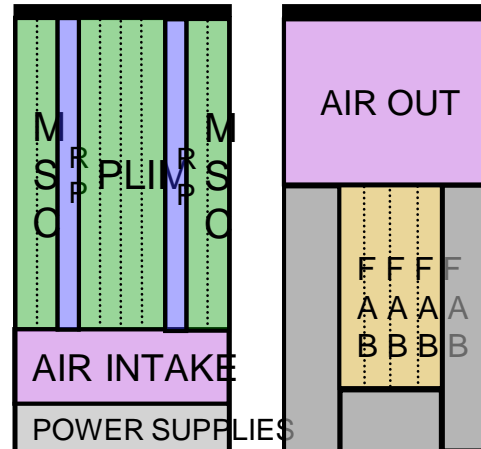
- PLIM (interfaces)
- Line Cards (“controller”, ex: MSC)
- Route Processors (2)
- Fabric Cards (4)
- Fan Trays (2)
- Air Intake and Exhaust - Cooling
- Power Shelves / Power Modules
- Dimensions:
  - 44.5 W x 93 D x 97.8 H (cm)
- Weight: ~ 275kg
- Rack mountable



# CRS Portfolio: Various form-factors

## 4-slot Chassis

- **EoX cycle started, Bundles exist to replace with 8-slot chassis**
- Front:
  - 2 RP slots (same as CRS8)
  - 4 PLIM slots
  - 4 LC/MSC Slots
- Back - 4 Fabric cards
- Dimensions: 44.8 W x 76.9 D x 76.2 H (cm)
- Rack mountable

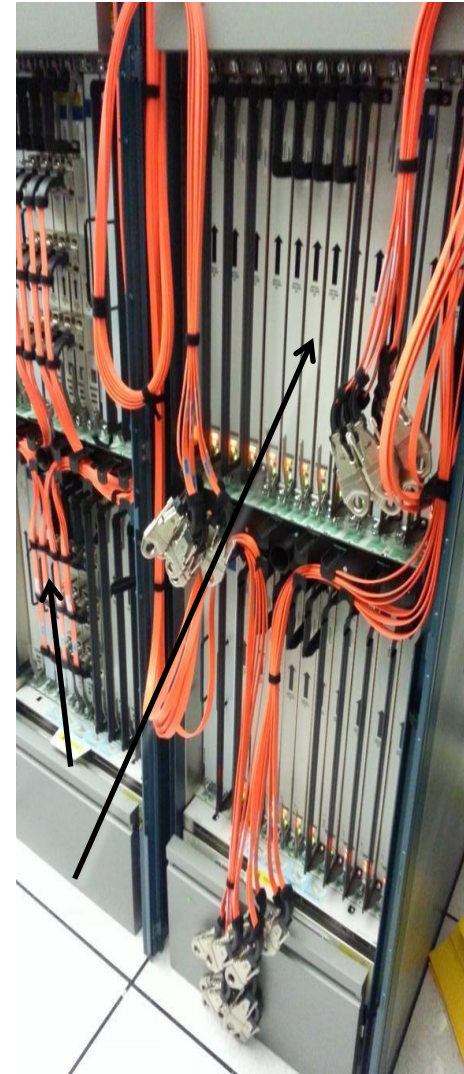




# CRS Portfolio: Various form-factors

## Fabric Cards 40G / 140G / 400G

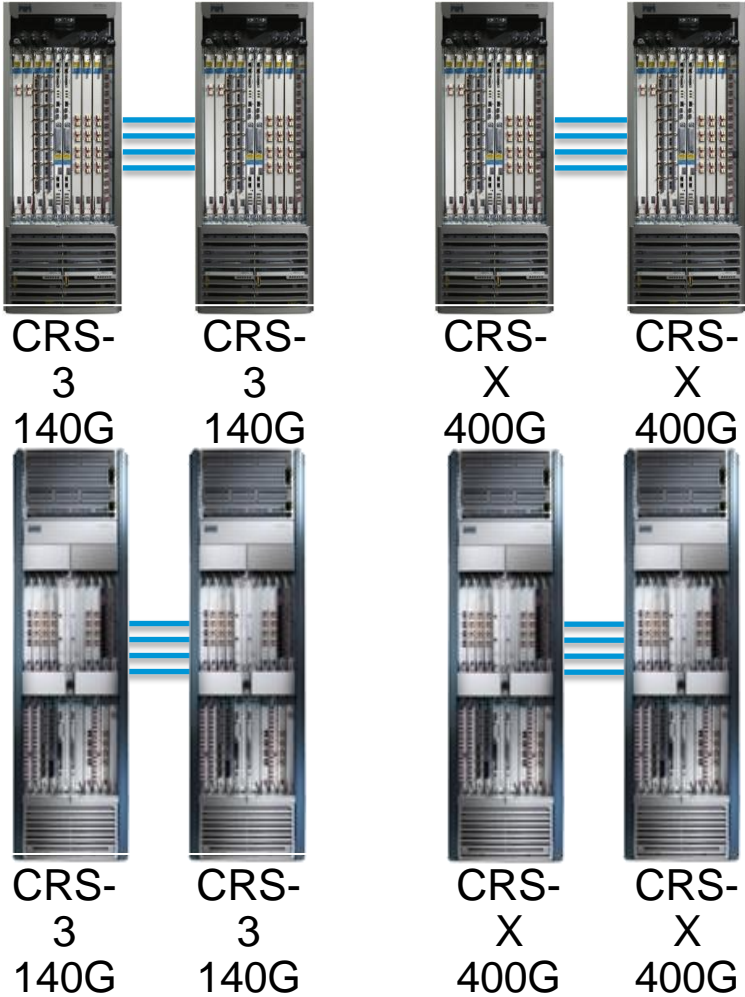
- Fabric is what makes a CRS-1, CRS-3 and CRS-X
  - CRS-1 is using Fabric Cards at 40G
  - CRS-3 is using Fabric Cards at 140G
  - CRS-X is using Fabric Cards at 400G
- /S for Single Chassis
- /M for MultiChassis and B2B
- Same Fabric Cards but different “Inter-Chassis Bundle” fibers between MC and B2B
- No B2B for CRS-1



# CRS Portfolio: Various form-factors

B2B and MC Supported Configurations

- We support:

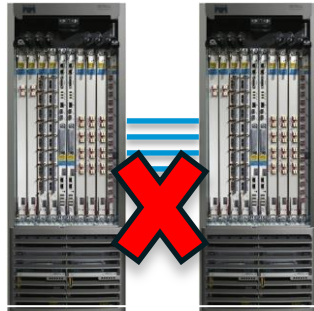




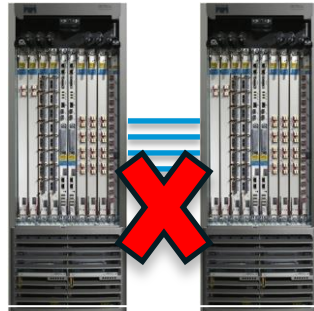
# CRS Portfolio: Various form-factors

B2B and MC Not Supported Configurations

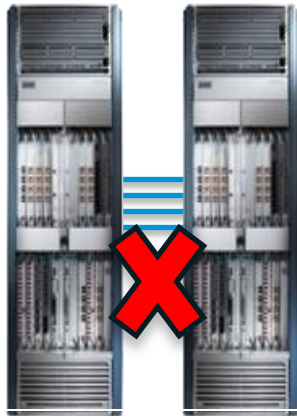
- We do NOT support:



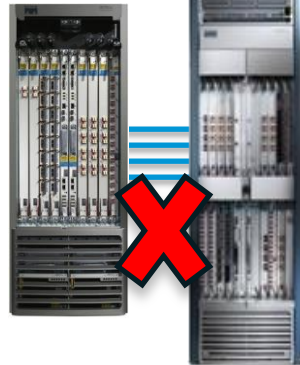
CRS-3 140G      CRS-3 400G



CRS-1 40G      CRS-1 40G



CRS-3 140G      CRS-X 400G



8-slot      16-slot

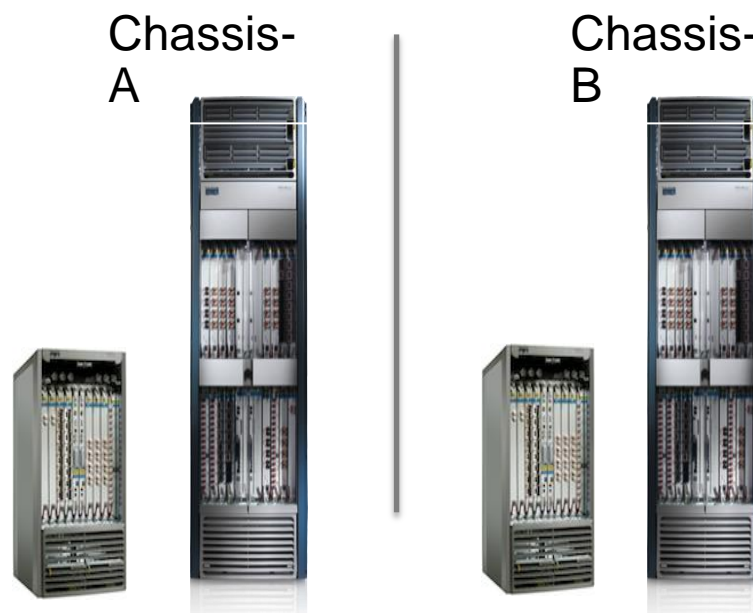


LCC 400G    LCC 140G    FCC 140G    FCC 40G    LCC 400G    LCC 400G



# Legacy and Enhanced Chassis

- Starting from 2011, we introduced new and “enhanced” version of 8-slot and 16-slot chassis
- -A is often referred as Legacy, -B being Enhanced
- Not for 4-slot Chassis



- Main difference being the power and cooling capabilities
- With 400G fabric cards, the –A chassis can only support 200G per line card slot.

# In Summary

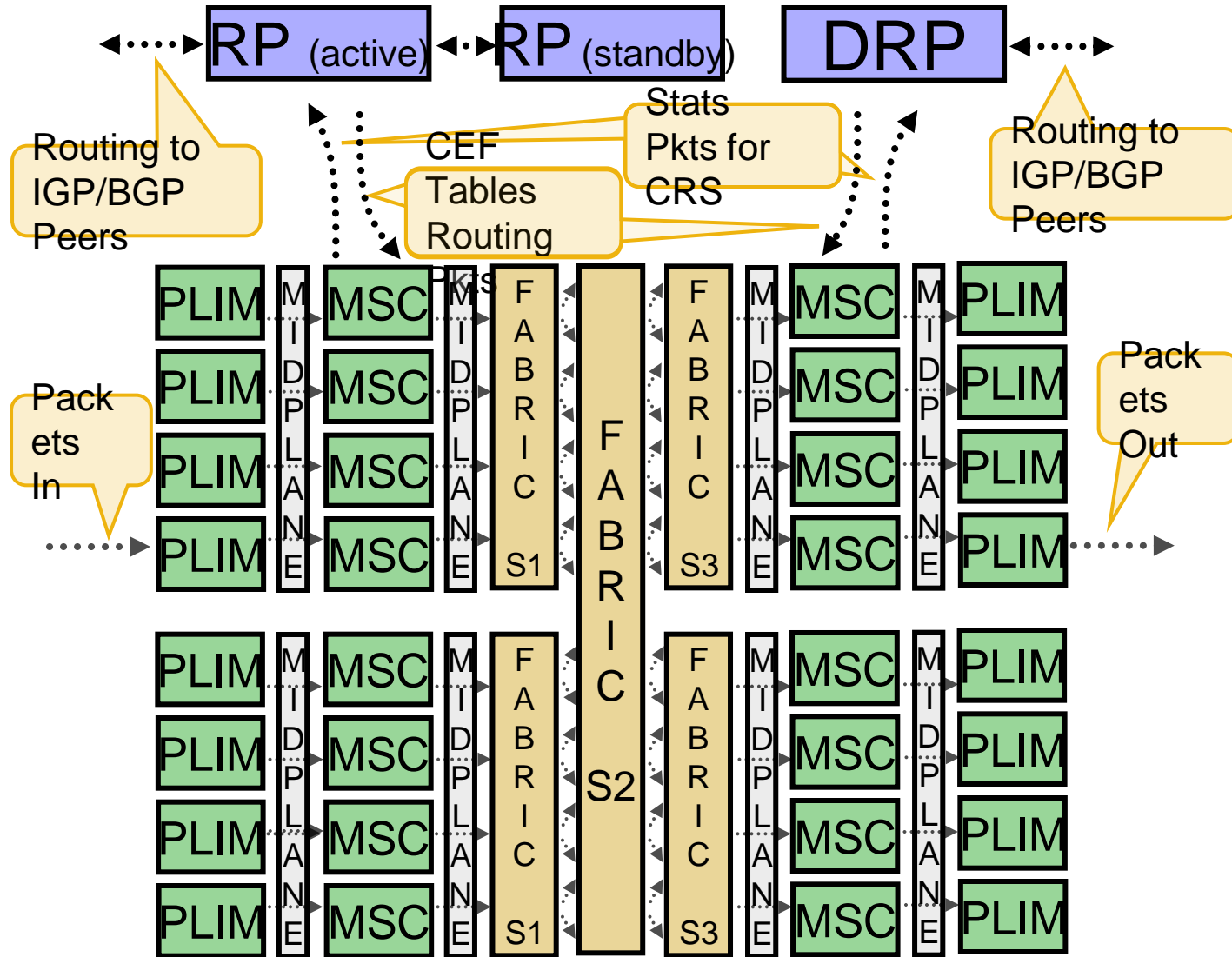
- CRS-1 / CRS-3 / CRS-X depends on the Fabric Cards generation
  - Offering 40G, 140G and 400G per slot.
- Many form factors:
  - Single Chassis: 4-slot, 8-slot and 16-slot
  - Multi Chassis: Back-to-Back 8-slot and 16-slot, MC from 2+1 to 8+2
- Legacy and Enhanced 8-slot and 16-slot chassis



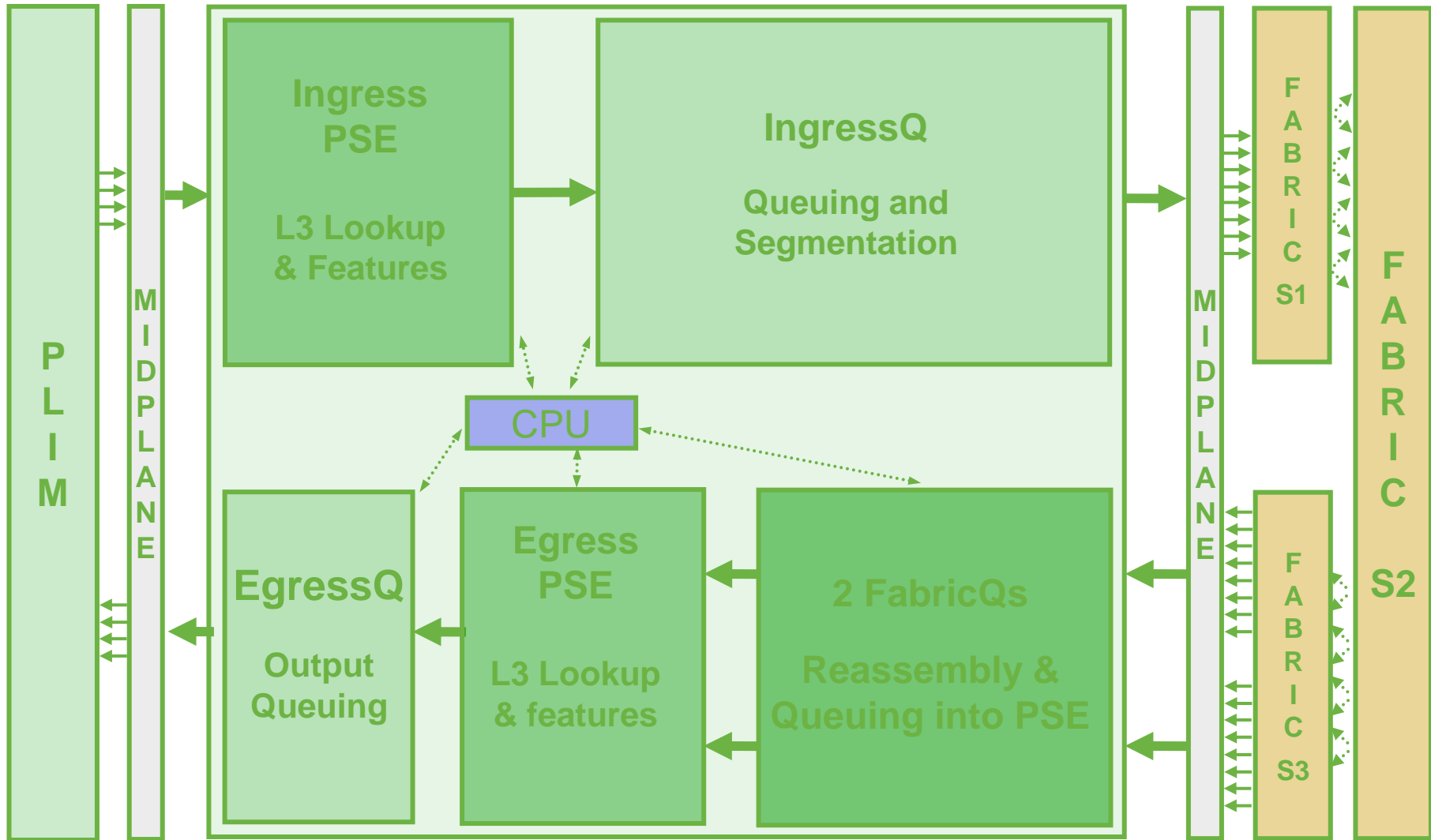
# CRS-1/3 Overview



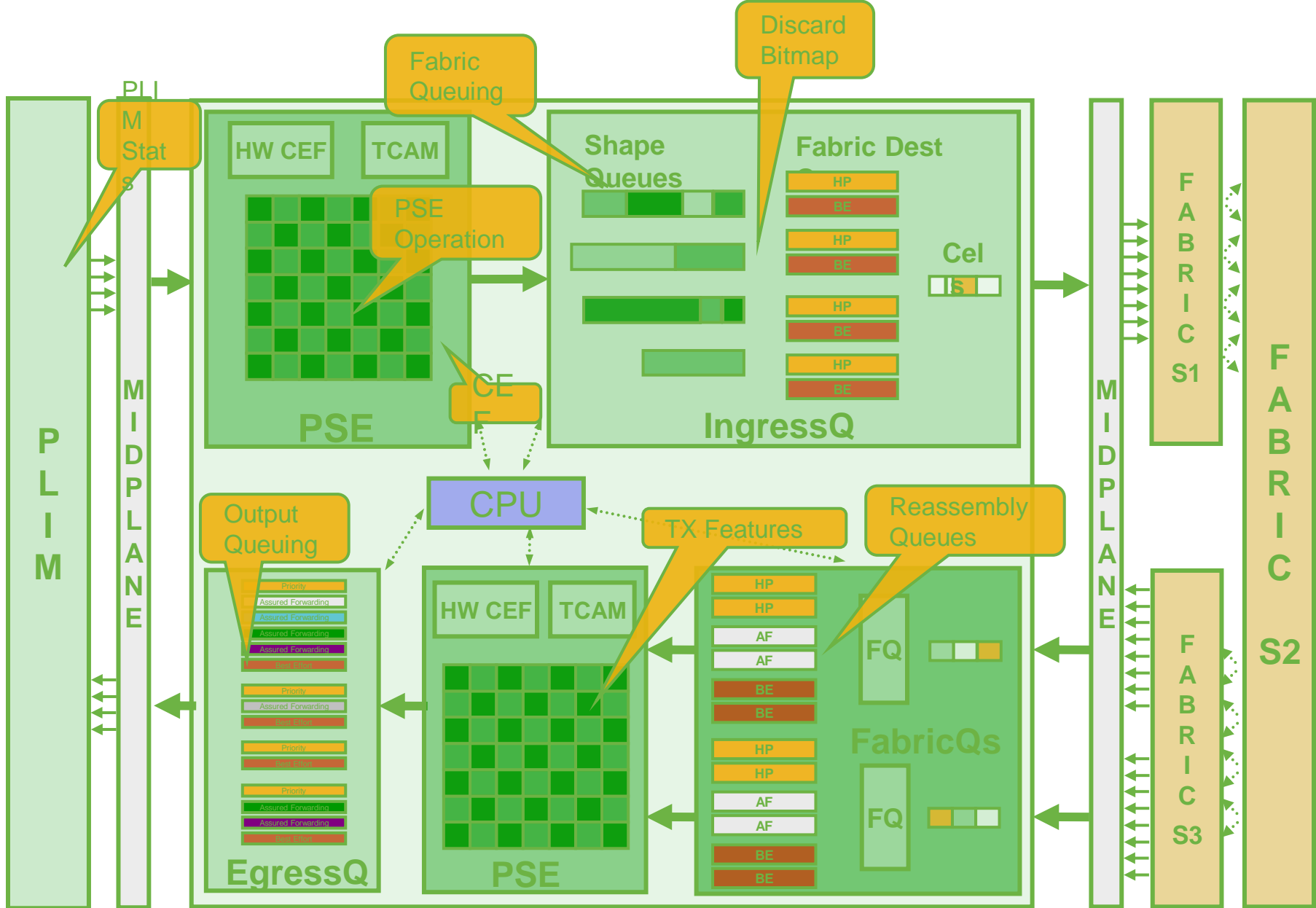
# High-Level CRS Hardware Architecture



# CRS Architecture Review – Line Card



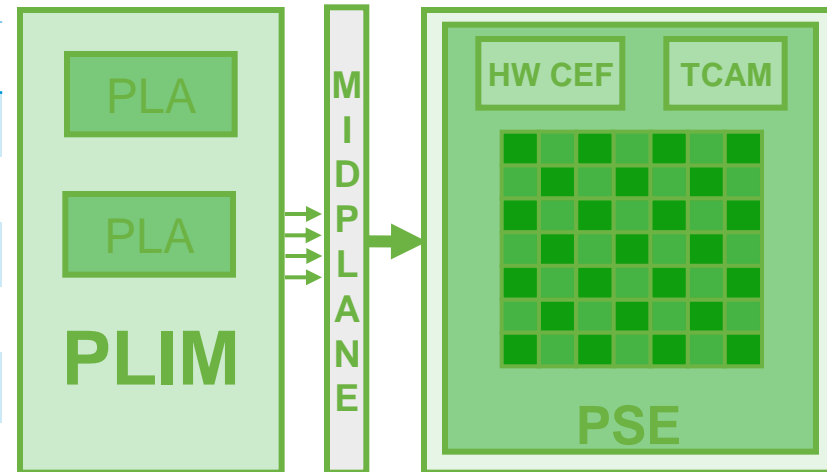




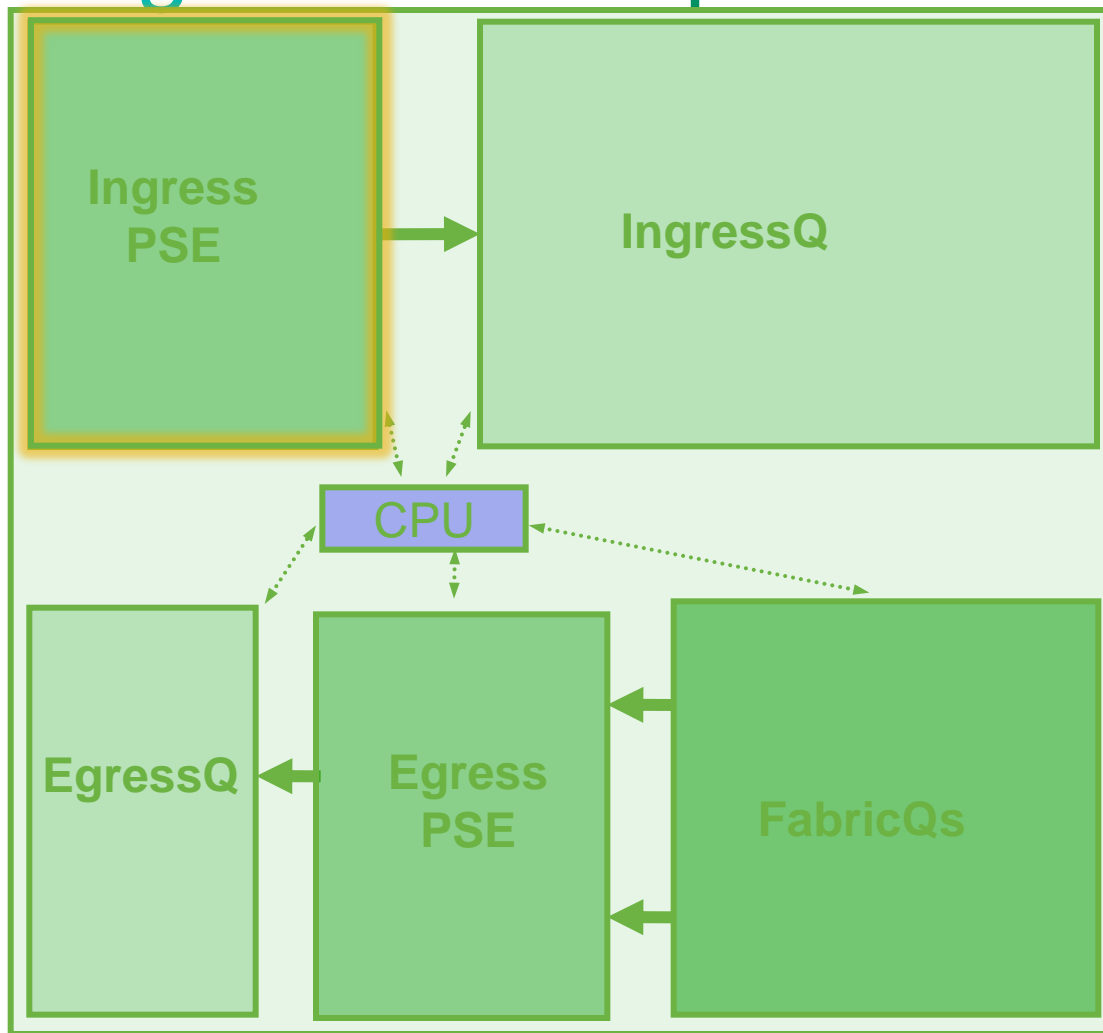
# Ingress PLIM

- 1 – 4 PLIM ASICs (PLAs) depending on card type
- Nominally 40 Gbps for CRS-1, 140G for CRS-3
- Oversubscription allowed when all Ethernet
- 96 Gbps aggregate bandwidth into PSE
  - No bandwidth bottleneck even when oversubscribed
- CRS-3 adds PLIM QoS

PLIM	PLAs	BW to PSE (per PLA)
4xOC192 POS	2	48 Gbps
16xOC48 POS	4	24 Gbps
1xOC768 POS	1	96 Gbps
4/8xTenGE	2	48 Gbps
SIP-800	2	48 Gbps

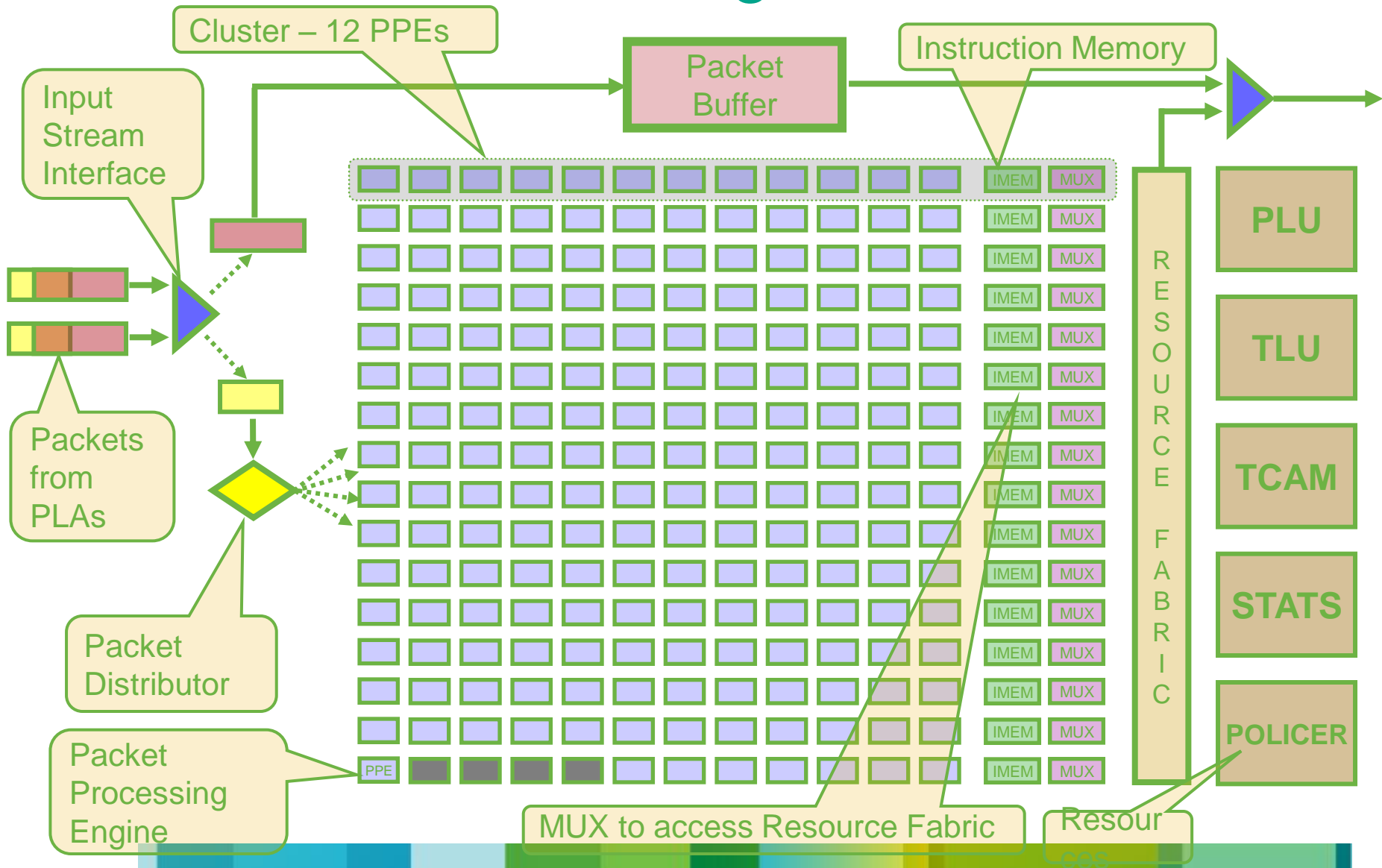


# Ingress PSE Topics



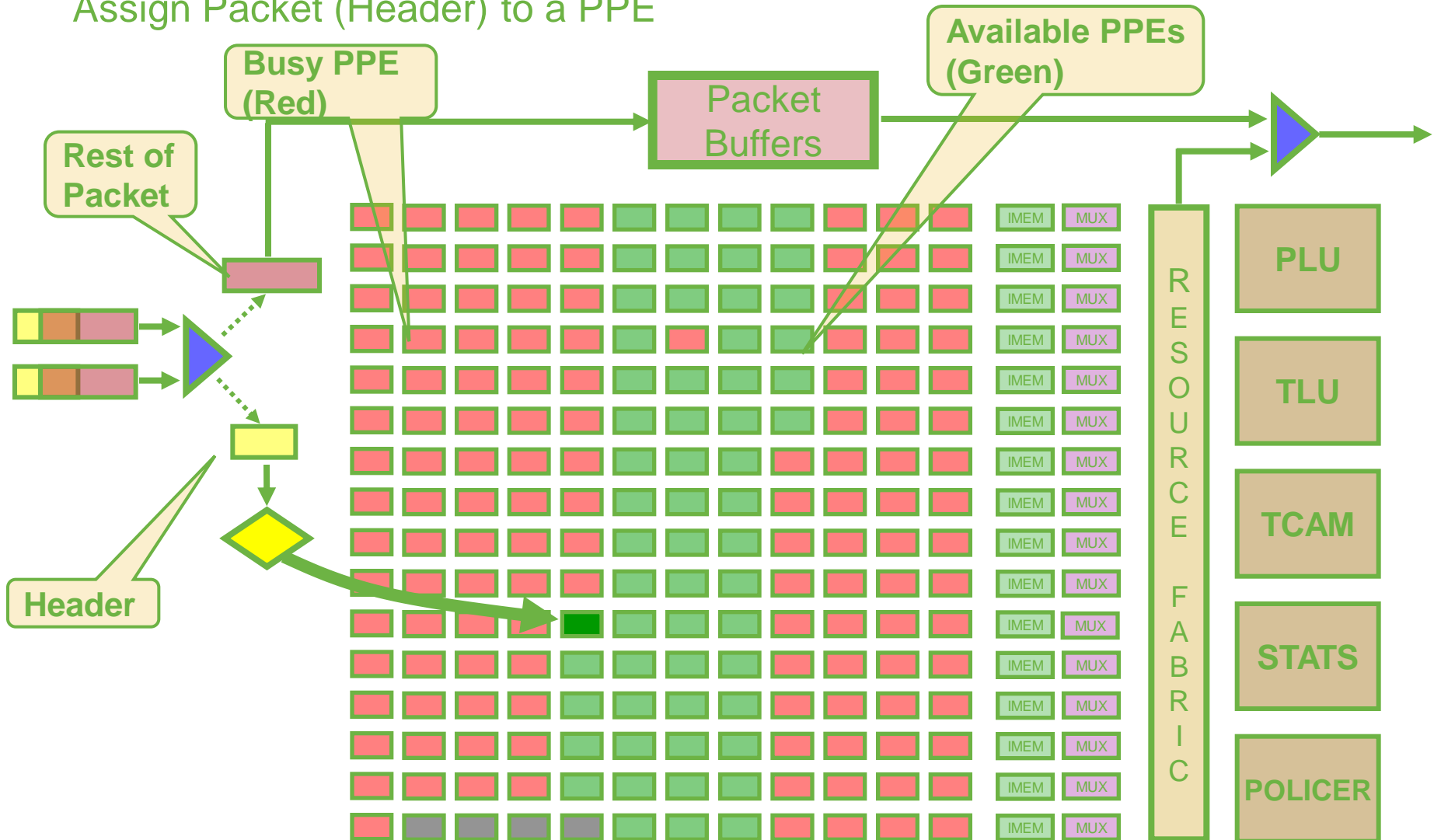
- ASIC Architecture
- PPE Operation
- TCAMs/ACLs
  
- Features
  - CEF - IP/MPLS
  - ACLs
  - Policing\*
  - Netflow\*
  - uRPF\*

# CRS-1 PSE Forwarding ASIC Architecture



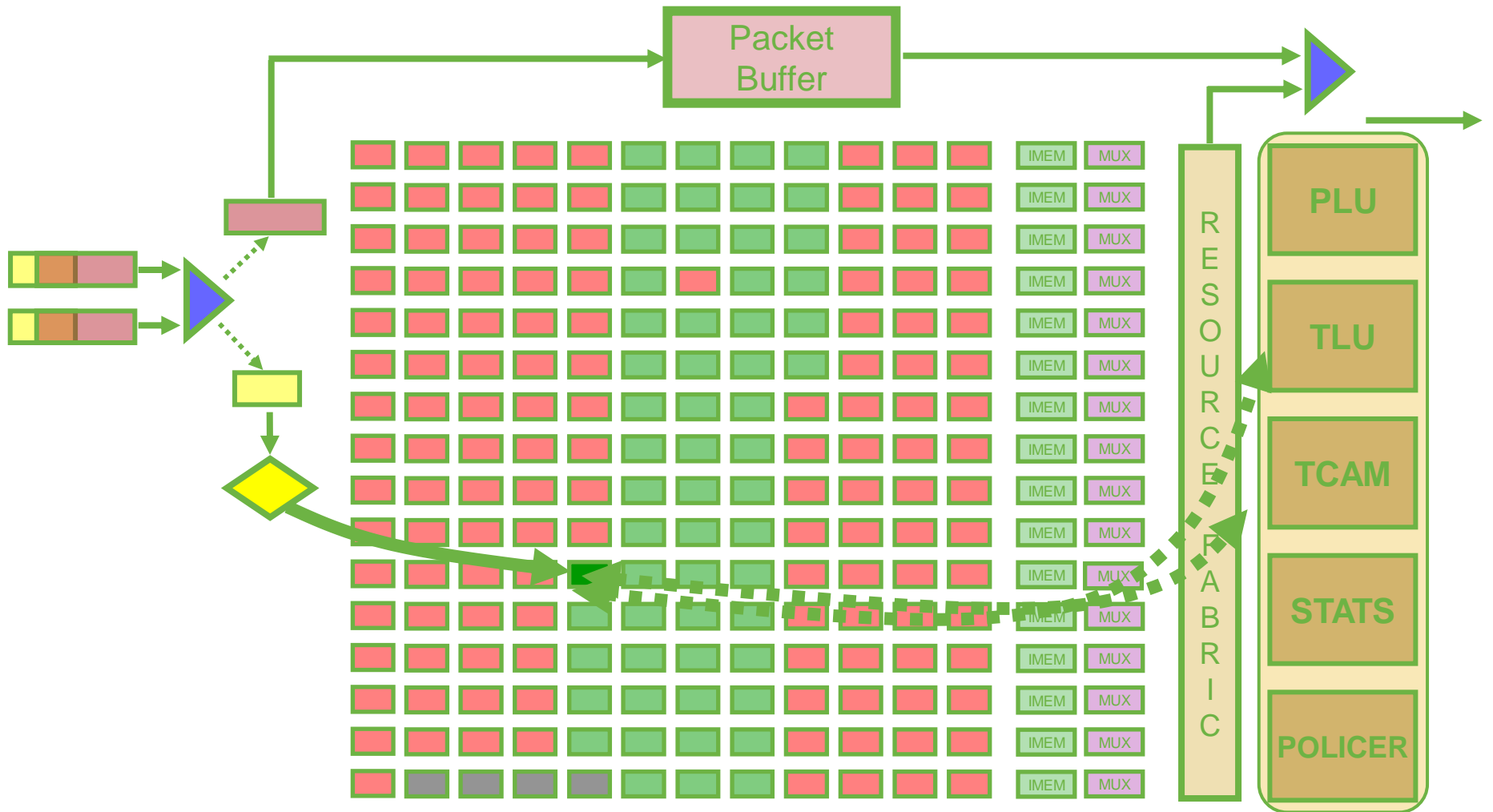
# Packet Path Within PSE

Assign Packet (Header) to a PPE



# Packet Path Within PSE

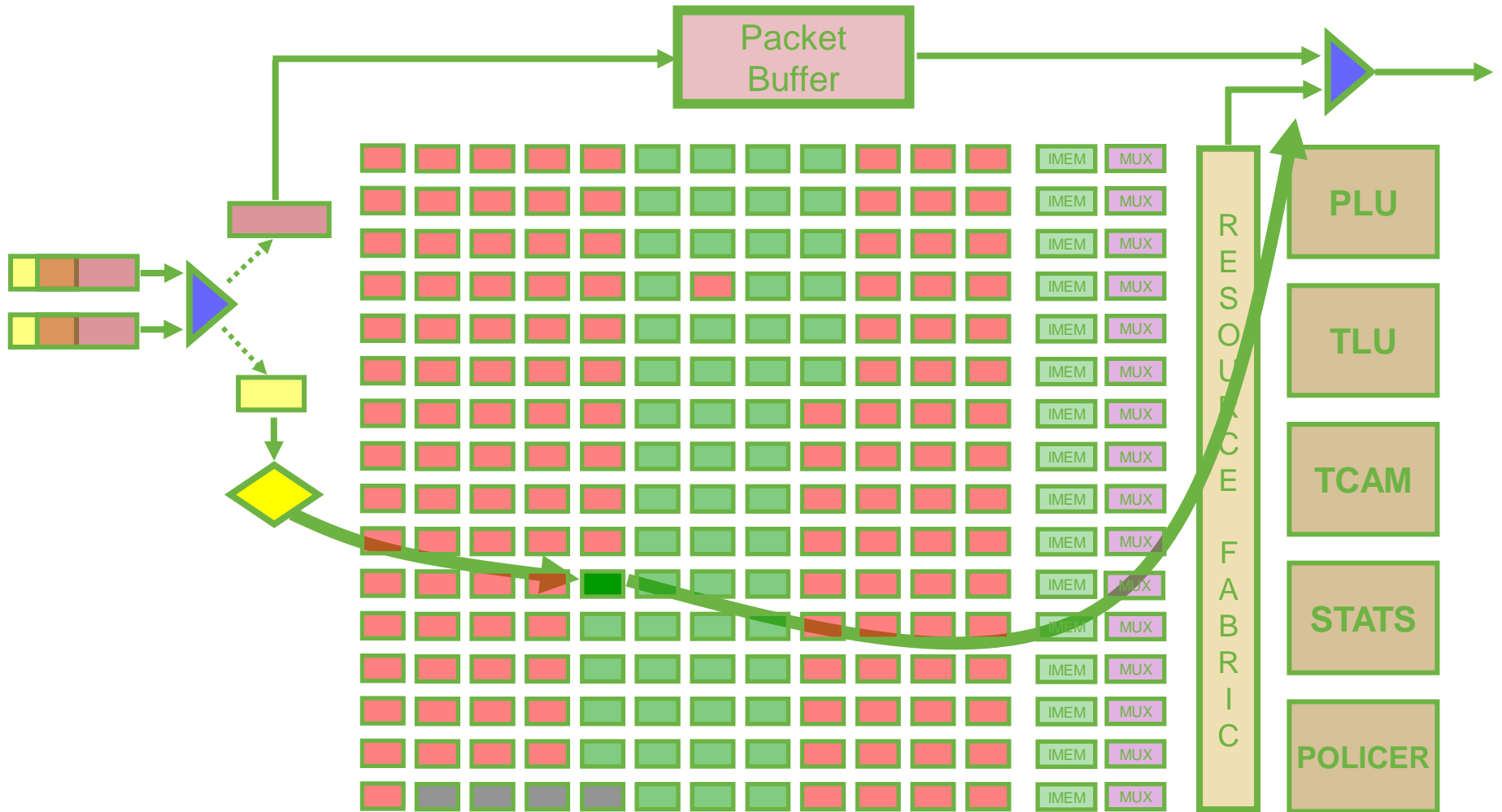
Perform Lookup and Features Using Resources



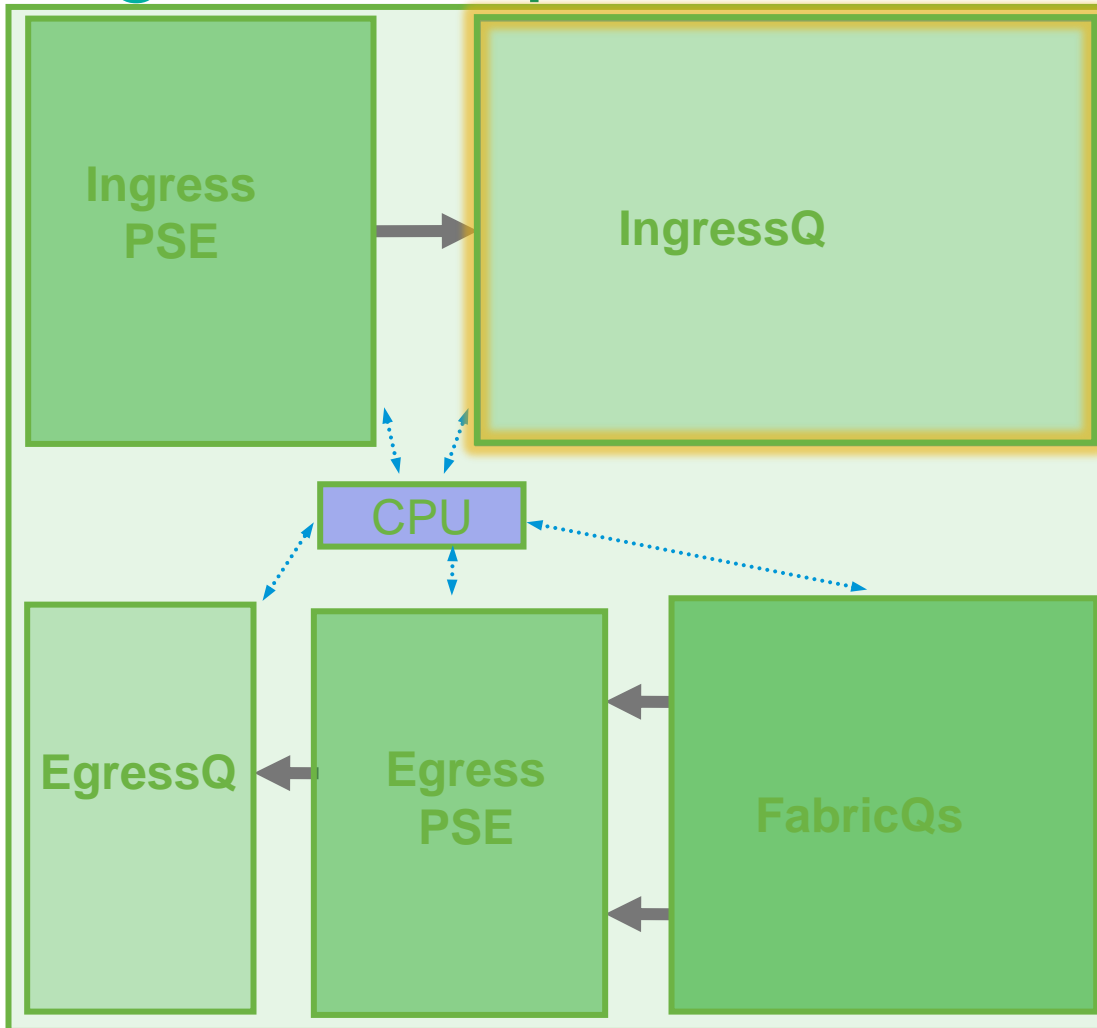


# Packet Path Within PSE

Recombine Header and Tail and Send to IngressQ



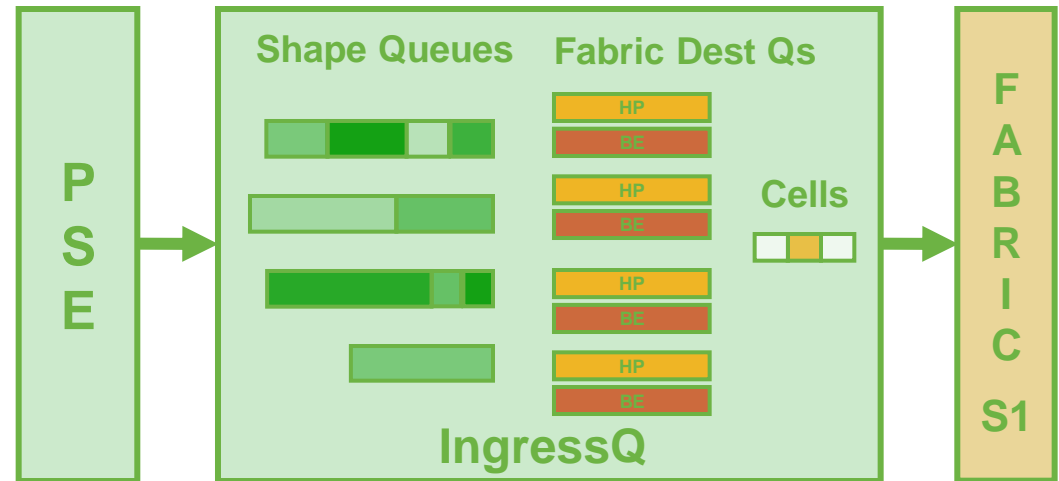
# IngressQ Topics



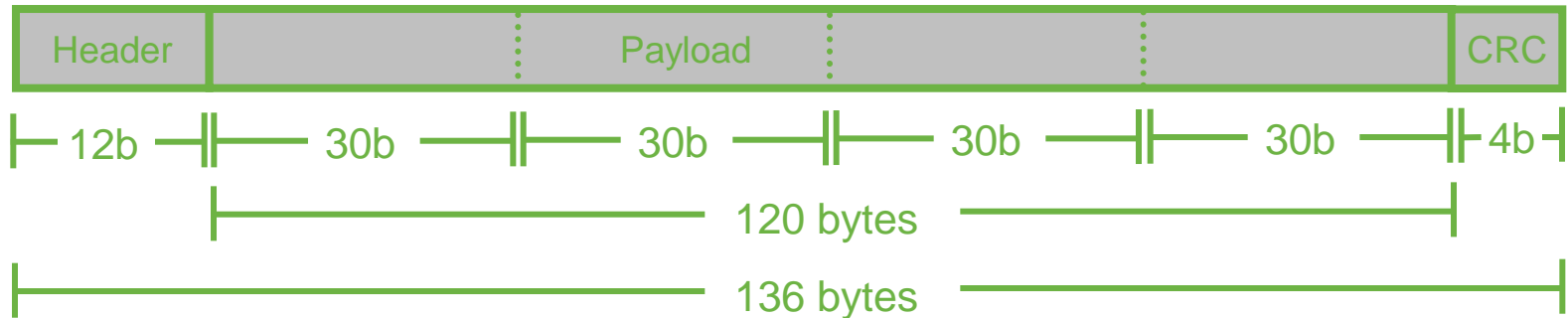
- Ingress Shaping
- To-Fabric Queuing
- Discard Bitmap\*
- Fabric QoS\*

# IngressQ

- Input Shaping Queues
  - Per Interface
  - HP & LP if configured
- Fabric Destination Queues
  - HP & LP for every FabricQ in system
  - 4 queues for every MSC in entire system
  - Queue determined by Ingress QoS or Fabric QoS
- Segmentation of packets into cells
- 45 Gbps limit between Shape Queues and Fabric Destination Queues
  - 140 Gbps in CRS-3
- Discard bitmap (discussed later but occurs between these two queues)



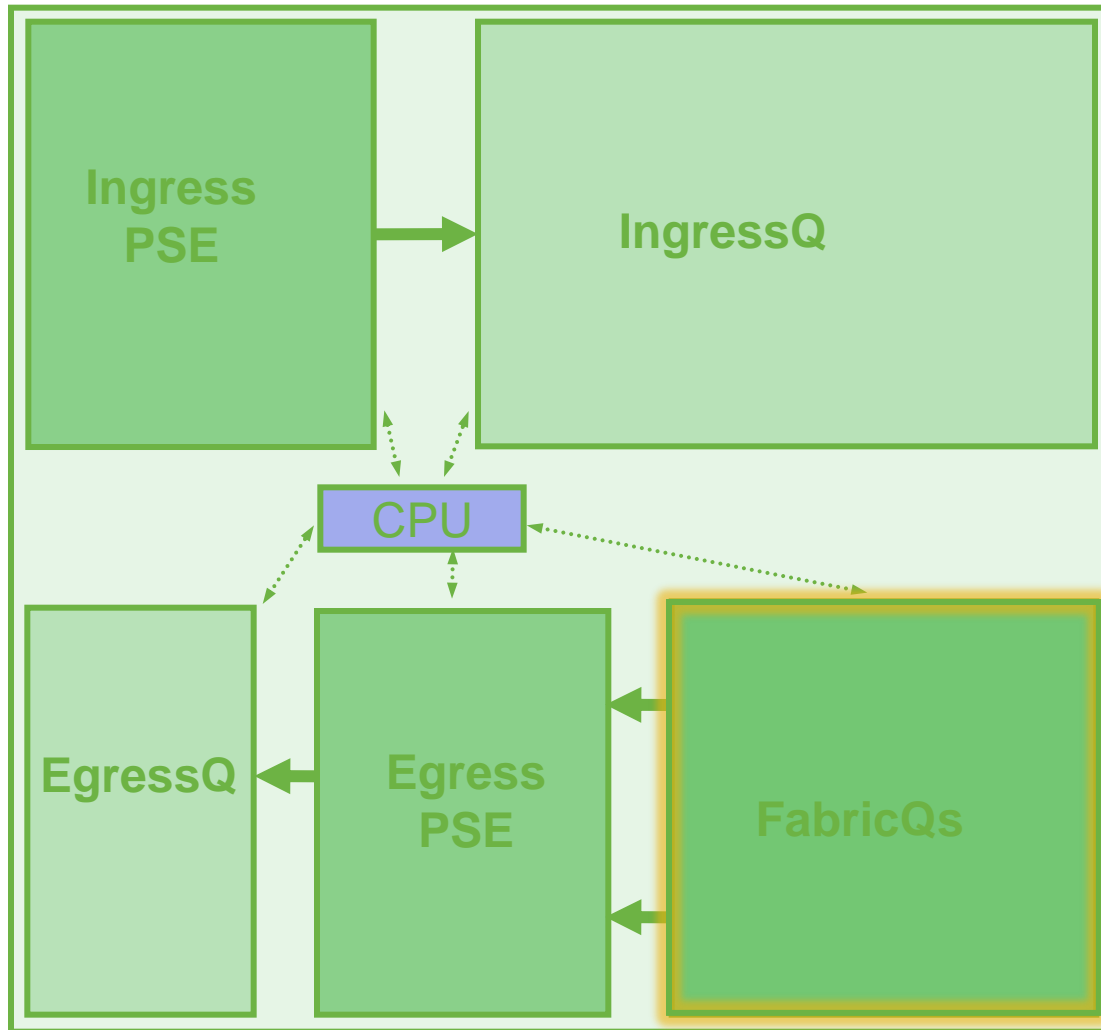
# CRS Fabric Cells



- 136 byte cells with
  - 12 byte header, 120 byte payload, 4 byte CRC
- 1 or 2 packets per cell
  - Packets must start on a 30 byte boundary
  - Packets sharing a cell must be same priority and cast
  - Entire cell travels over 1 fabric plane
- Round Robin among 4 or 8 fabric planes

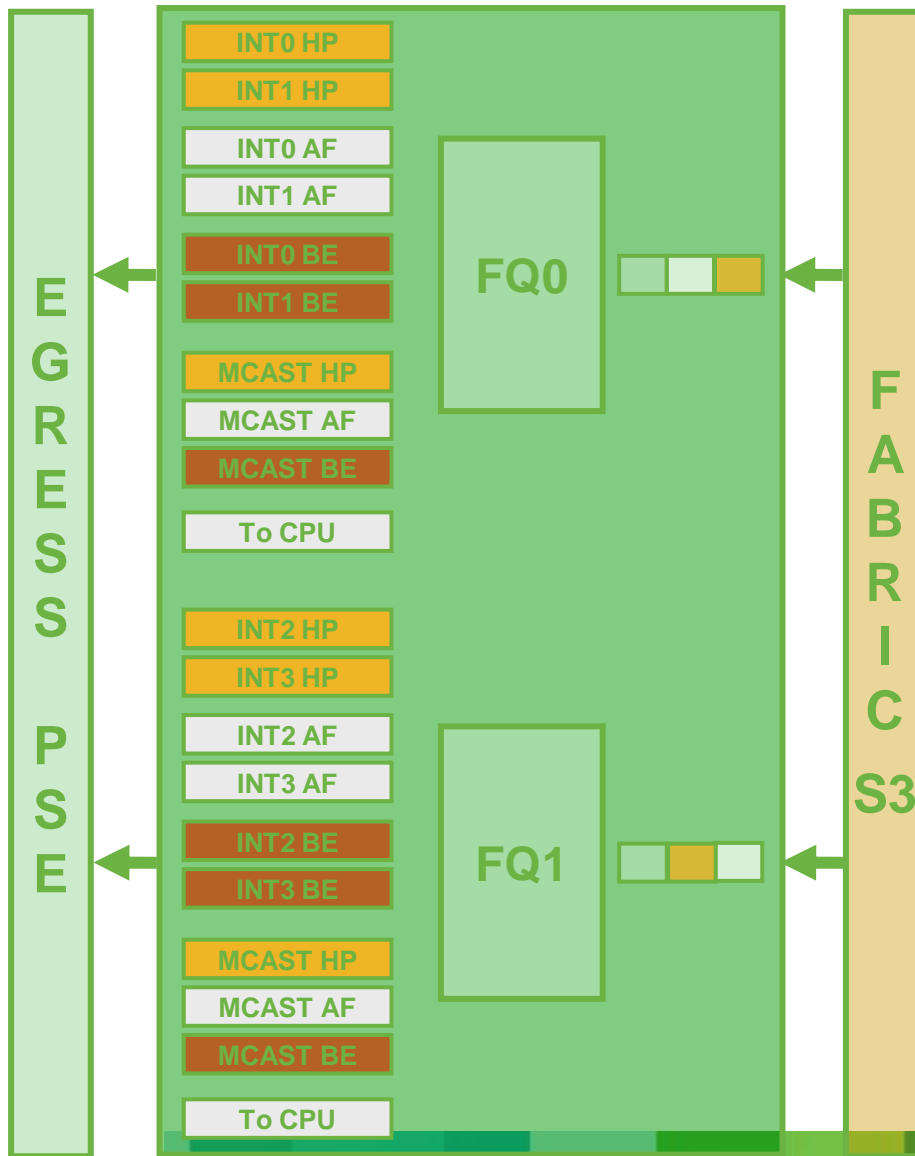


# FabricQ Topics



- FabricQ queues
- Monitoring queues
- Discard Bitmap
- Fabric QoS

# FabricQ Queues – Before Egress PSE



- Cells reassembled into packets
- Packets queued prior to PSE
- Unicast queues
  - Per type of service (HP/AF/BE)
  - Per output interface
- Multicast queues
  - Per type of service
- Raw (to CPU) queues
  - 8 queues



# CRS-3 Overview

## 140G/slot & 100GE technologies for CRS-1

### ■ 140G/Slot

- All models supported (4 slot, 8 slot, 16 slot, Multi-chassis)
- PLIMs (initial): 1x100GE, 14x10GE, 20x10GE (oversub)
- Service Cards: MSC140, FP140
- Fabric: Redundant 140G fabric

### ■ Varied Applications

- MSC140: Large Core, High Speed Edge apps
- FP140: Thin Core, Peering apps

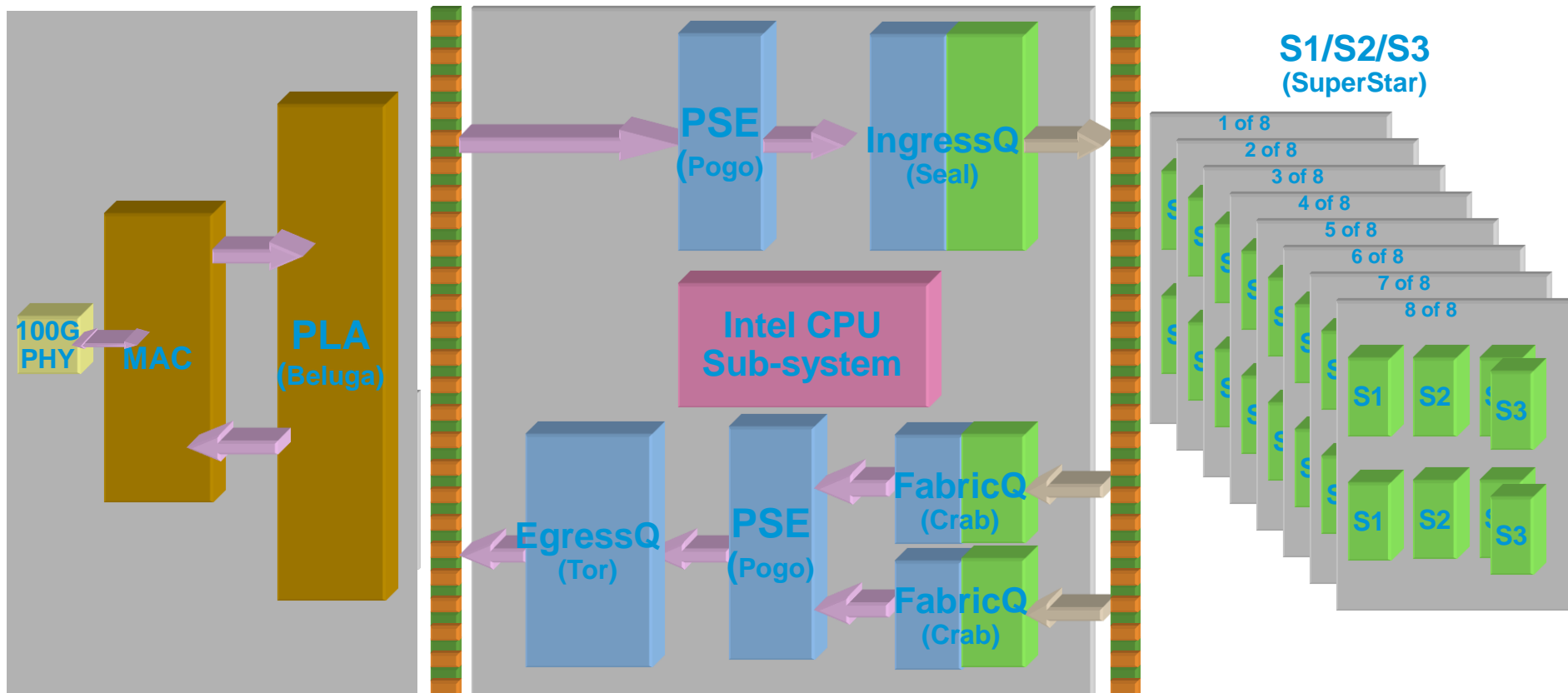
### ■ Rich Feature Set

- 3M+ prefixes
- 12,000 customer connections (VLANs, VPNs) with H-QoS (64k queues)
- L2 VPN (VPLS, VPWS), L3 VPN support
- Hardware-accelerated OAM protocols for higher scale & faster convergence
- Integrated hardware time-stamping for critical SLA monitoring (latency, jitter)



# CRS-3 System Overview

- Same architecture as CRS-1 but at 140G
- Compatible with all CRS-1 chassis sizes
- Standalone or Multi-chassis



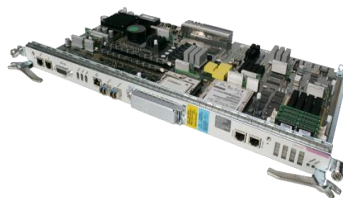
PLIM MSC/FP Fabric

# System Attributes Comparison

	MSC40	MSC140
Bandwidth	40 Gbps	140 Gbps
Max Packet per second	80 Mpps	125 Mpps
FIB Scalability	2M IPV4, 1M IPV6	4M IPV4, 2M IPV6
BW Modes	20/40 Gbps	40/140 Gbps
Queues/Groups/Ports	8k/2k/768	64k/16k/128
Supported Fabric Cards	40G and 140G fabric	140G fabric only
PLIMs	8 x 10GE 16xOC48, 4xOC192, 1xOC768 Modular 6 x SPA	14 x 10GE 20 x 10GE (oversubscribed) 1 x 100GE SONET and Modular PLIMs (future)
Ethernet Feature Set	Minimal	MTP support Shaped and CBFC pause frame
Other HW features	None	Timestamping Stateful feature support
Power	LC (MSC + PLIM): 530W Switch Card (4/8/16 slot): 102/185/206W Total (4/8/16-slot): 3.5/6.6/11.5KW	LC (MSC + PLIM): 600W Switch Card (4/18/16 slot): 49/90/94W Total (4/8/16 slot): 4/7.5/14 KW

# Route Processor Cards

- 2 major generations: RP and PRP
- Different for 16-slot Chassis and for 4-slot/8-slot Chassis
- PRP exists in 6GB and 12GB
- MUCH faster boot up and convergence time than legacy RP
- PRP mandatory
  - After 5.x.x and highly recommended in 4.3.x (limited scale)
  - For B2B
- PRP comes for free in bundles with Fabric
- Particular Case: DRP-CPU and DRP-PLIM



CRS-8/4-RP



CRS-8/4-PRP



CRS-16-PRP

# CRS-3 PLIMs and Line Cards

- At FCS, CRS-3 offered three types of PLIMs and two types of line cards

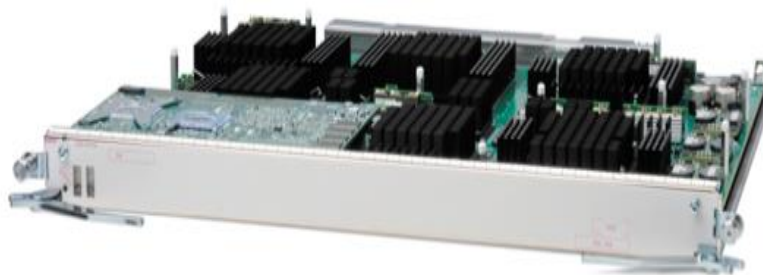
PLIM 14x10G

PLIM 20x10G

PLIM 1x100G Grey

MSC-140

FP-140



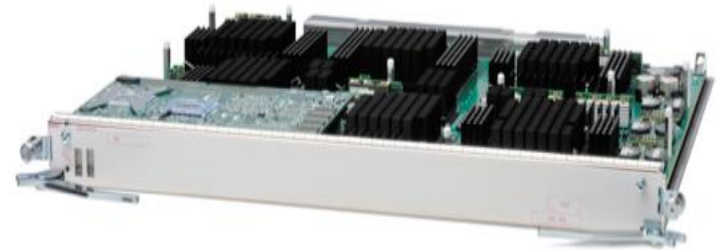
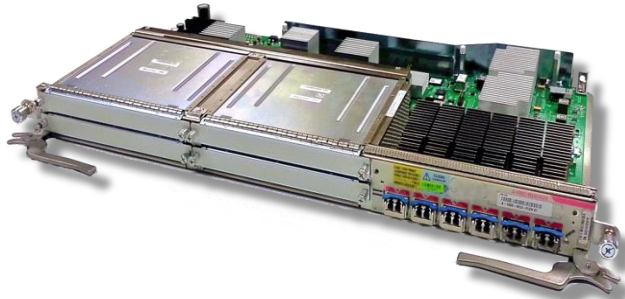
# CRS-3 PLIMs and Line Cards

LSP Line Card

Flex PLIM (4 x SPA + 6 x 10G)

PLIM 1 x 100G IPoDWDM

PLIM 4 x 40G OTN





# MultiChassis



# CRS-3 Full Rack MultiChassis

## Fabric Card Chassis (FCC)

- Optical Backplane
- Redundant Fans/Power
- Supports CRS-16

### FRONT:

- 24 Fabric cards
- 2 Shelf Controllers
- Control Ethernet Connections

### BACK:

- 24 Optical Interconnect Modules (OIM)
- 2 OIM LED Module
- Array Cable Connections



100m



## Line Card Chassis (LCC)

- Mid-Plane Design
- Redundant Fans/Power
- 140G per Slot

### FRONT:

- 16 Interface Slots
- 2 RP Slots
- 2 Controller slots

### BACK:

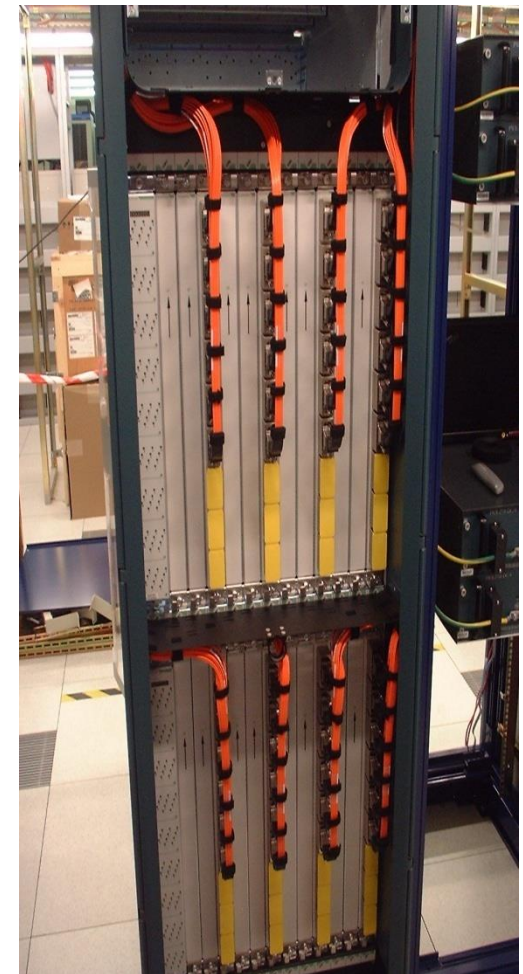
- 16 LC Slots
- 8 Fabric Card Slots



- Each Line Card Shelf add 4.48 Tbps to the MultiChassis system
- Fabric architecture can support 72 LC Chassis and 8 Fabric Chassis - 322 Tbps
- Hitless Single Chassis to MultiChassis Upgrade

# CRS Fabric Chassis (FCC)

- Front & rear access – mini-midplane for control network access
- Front
  - 24 Fabric card
  - 2 Shelf Controllers
- Back
  - 24 OIM slots
  - 2 Fiber LED modules
- Dimensions:
  - 23.6" W x 41" D x 84" H
  - (60 W x 104.2 D x 213.36H (cm))
- Power: ~9 KW DC, 11.1 KW AC
- Weight: ~1550 lbs/704kg
- Heat Dis.: 27600 BTUs

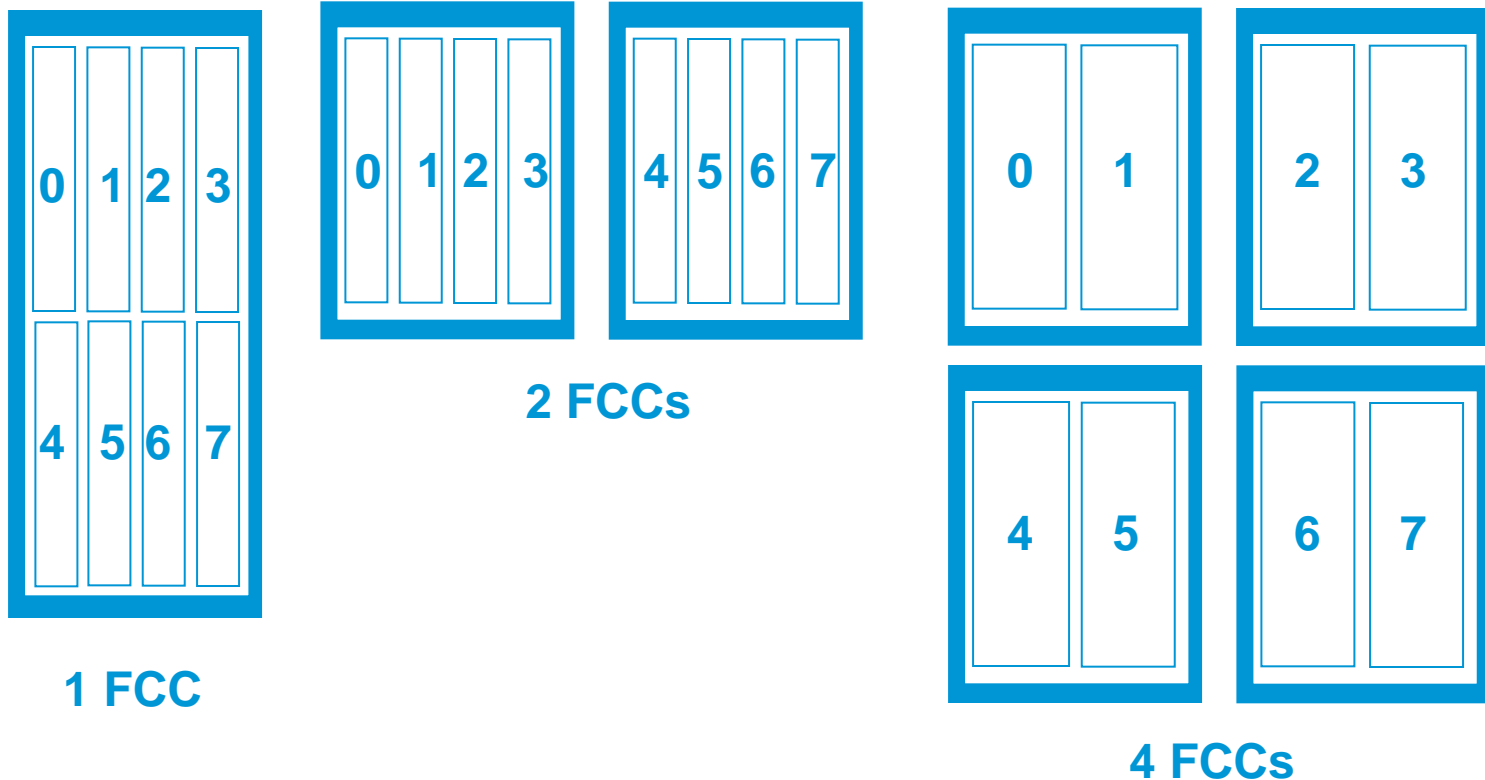


# MultiChassis technology building blocks

- Key points:

A fabric plane can span multiple S2 fabric boards

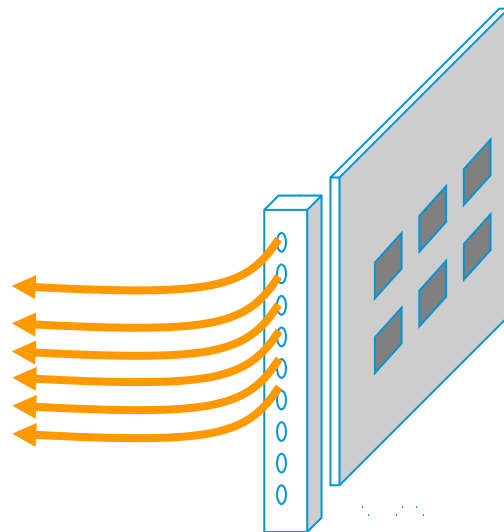
Multiple fabric planes can be supported in a Fabric chassis



# Fabric Plane Configuration

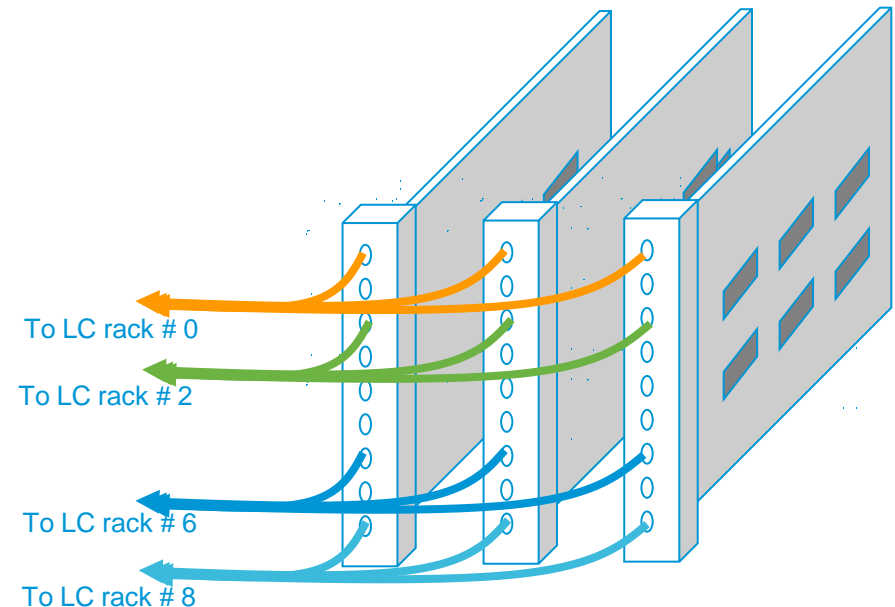
## Single Module Topology

- Full Plane Configuration
- Each S2 SFC serves one plane
  - The array cables for each fabric plane are connected to a common SFC
- Each S13 card is connected to a single S2 card



## Multi-Module Topology

- Three Plane Configuration
- Three S2 SFC's are used to create a plane
  - The array cables for each fabric plane are connected to all three cards
- Each S13 card is connected to 3 different S2 cards



Switch Fabric Card (SFC)

Switch Fabric Cards



# MultiChassis Cabling

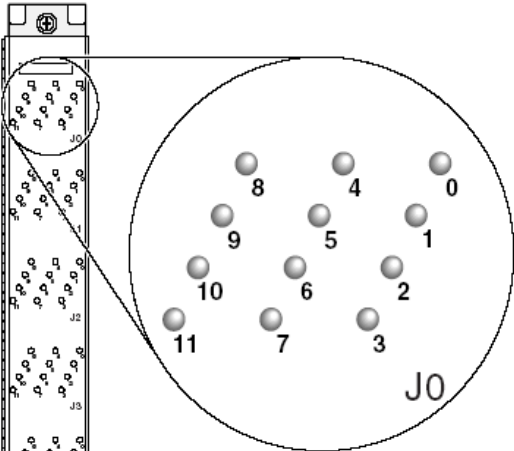
- 4+4 Configuration
  - 4 CRS-16 Line Card Chassis
  - 4 Fabric Card Chassis



- 2+1 Configuration
  - 2 CRS-16 Line Card Chassis
  - 1 Fabric Card Chassis



# OIM LED Module



OIM LED Module

- OIM LED Module provided to aid Operations
  - Two OIM LED Modules per Fabric Chassis
  - LEDs provide visual indication of the status of each array cable
  - Each tri-color LED shows 5 states:
    - OK (green)
    - no signal (none)
    - misconnected (blinking red)
    - signal fault (yellow)
    - connect here (slow blinking green)
- OIM LED Module provides an installation aid to help identify misconnected fiber bundles
  - Based on fabric configuration, a point-point connection map will be generated by the router
  - If only 1 fiber bundle is misconnected, specifies the correct place to connect or reconnect a fabric cable.

# CRS Switch Fabric Interconnect

## Optical Interconnect

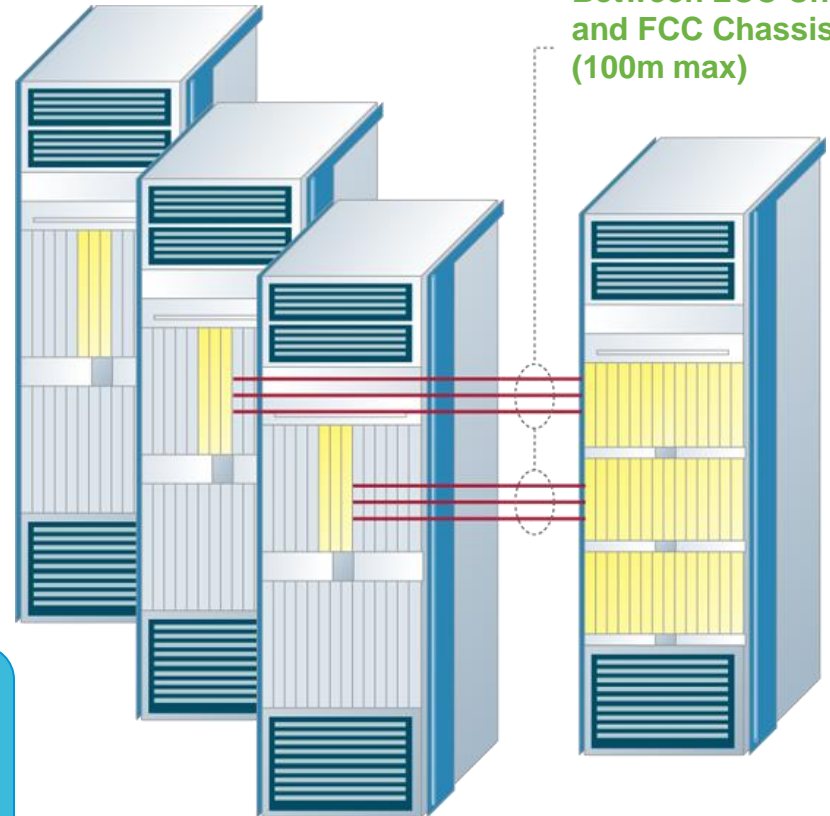
### FIBER BUNDLE

12 fibers per ribbon cable

6 ribbon cables per bundle = 72 fibers per Array cable



Array cable connections on an S13 card

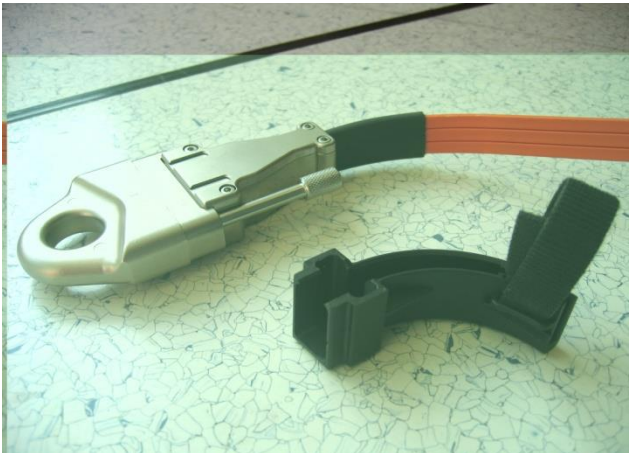
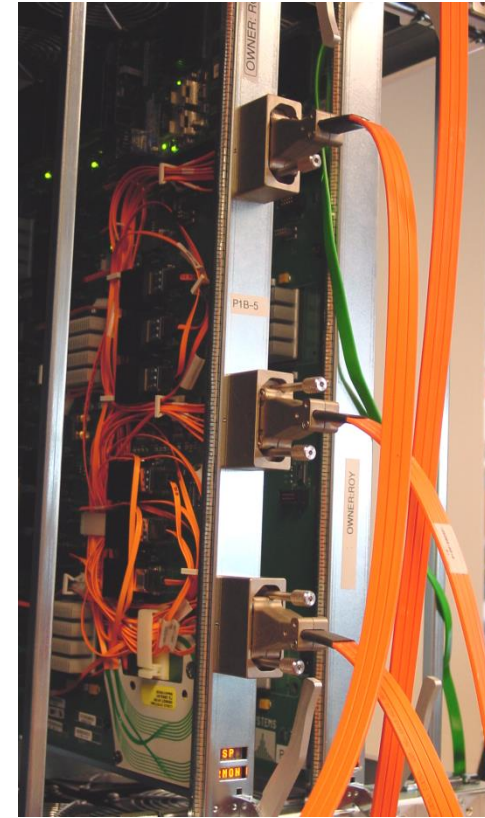


Multiple cables Between LCC Chassis and FCC Chassis (100m max)



# CRS Array Cables Details

- Array Cables connect FCC to the LCC
- 24 Array cables required for each LCC (3 per fabric plane x 8 fabric planes)
- Array Cable composed of individual Fibers
  - Each cable consists of 6 ribbon cables with 12 fibers each
  - 10m, 15m, 20m, 25m, 30m, 40m, 50m,60m,70m, 80m, 90m,100m length variants
  - Terminated at each end with a Square keyed Connector



# Array Cables



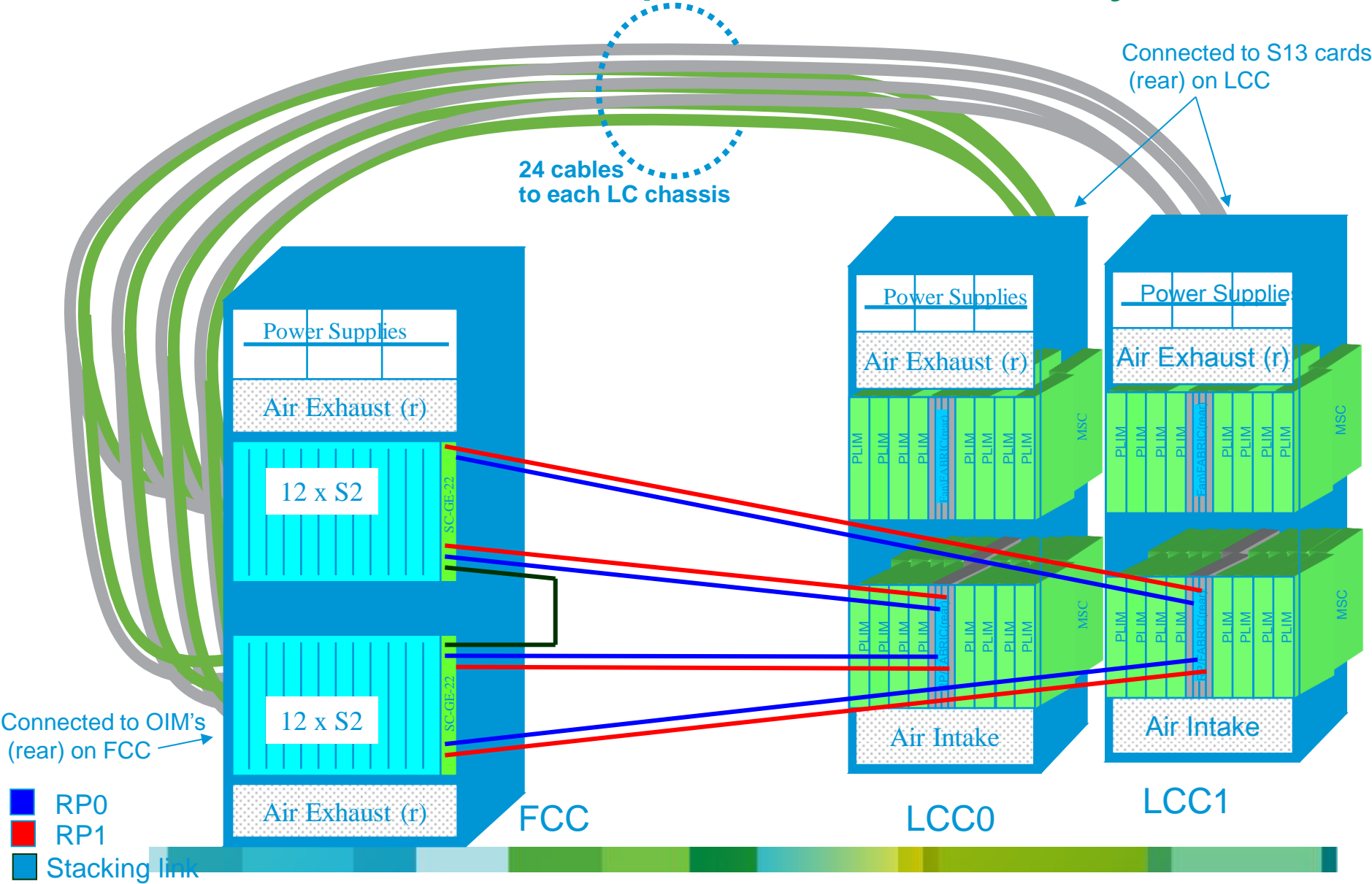
Turn Radius collar attachment

# CRS MultiChassis Control Ethernet

- **MC control Ethernet functions**
- **MC control Ethernet HW – Integrated Shelf Controller**
- **MC control Ethernet connectivity**



# 2LCC +1FCC MC Component connectivity via SC





# CRS-1 MultiChassis Control Ethernet

- All communication from the Line card RPs to integrated switch is over the Control Ethernet
  - The integrated switch is not connected to the fabric
- The Control Ethernet is used for many purposes
  - System Boot
  - Node availability (Heart beat) checks
  - All communication from the LCC to the FCC.
- The Control Ethernet is redundant and must be connected in a fully meshed configuration to all active and standby RPs and SCs
  - 2+1 Systems requires 9 cables – 8 RP to SCGE and 1 SCGE to SCGE
  - 2+2 System requires 15 cables – 8 RP to SCGE and 6 SCGE to SCGE
  - 2+4 System requires 36 cables – 8 RP to SCGE and 28 SCGE to SCGE
- The Control Ethernet uses Spanning Tree (STP) to determine which paths to use for communication

# Integrated Shelf Controller

- Product Description:

- Fabric chassis houses 2 Shelf Controllers (SC) acting as Primary and Secondary

- Provides local management of fabric chassis components

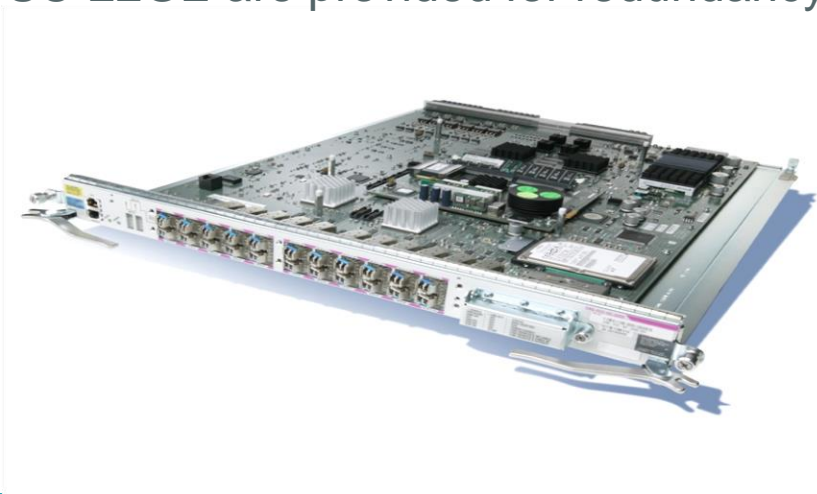
Boot and initialization of the SFCs, the optical interface module LED (OIM-LED) card, alarms, power supplies, and fans

- Integrated GE Ethernet Switch Interface

Provides out of band inter-chassis control network

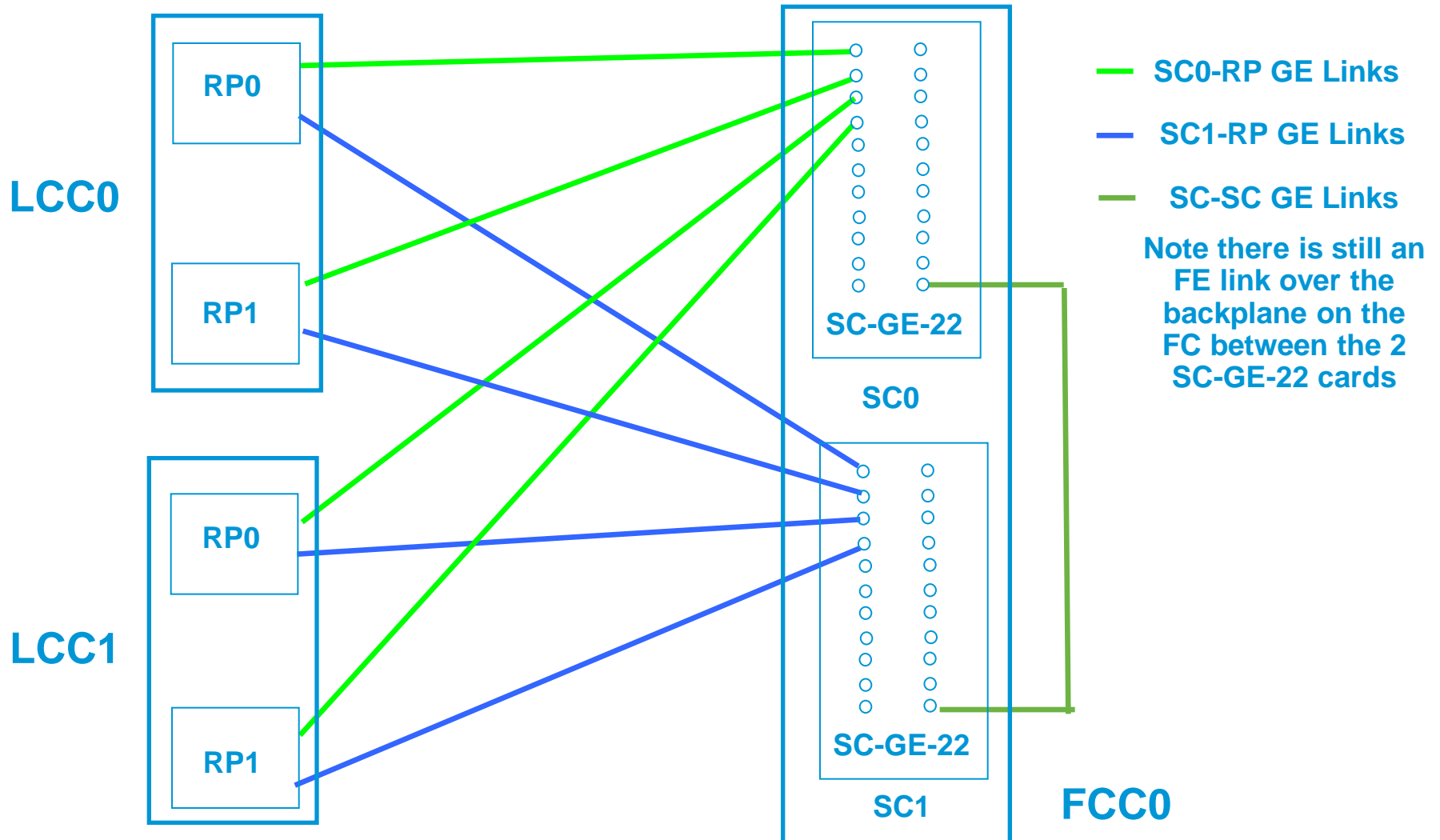
Full Mesh connectivity required for control Ethernet

- Two SC-22GE are provided for redundancy

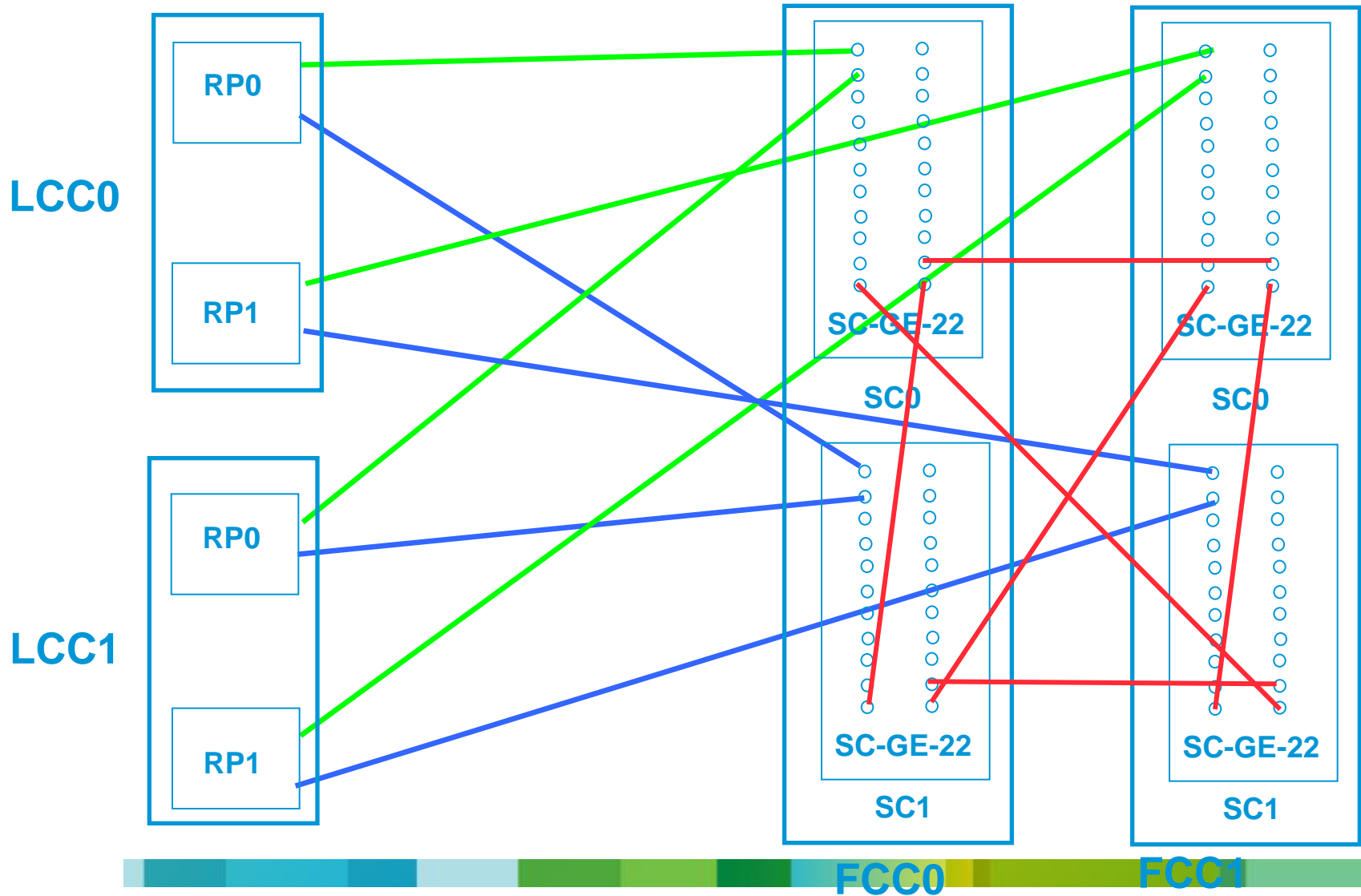


Product ID:  
CRS-FCC-SC-22GE

# 2+1 MC External GE Connections



# 2+2 MC External GE Connections





# CRS MultiChassis Configuration

- MC dSC
- MC SW distribution
- MC rack configuration
- MC fabric plane topology configuration



# CRS MC - Introducing the dSC

- dSC = designated System Controller
- dSC is responsible for overall system control, configuration and operation
- Image download & synchronization to all devices in system is controlled by dSC
- By default, dSC is the Primary RP in the first rack (LCC) that boots. If RP fails, Secondary RP assumes the role
- If the LCC housing the dSC were to fail, today, MC system will reboot and dSC will come active on one of other LCC's connected to the system
- Eventually, dSC functionality will be able to move between racks in a graceful manner
- No specific configuration is required to become dSC



# IOS-XR SW version on MC

- dSC determines the IOS XR version on all components of the system
  - Adding a new LCC with different IOS XR version will install the version running on the dSC
  - Upgrading from single to MC – the FCC will install the IOS XR version running on the dSC
  - Inserting new MSCs – same as on single chassis



# CRS MC – configuration Rack numbers

- Assign a rack number to the S/N of each chassis
- Serial numbers can be obtained from 'sh diag chassis'
  - From rommon with “dumpplaneeprom”
  - From rear of the FCC and front of the LCC
- Rack 0 or 1 = dSC (also an LCC)
- Rack 1->127 = LCC
- Rackf0-> f4 = FCC
- Note: Rack numbers have to be unique

Example:

```
RP/0/RP0/CPU0:CRS3#admin show run | i dsc
Building configuration...
dsc serial TBA10080000 rack 1
dsc serial TBA10090001 rack 0
dsc serial TBA10120000 rack f0
```



# CRS MC Fabric – Config Plane Topology configuration

- Planes are not tied to specific slots in fabric rack
- Admin-level config defines the slots each plane uses (and how big the plane is)

```
controllers fabric plane 0
```

```
oim count 1
```

```
oim width 1
```

```
oim instance 0 location F0/SM0/FM
```

**Count 1- All cables in plane connect to the same OIM. Count 3 - The cables from each LCC for that plane connect to different OIMs**

**Position of 1<sup>st</sup> card within the plane in fabric rack:  
plane 0 uses rack f0 slot sm0**

# CRS MC Fabric Configuration Example

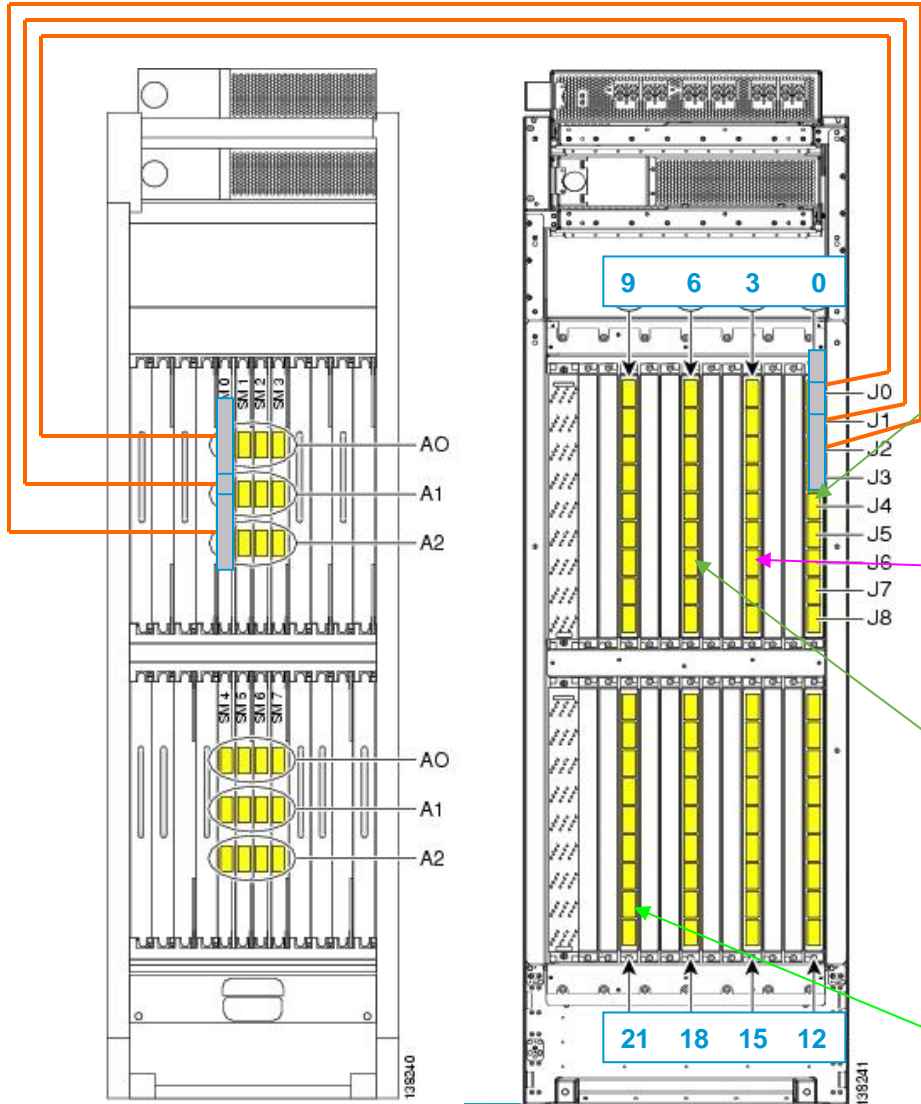
```
RP/0/RP0/CPU0:CRS3#admin show run
Building configuration...
```

```
dsc serial TBA10080000 rack 1
dsc serial TBA10090001 rack 0
dsc serial TBA10120000 rack F0
controllers fabric plane 0
  oim count 1
  oim width 1
  oim instance 0 location F0/SM0/FM
```

```
!
controllers fabric plane 1
  oim count 1
  oim width 1
  oim instance 0 location F0/SM3/FM
```

```
!
controllers fabric plane 2
  oim count 1
  oim width 1
  oim instance 0 location F0/SM6/FM
```

```
!
[SNIP]
controllers fabric plane 2
  oim count 1
  oim width 1
  oim instance 0 location F0/SM21/FM
!
```

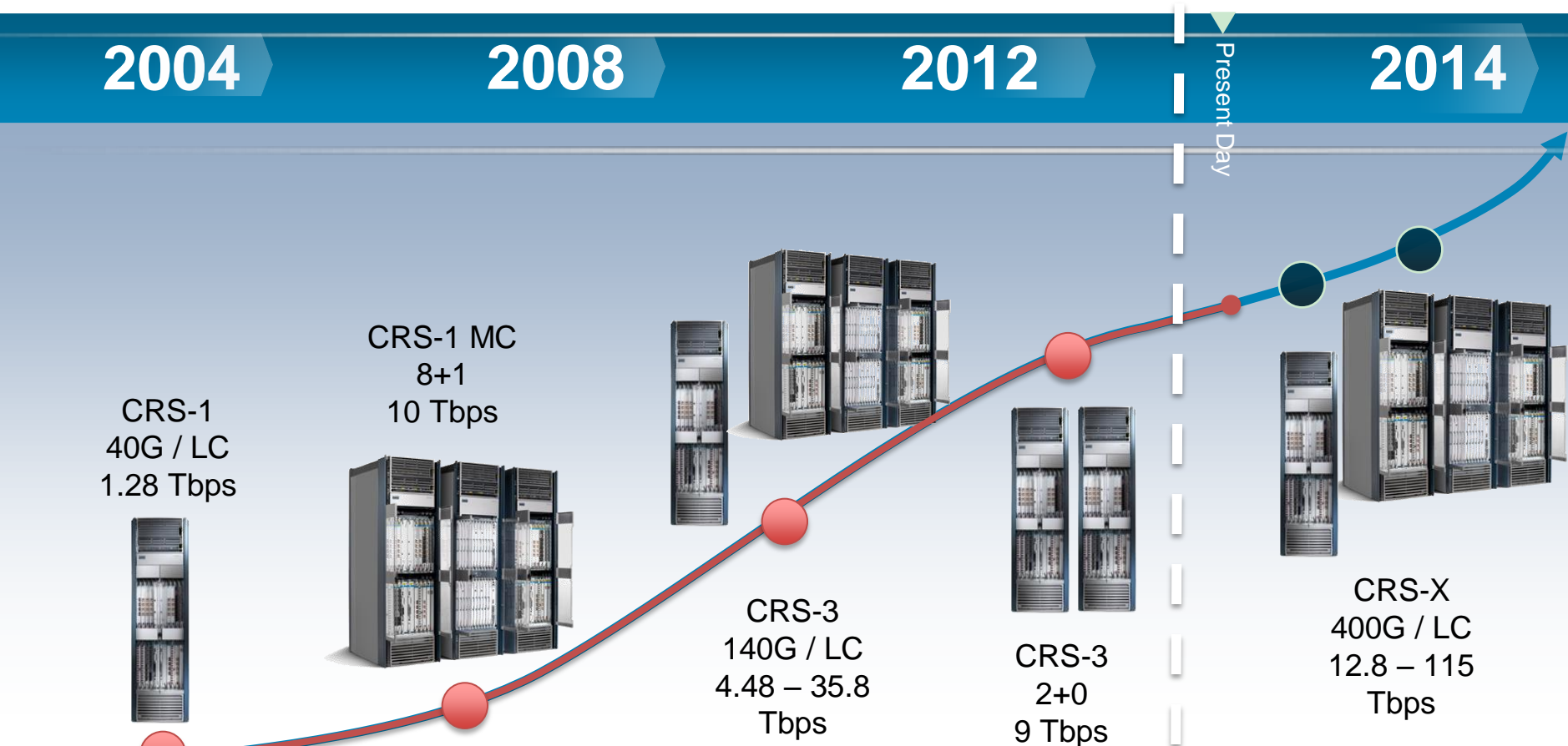


# CRS-X Overview



# Investing in the Future

10x Capacity Gains over 10 Years



- Future proof architecture (forwarding-, switching-, control-plane separation) allows hitless evolution from 40G 140G 400G and from Single Chassis



B2B and MC

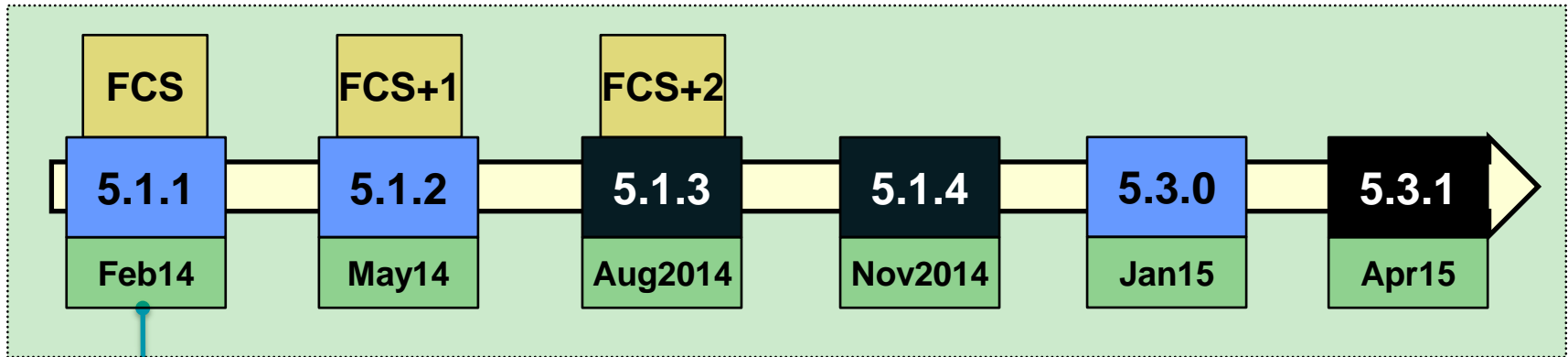




# CRS-X Program Planning

Single Chassis Only  
 B2B, B2E and MC

- IOS XR planning and CRS-X program phases Software/Hardware



Single Chassis  
 40x10G 4x100G



x100+5x40 B2B 16S  
 MC up to 4+1



MC up to 8+2



B2B 8S

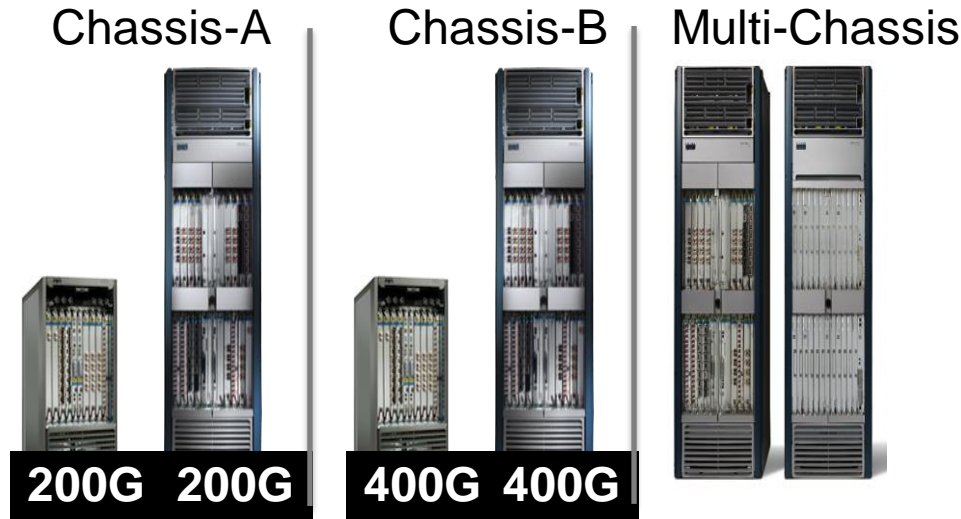


First phase: only 4 LCC supported with FCC-400G

# CRS-X Program Overview

400G per Slot in CRS Chassis

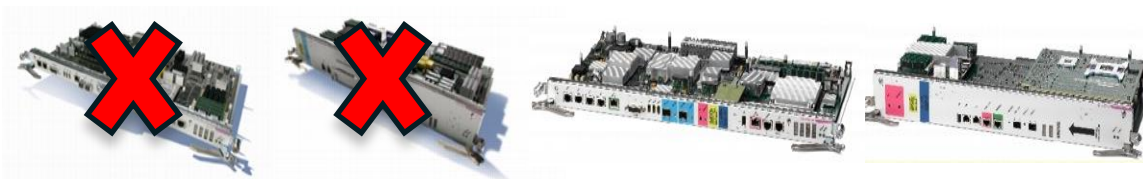
- Support on 16-slot legacy chassis (**200G**), on 16-slot enhanced chassis (400G), on 8-slot legacy chassis (**200G**), on 8-slot enhanced chassis (400G)



- but NOT on 4-slot chassis** because of hardware limitations.



- PRP route processor cards will be mandatory, **RP are not supported**



# CRS-X Program Overview

400G per Slot in CRS Chassis

- At FCS, with 5.1.1.  
We will support 8-slot and 16-slot chassis



- At FCS+2 (5.1.3), we'll add the support of B2B16-slot and MultiChassis (up to 4+2, larger systems will be supported later)



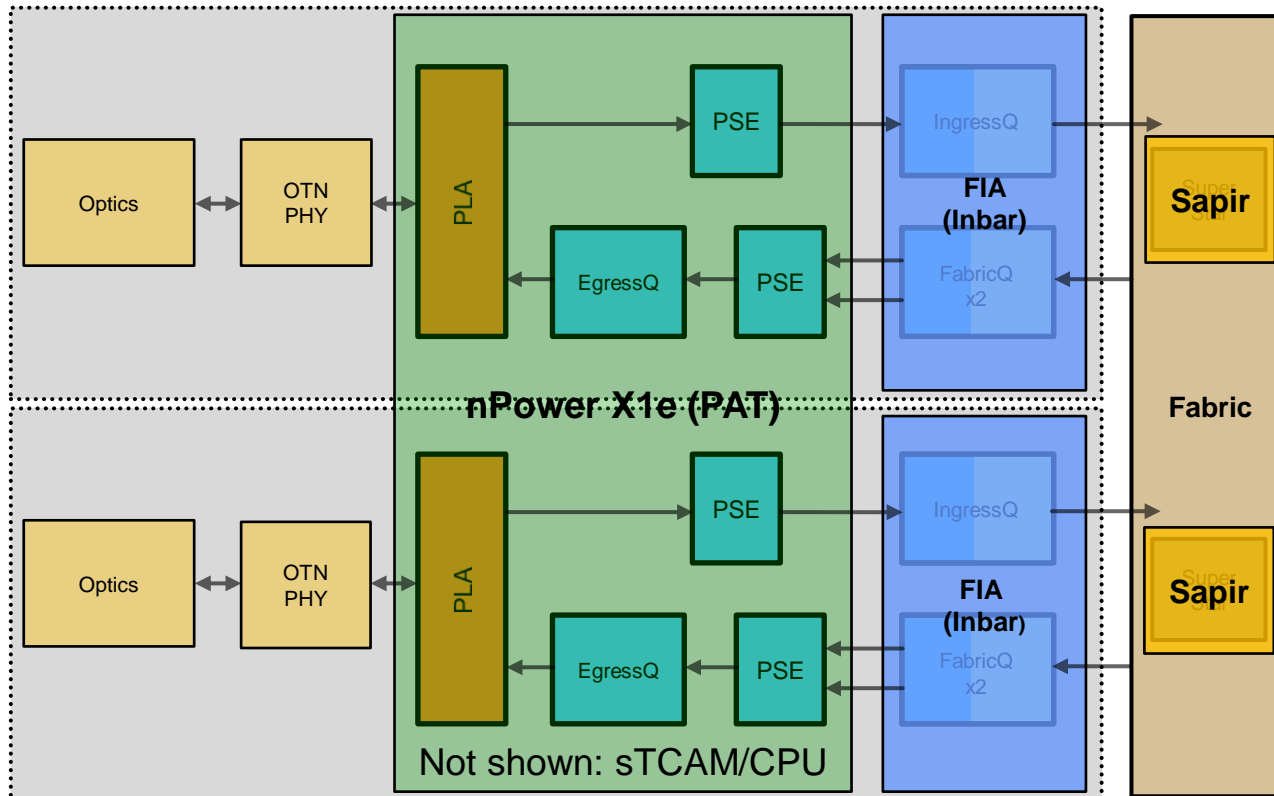
- In 5.3.1, we'll add the support of B2B 8-slot



# CRS-X ASIC Integration

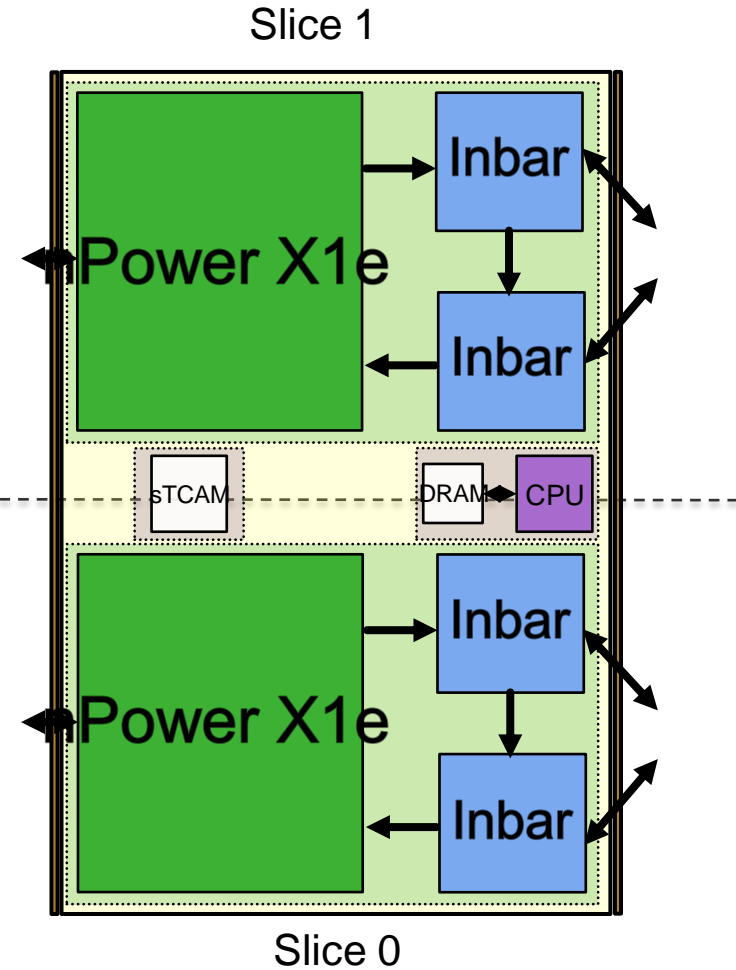
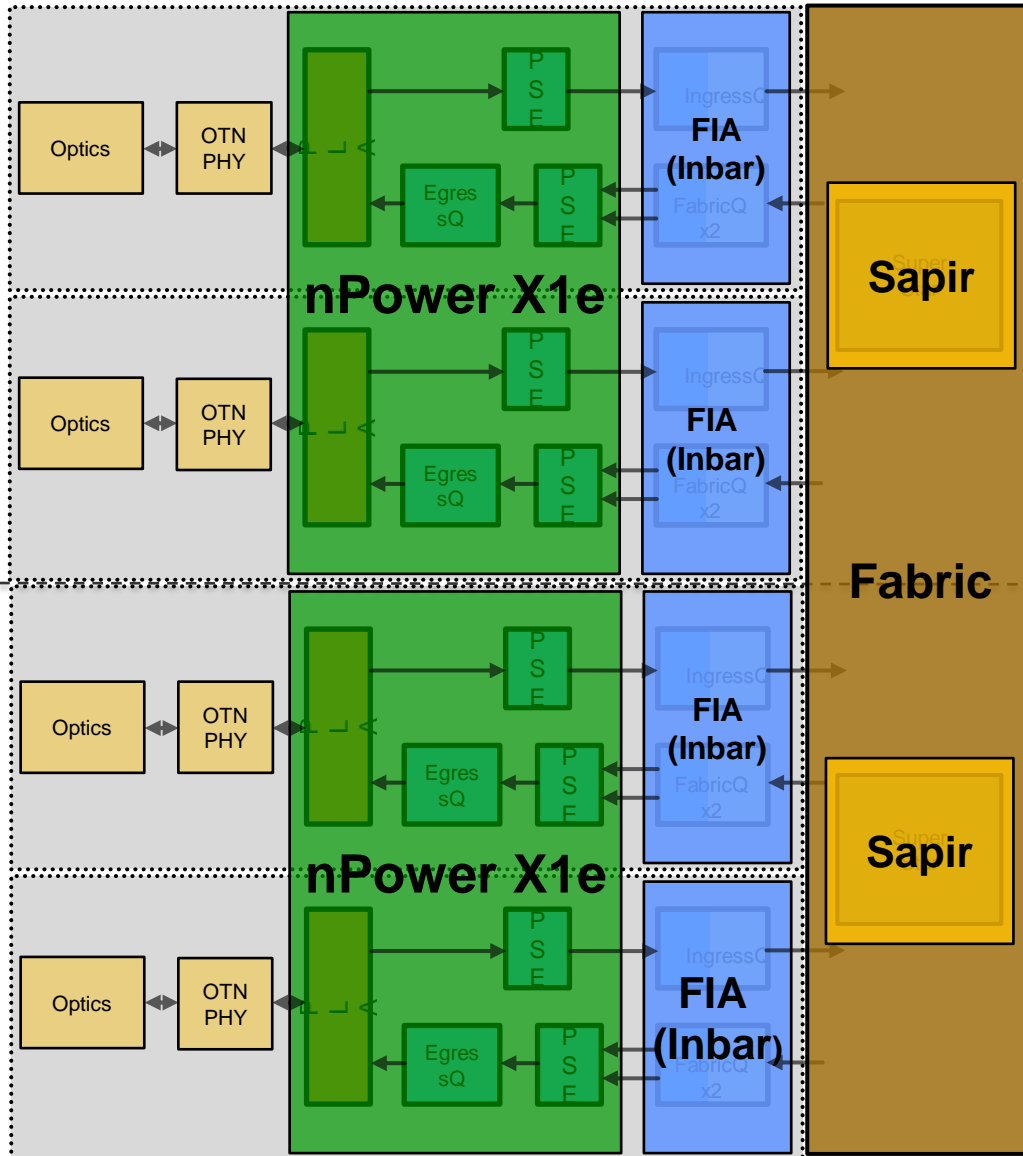
## New Generation ASICs Integrates Former Roles

- CRS-X collapses former CRS-3 ASICs in a lower number of elements
- nPower X1e (PAT) replaces 2 x PLA + 2 x EgressQ + 4 x PSE
- Inbar replaces IngressQ + FabricQ
- Sapir replaces the SuperStar fabric ASICs
- sTCAM @ 500MHz replaces TCAM4 @ 400MHz
- Atris Memory LLDDRAM3 @ 800MHz replaces QDR BL2 and BL4



# Introducing the Slice concept

Two slices per line card



# SC-GE-22

Code name: "Trishul"

- Prior to perform MultiChassis fabric upgrade to CRS-X, it will be mandatory to replace SC-GE-22 cards in FCC by newer version: Trishul boards.
- Intel-based cards. Replacement driven by Lead(Pb)-free requirements
- Legacy (PPC based) and newer boards (x86 based) can not be mixed.
- It will NOT be necessary to shutdown the FCC to replace them
- Trishul boards will re-use the existing Kensho-GL daughterboard
- Part-number: SC-GE-22-B
- PID: CRS-FCC-SC-22GE-B



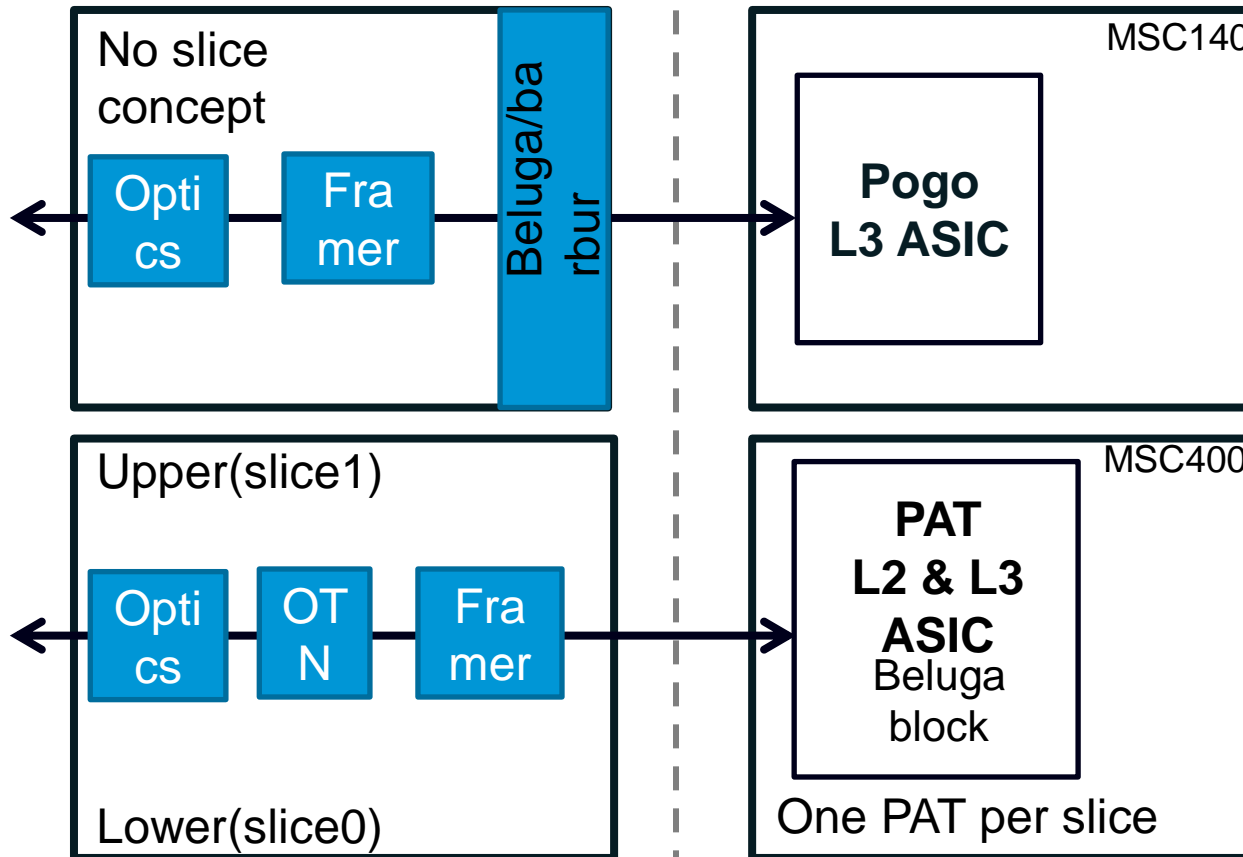
# CRS-X Hardware Details

## PLIM



# CRS-3 vs CRS-X PLIMs

- CRS-3: PLIM is L1+L2
- CRS-X: PLIM is L1 only, L2 has migrated to PAT

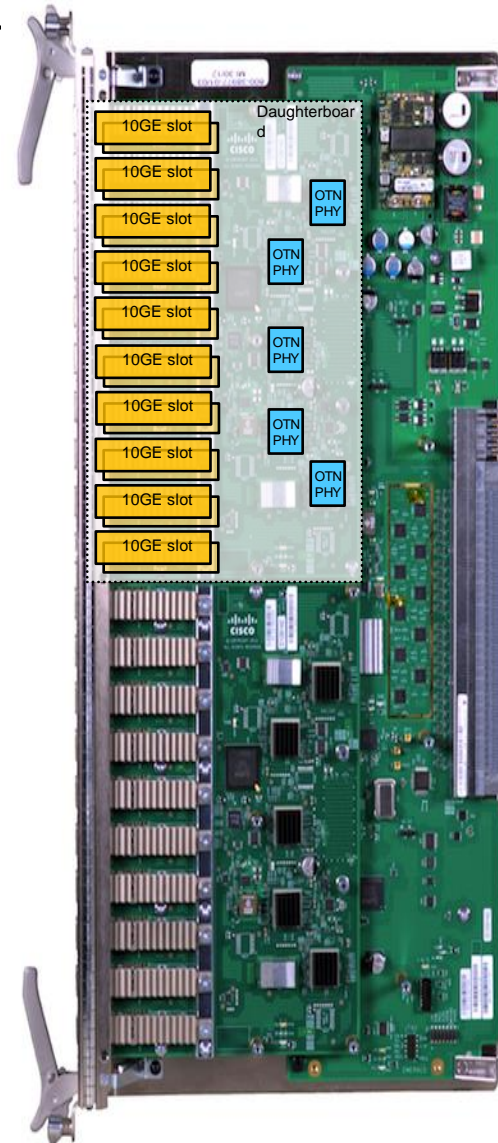
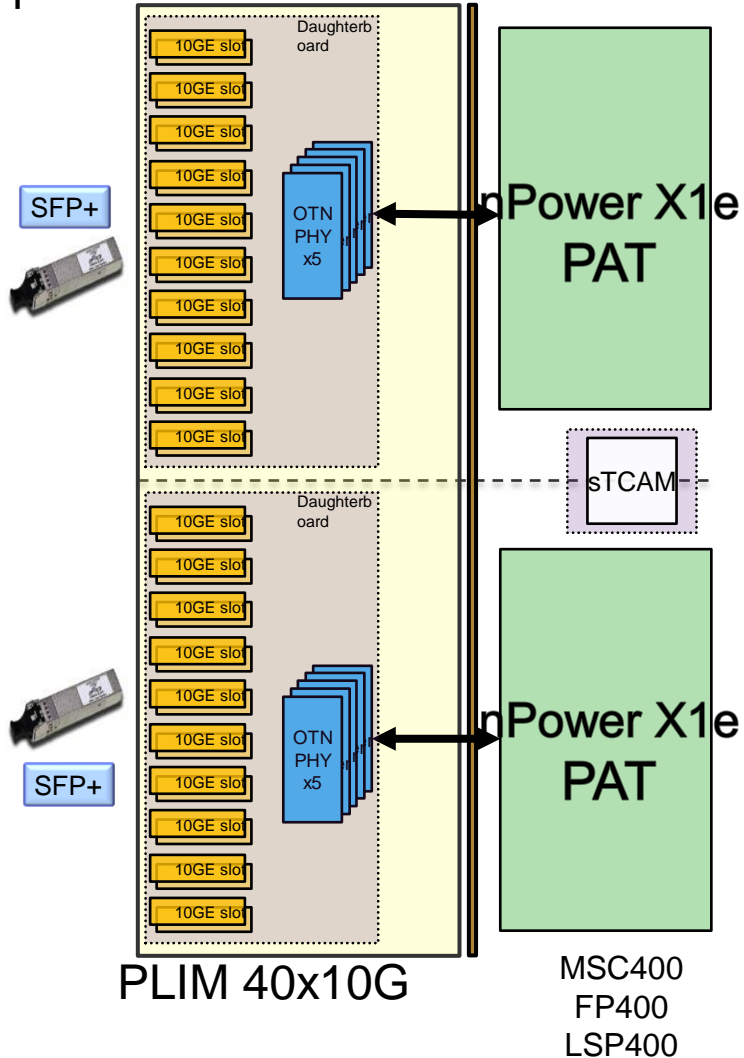




# PLIM 40 ports 10GE

40x10GE-WLO

- 40 SFP+ slots with LAN, WAN and OTN (G.709 F support).



# PLIM 40 ports 10GE

40x10GE-WLO

- Available at FCS (5.1.1)
- Physical layer functions (including OTN Framing)
- Standards: WAN/LAN PHY, G.709 OTU
- FEC: G.709 FEC , G.975 I.4 EFEC (AP)
- G.975 I.7 EFEC (Cortina)
- SFP+ Optics



(Printings/color on the face plate are not accurate)

Supported SFP+	Distance	Roadmap	Power
SFP-10G-SR / SFP-10G-SR-X	Up to 400m	5.1.1	1W
SFP-10G-LR / SFP-10G-LR-X	10km	5.1.1	1W
SFP-10G-LRM	Up to 300m	Target 5.5.0	1W
SFP-10G-ER	40km	5.1.1	1.5W
SFP-10G-ZR	80km	5.1.1	1.5W
DWDM Tunable	Up to 70km	Target 5.5.0	1.8-2.2W

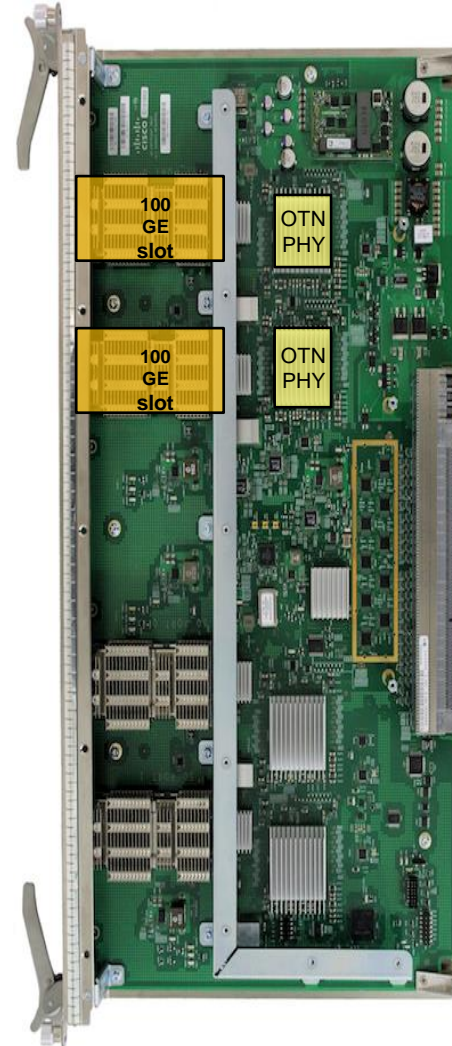
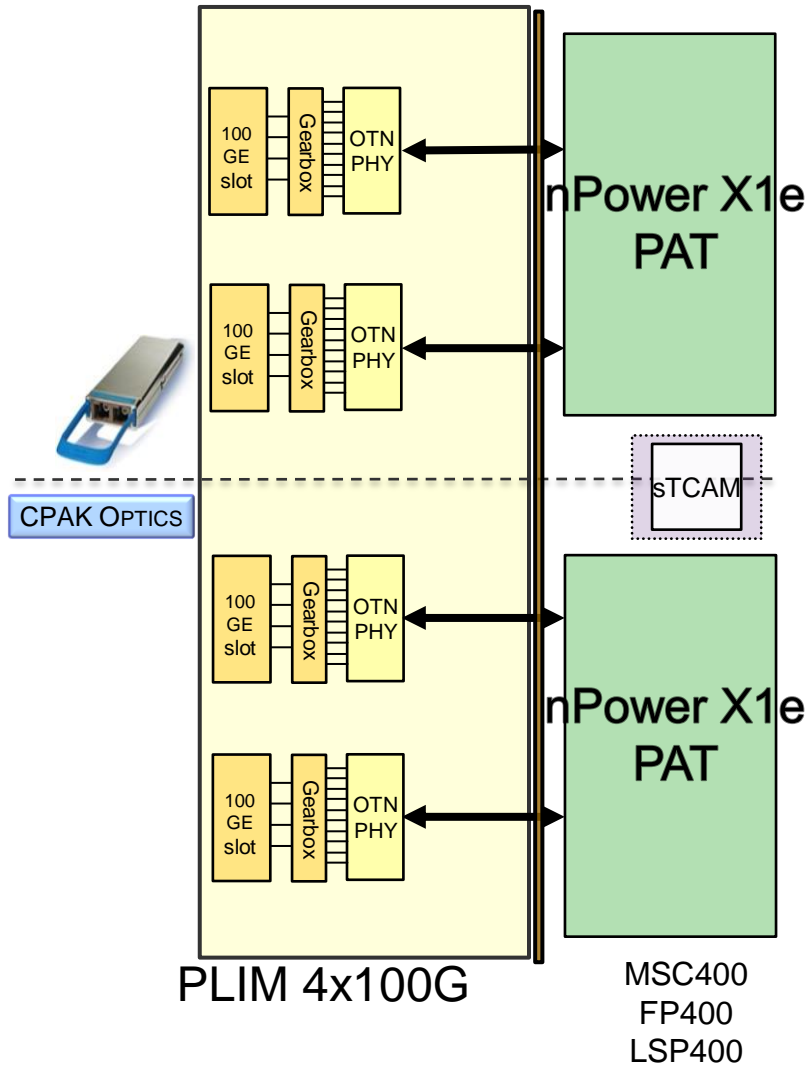


- Low-powered and high-powered SFP+  
We plan to support a fully populated P

# PLIM 4 ports 100GE

4x100GE-LO

- 4x100GE-LO leverages the low power consumption and reduced form-factor of our CPAK optics.

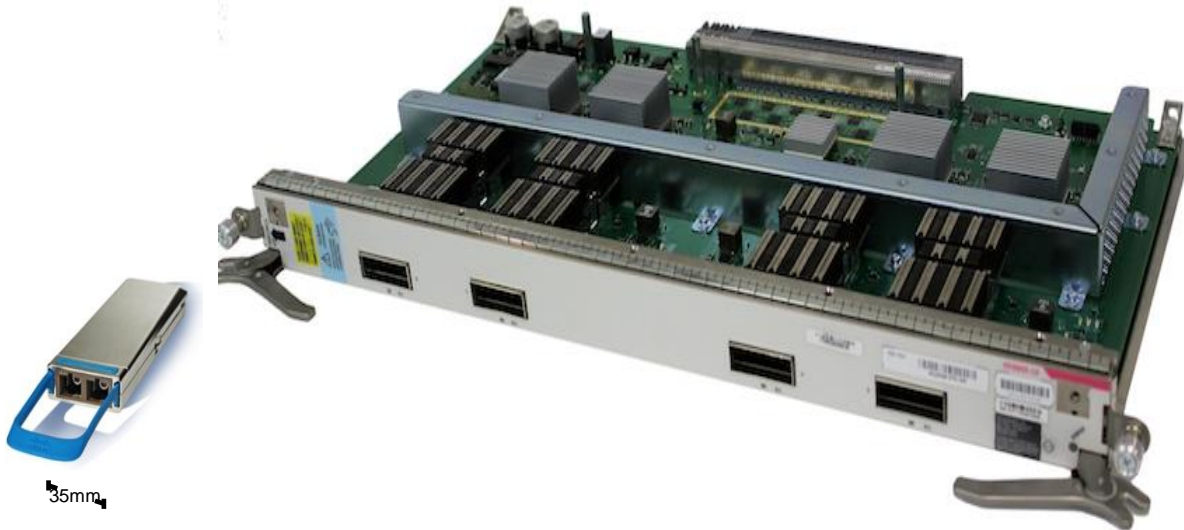


# PLIM 4 ports 100GE

4x100GE-LO

- Available at FCS (5.1.1)
- Physical layer functions (including OTN Framing)
- Standards: 100GBASE-R LAN PHY, OTU-4, G.709 FEC
- CPAK Optics (Lightwire acquisition)

Supported CPAK	Roadmap
CPAK 100G SR10	5.1.1
CPAK 100G LR4	5.1.1
CPAK 100G ER4 Light	Target 5.3.0
CPAK 100G CSR10	Future
CPAK 100G SMF-SR	Future



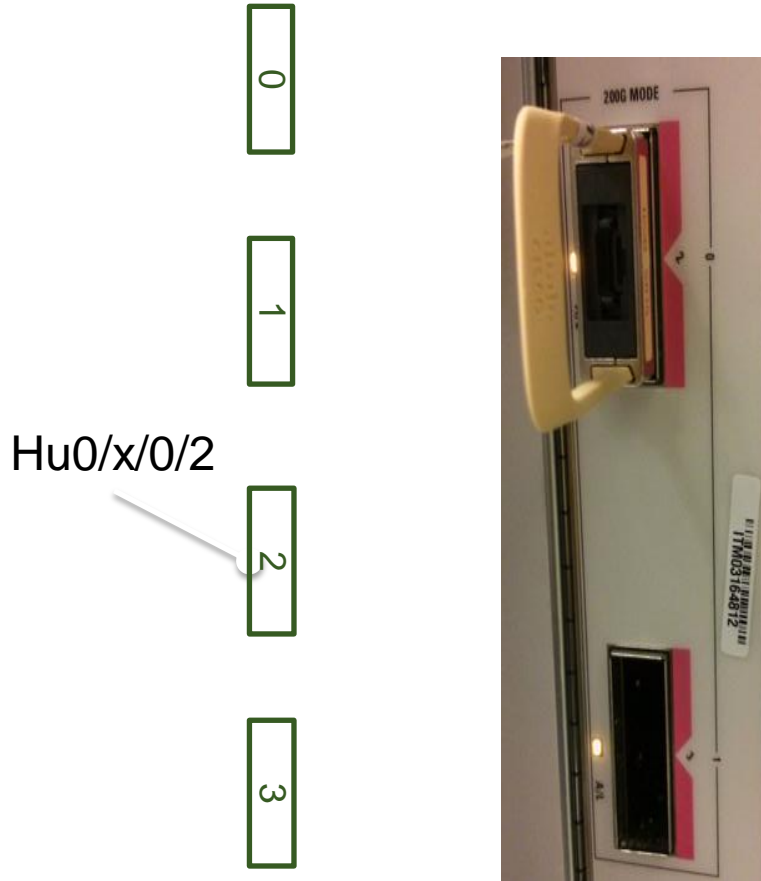
- Note: ER4lite supports up to 25km

# PLIM 4 ports 100GE

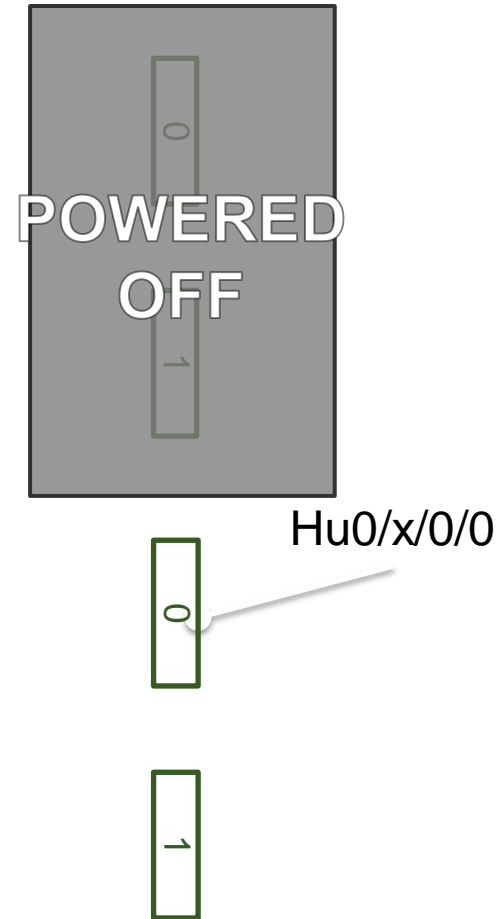
4x100GE-LO

- Port Numbering

Enhanced Chassis



Legacy Chassis

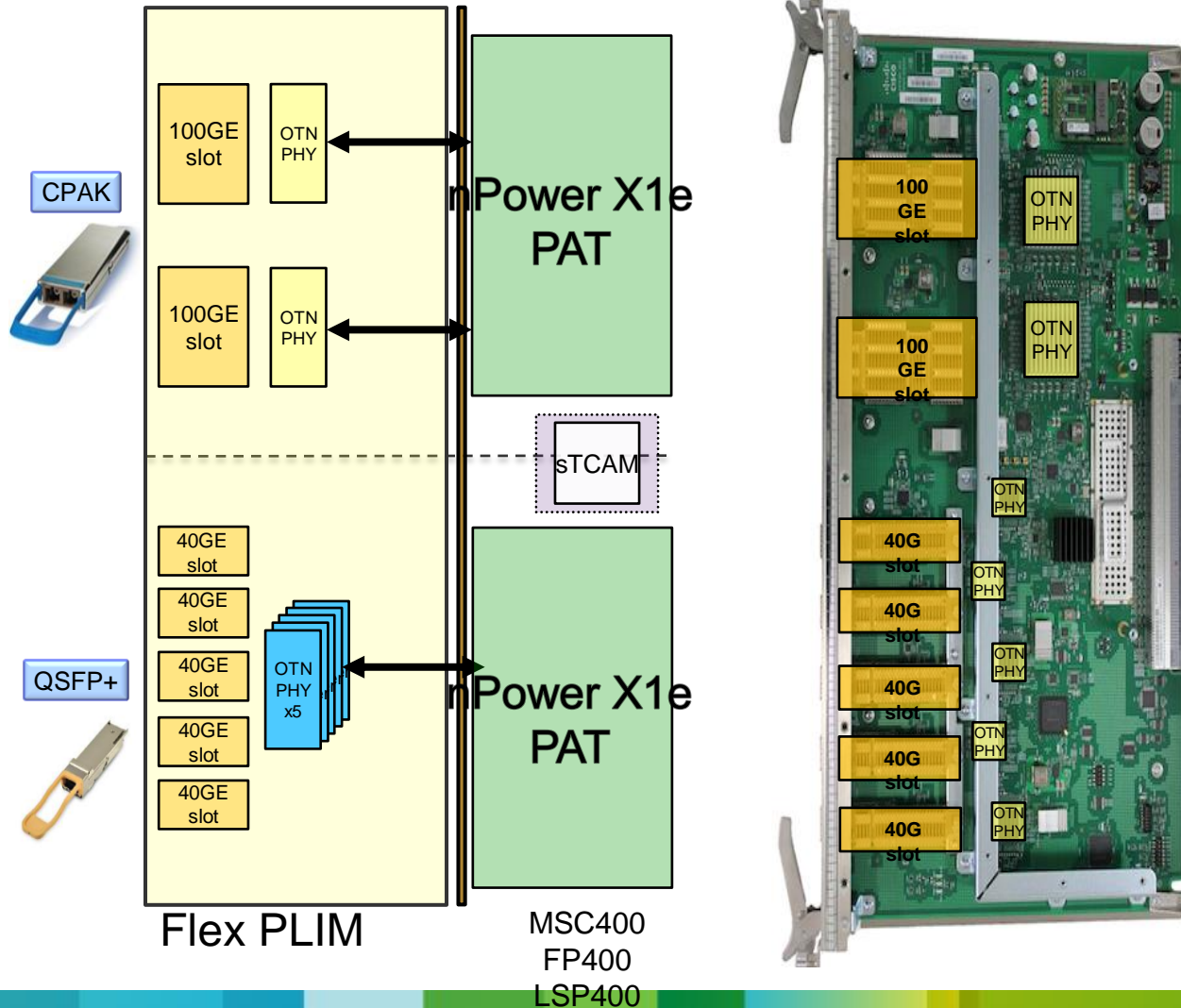




# FLEX PLIM 2 ports 100G / 5 ports 40G (5.1.3)

2x100GE-FLEX-40

- 2x100GE-FLEX-40 will offer 100G via CPAK, 40G via QSFP+ or more flexibility via 40G breakout-cables and patch panel boxes



# FLEX PLIM 2 ports 100G / 5 ports 40G

2x100GE-FLEX-40

- Available at 5.1.3
- Standards: 100GBASE-R, G.709 OTU-4 // 40GBASE-R, G.709 OTU-3
- FEC: G.709 FEC
- CPAK: cf 4x100G PLIM



Supported Optics	Roadmap
CPAK 100G SR10	5.1.3
CPAK 100G LR4	5.1.3
CPAK 100G ER4 Light	Future
CPAK 100G CSR10	Future
CPAK 100G SMF-SR	Future
QSFP 40G SR4	5.1.3
QSFP 40G LR4	5.1.3
QSFP 4x10G LR4 breakout	Post 5.3.1
QSFP 4x10G ER4	Post 5.3.1
QSFP 40G SMF-SR	5.3.x
QSFP 40G CSR4	Future



# CRS-X Hardware Details MSC-FP-LSP Cards

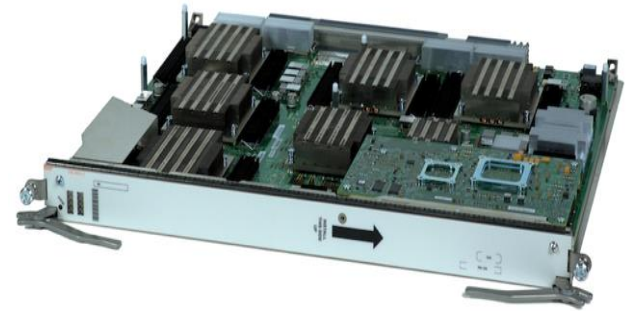
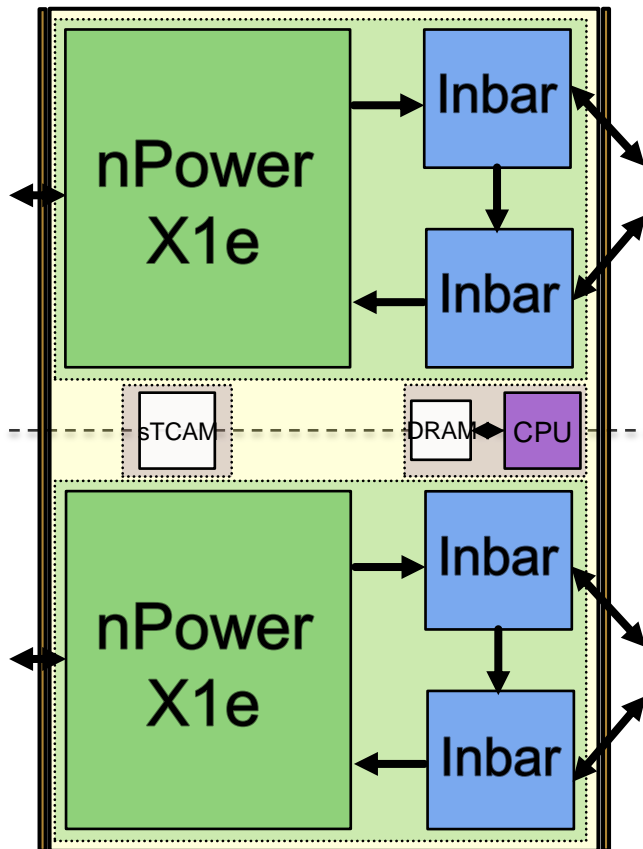




# MSC400/FP400/LSP400

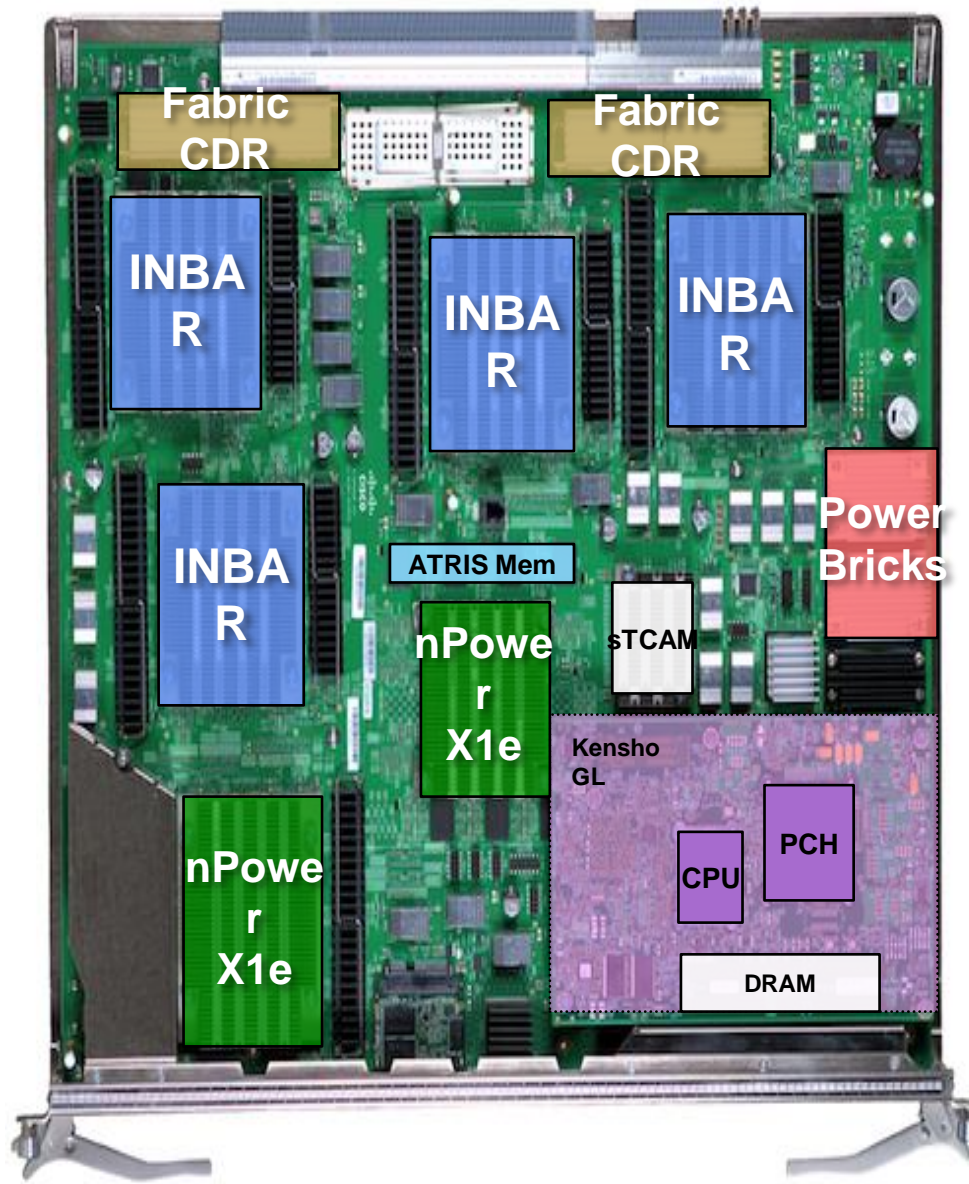
## Line Cards

- Paired with the CRS-X 400G PLIMs, they will offer similar roles to their Taiko and Metro counterparts with the addition of the MAC role.
- Contains nPower X1e, Inbar and Kensho daughterboard (CPU, memory, ...)



# MSC400/FP400/LSP400

## Details

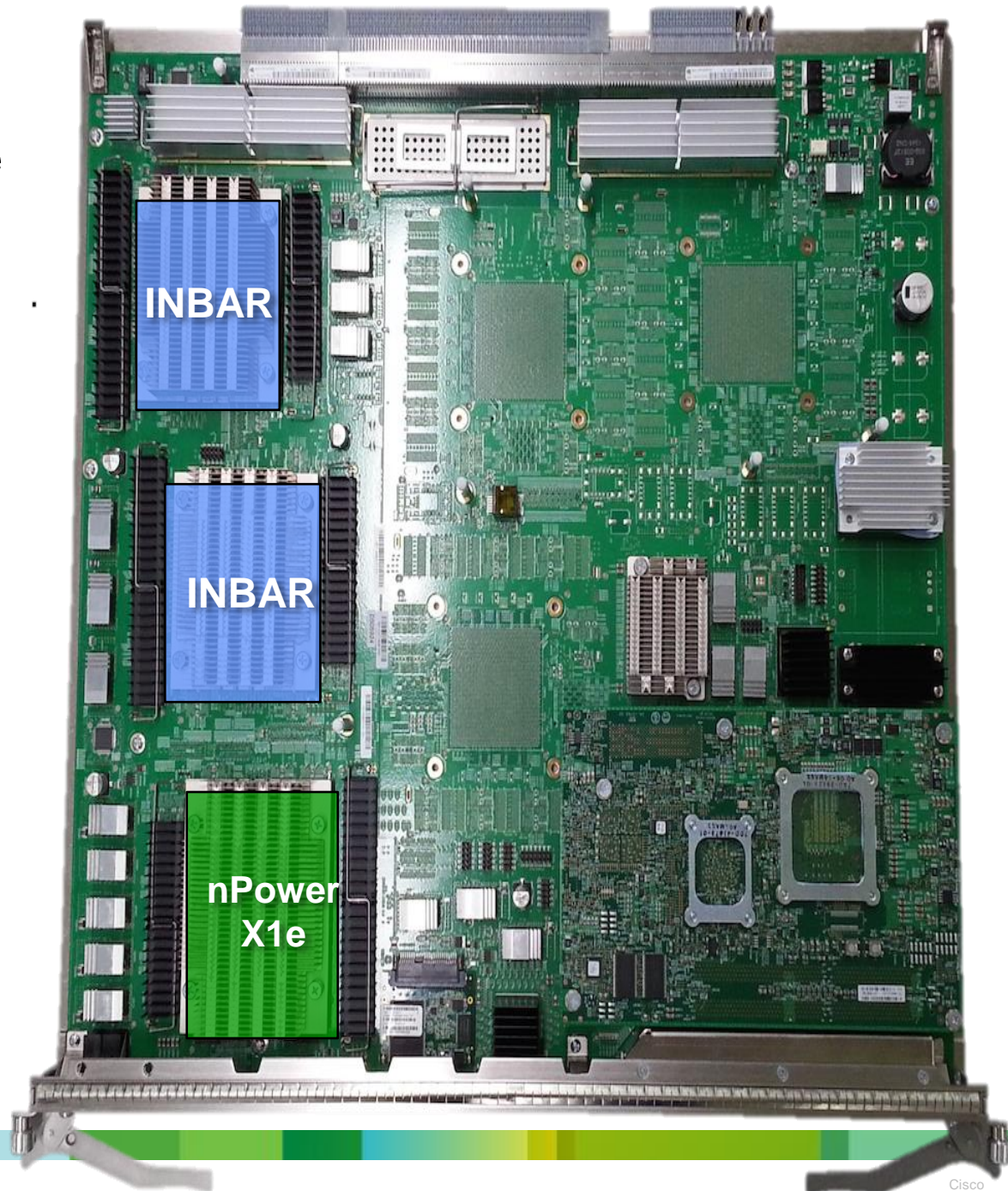




# FP200-Lite

Only in Legacy Chassis

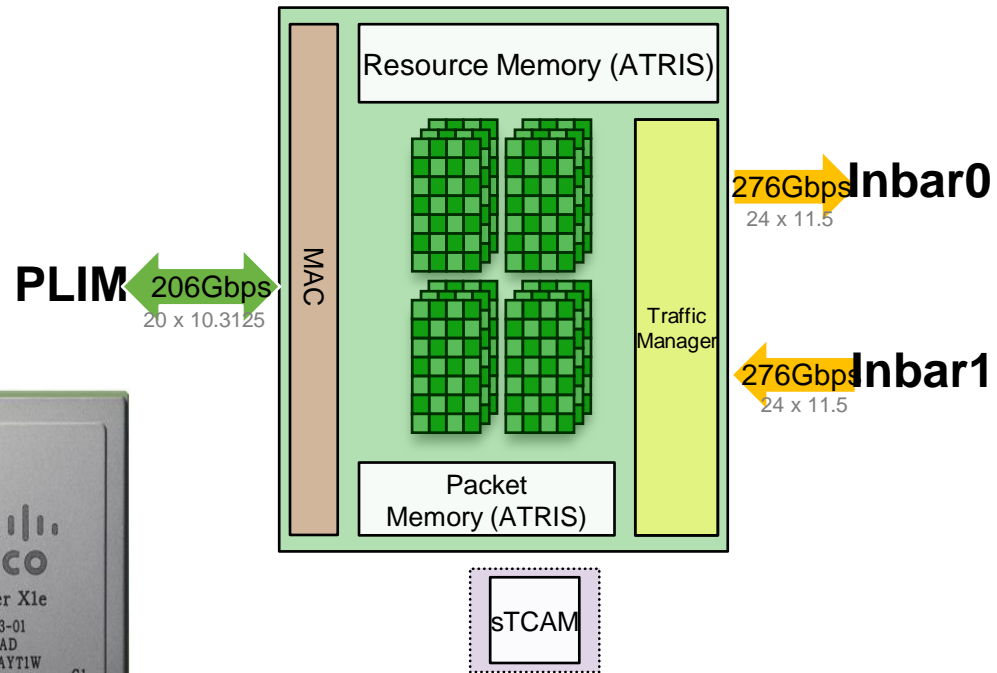
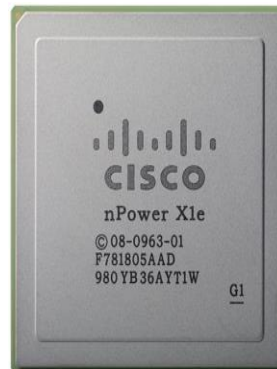
- Depop'd version of the FP400 and MSC-400 linecards
- One slice only
- TCO much lower than 400G linecard with 200G license



# nPower X1e (PAT) – Details

The industry's most advanced Network Processing Units

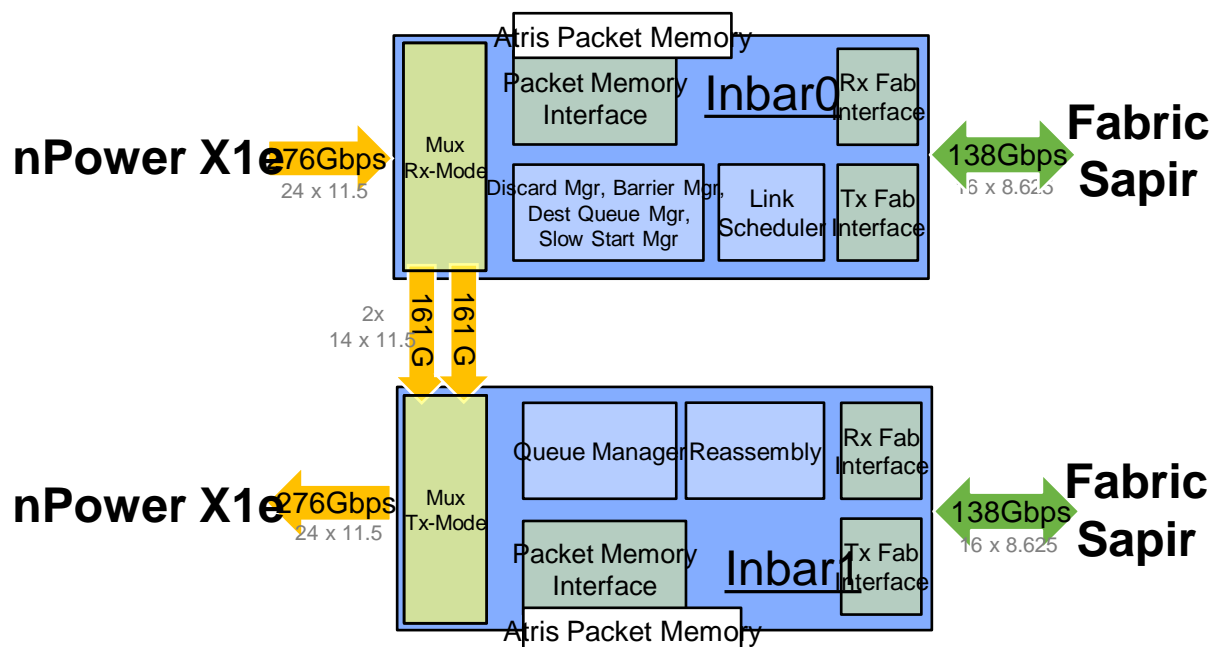
- TI / TSMC 40 nm ASIC evolved from nPower X1 (PITA) with edge features and increased scale
- 336 (21x16) micro-coded PPEs running at 800 MHz
  - 11 clusters are assigned to Ingress (125Mpps)
  - 10 clusters are assigned to Egress (117Mpps)
- Two threads per Packet Processing Engine
- 230 Mpps, 200 Gbps bandwidth
- Off-chip TCAM memory: 80Mb shared between 2 slices
- 9M IPv4/VPNv4 routes
- 4.5 IPv6/VPNv6 routes
- Enhanced Traffic Manager
  - 32K interfaces
  - 128K Queues
  - 5 Level QoS
- Integrated MAC logic
- Power: ~75W



# Inbar- Details

## Collapsing IngressQ and FabricQ in a Single ASIC

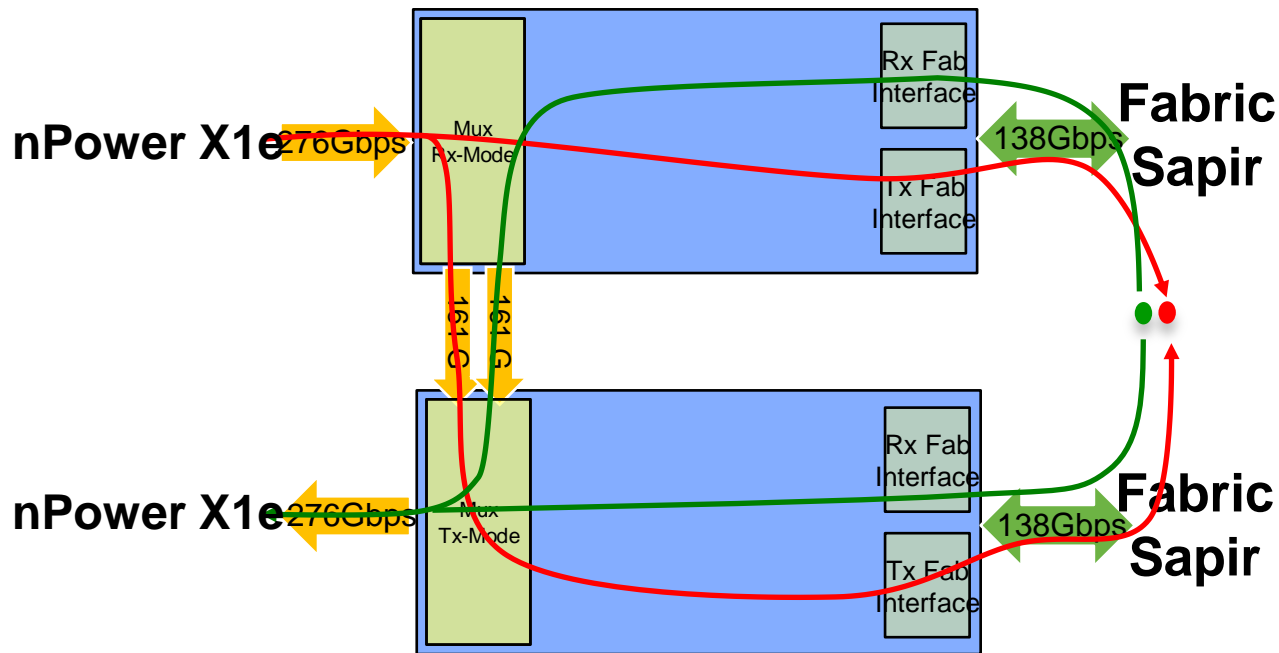
- TI / TSMC 40 nm ASIC
- Integrated Taiko Seal and Crab roles in each Inbar.
- Separate IngressQ, FabricQ via s/w drivers
- Support for
  - 400Gbps mode LC via 16x8.625Gbps Serdes
  - 200Gbps mode LC via 32x4.3125Gbps Serdes
- 26ms RTT buffer (IMIX)
- Power: 35W



# Inbar- Details

## Traffic Load Balancing to the Fabric

- Inbar will segment traffic in Cells before passing them to the fabric
- In Inbar we also have a Mux programmed as Rx or Tx. It will be used to split traffic based on ports.
- Performance (bidirectional)
  - 100Gbps
  - 125Mpps
  - 125Mcps
  - Extra 5Mcps headroom for CPU

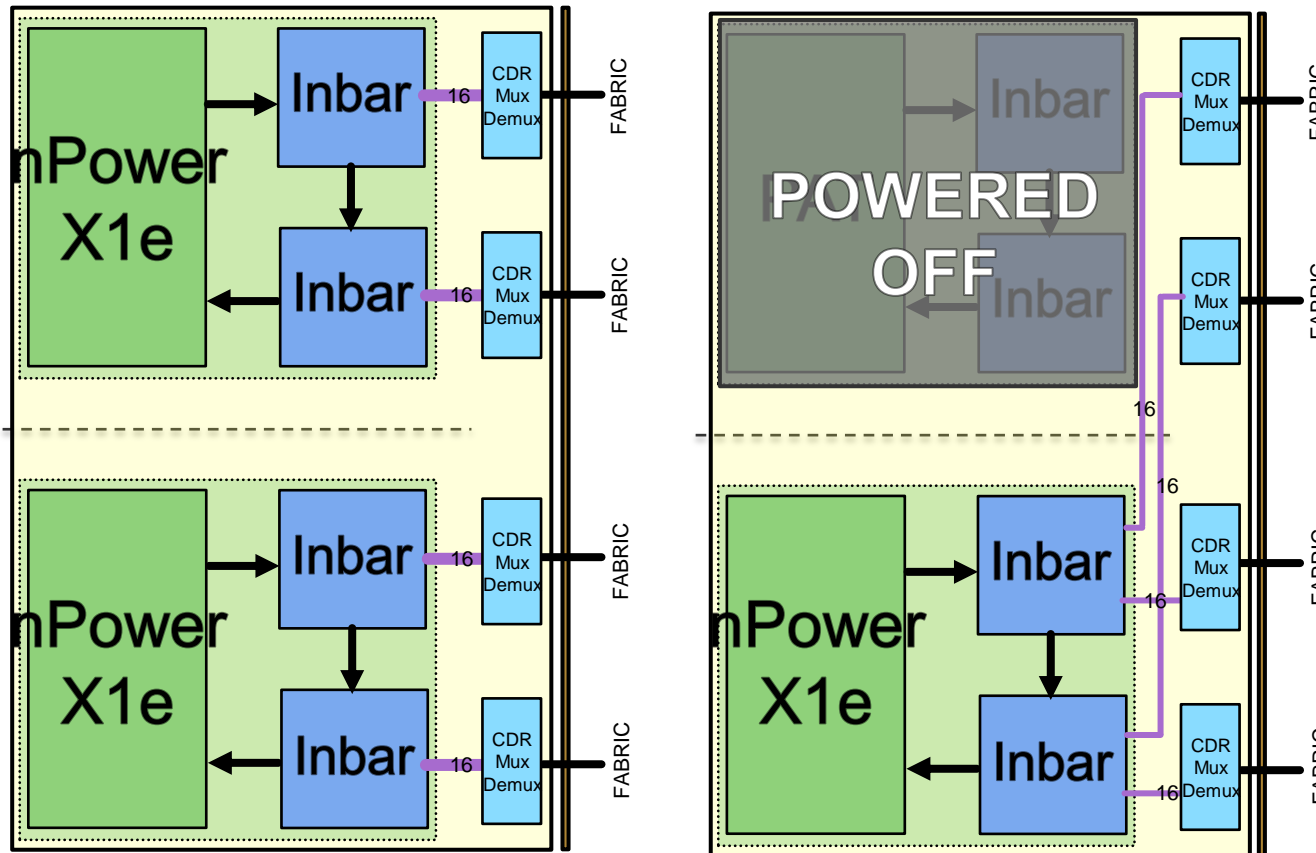


# Mode of Operation

Enhanced Chassis 400G / Legacy Chassis 200G

- 400G Mode: 2 active slices  
4 x 16 x 8.625 Gbps

- 200G Mode: 1 active slice  
4 x 16 x 4.3125 Gbps

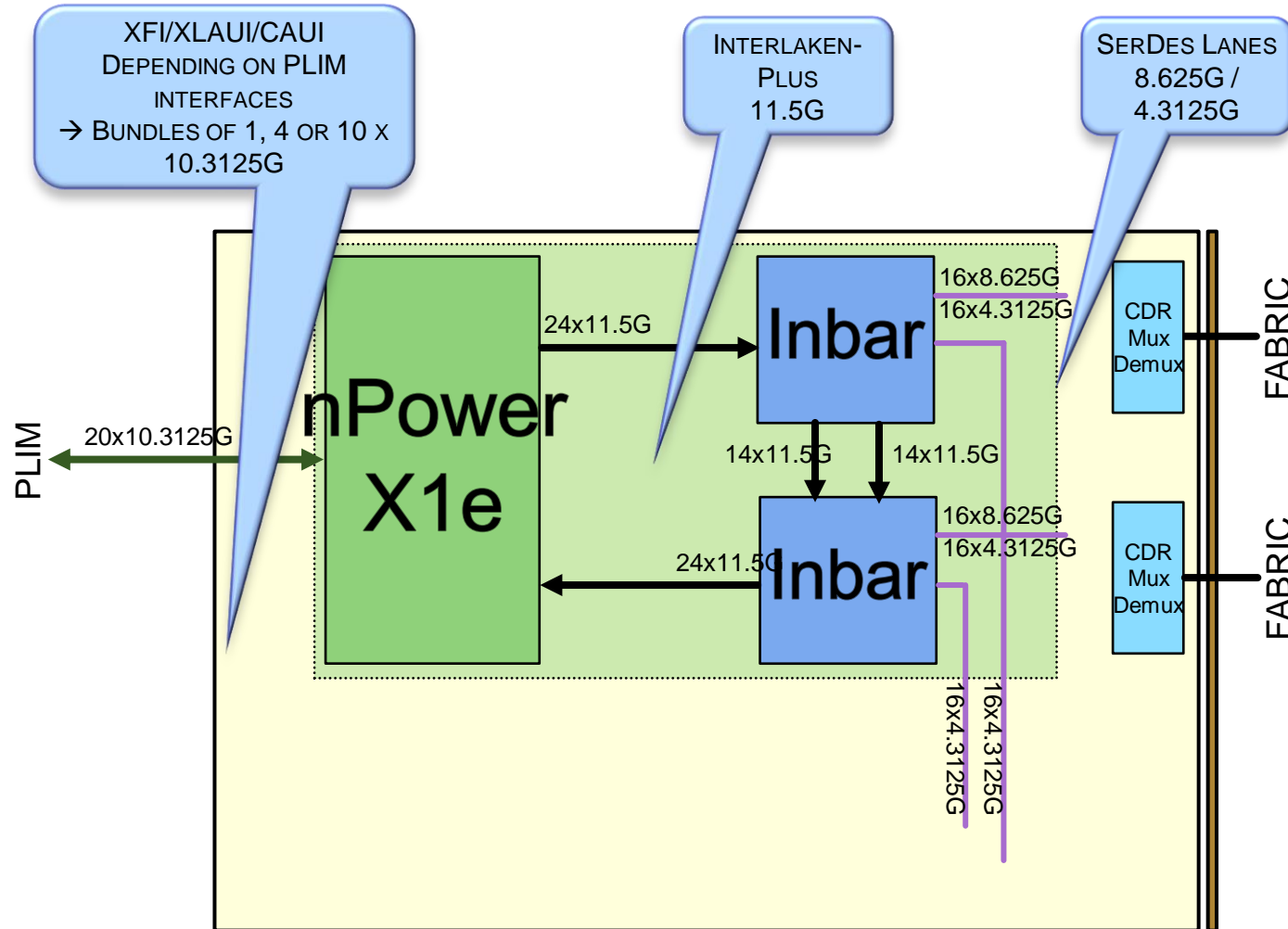


- CDR Mux are linear Equalizer and Retimer devices that include MUX functionality to support fabric connectivity in 200G mode

# ASIC Data Connection

Enhanced Chassis 400G / Legacy Chassis 200G

- ASICs are interconnected via different links and lanes

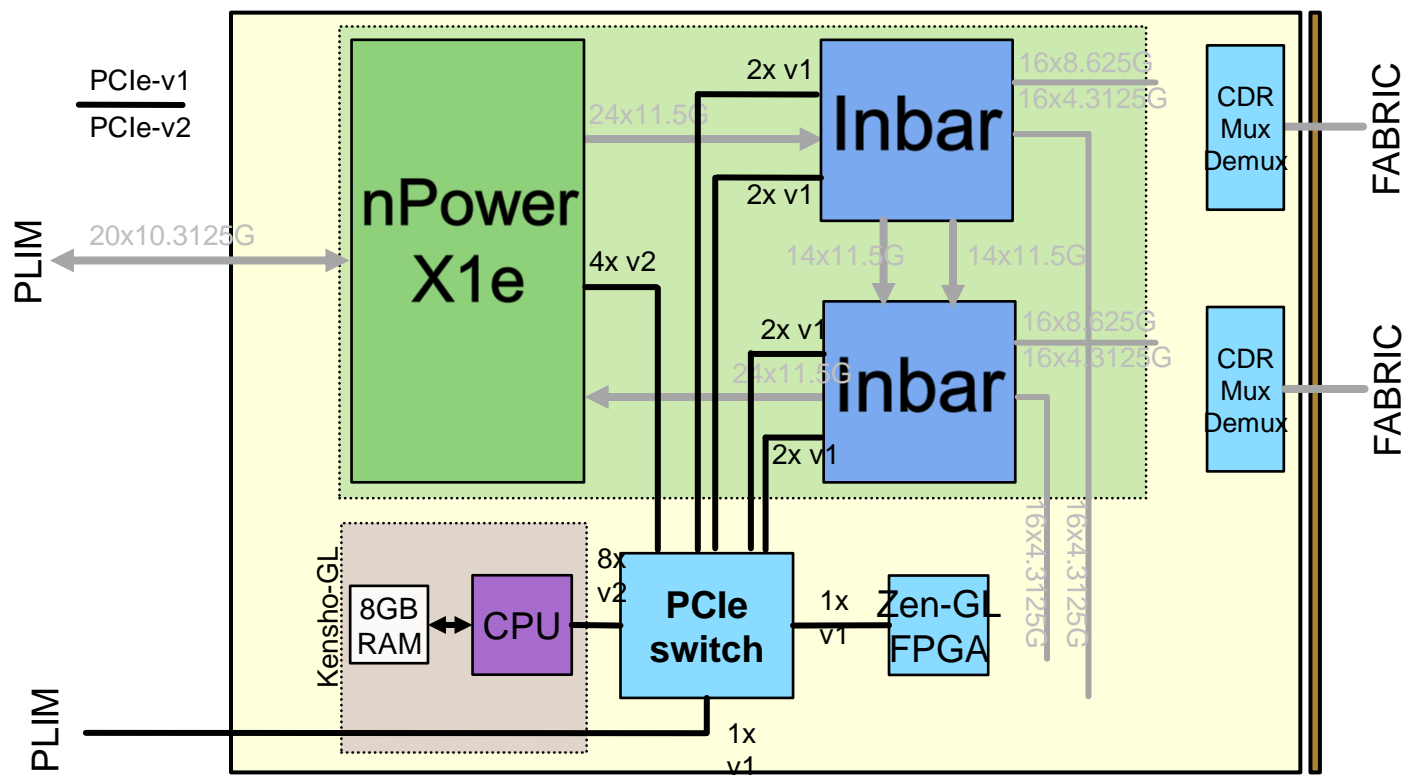




# ASIC Control Connection

Enhanced Chassis 400G / Legacy Chassis 200G

- a PCI-e v1/v2 switch connected to the various ASICs of the line card and to the PLIM too



- To Inbar, we have 2 lanes to Seal driver and 2 lanes to Crab driver of the ASIC



# CRS-X Hardware Details

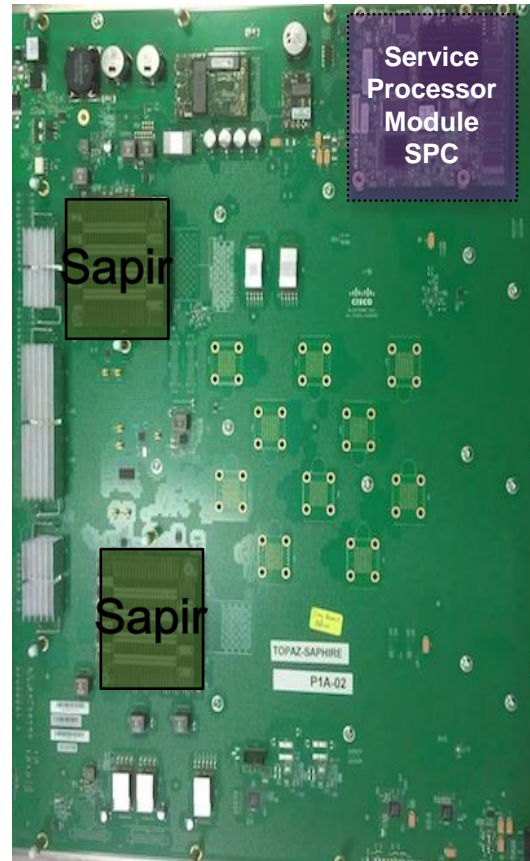
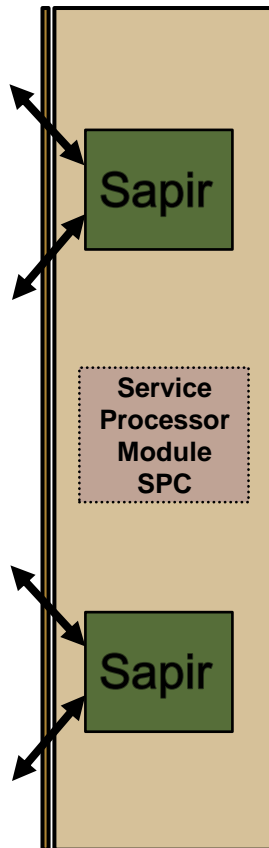
## Fabric Cards



# FC400

Fabric Cards

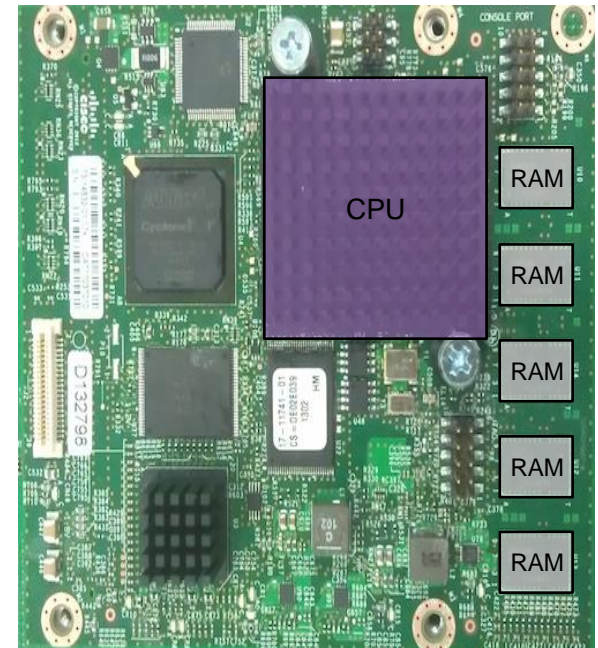
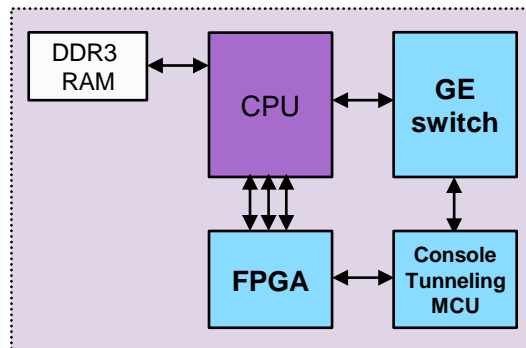
- At FCS, only for single chassis 8 or 16-slot
- At 5.1.3, for B2B and MC (same cards), 16-slot
- Contains Sapir ASIC and a new SP-C daughterboard (CP Memory, ...)



# FC400

## SP-C Daughter Cards

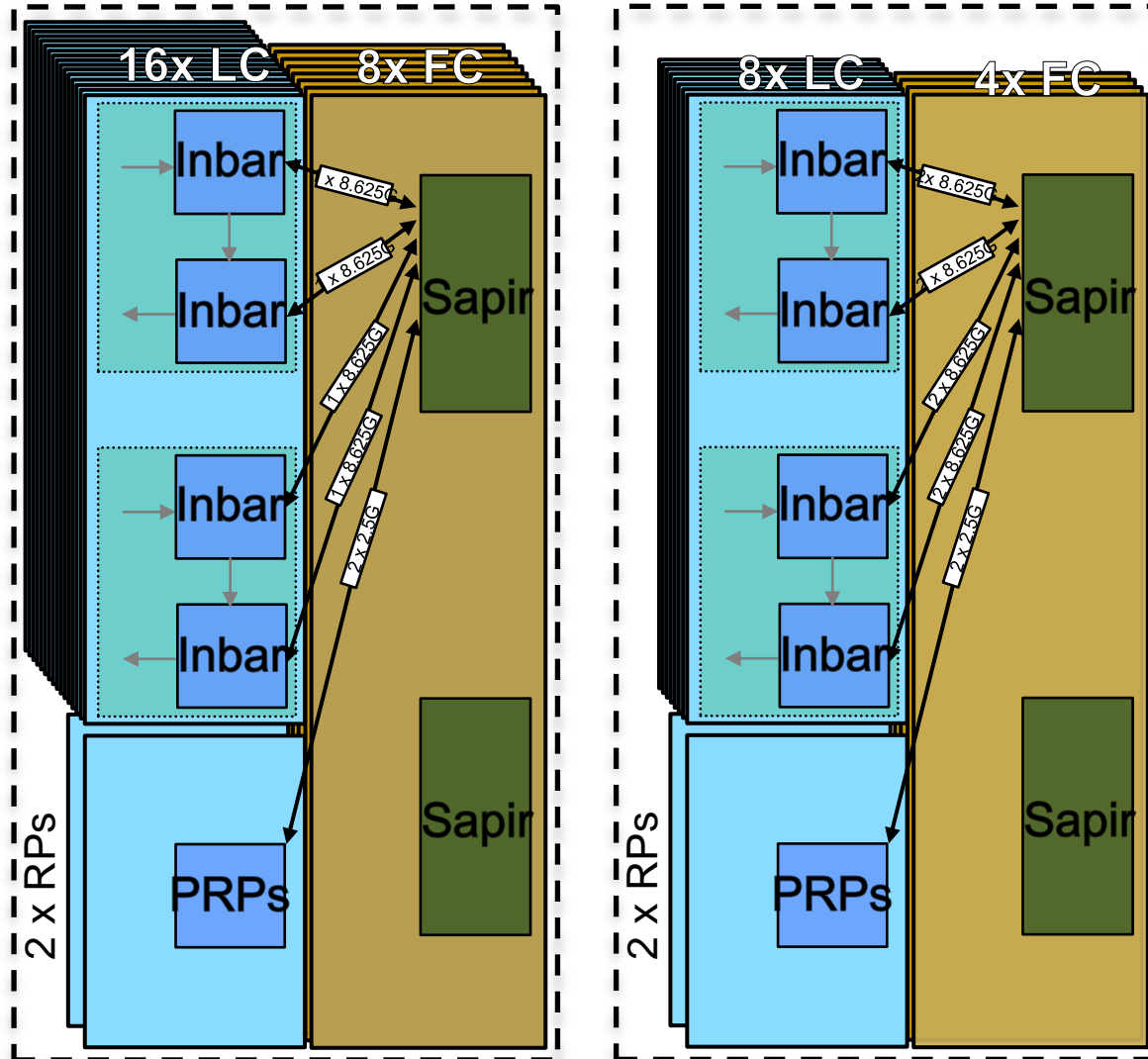
- SP-C role:
  - Run the local control plane software or the offline diagnostics
  - Power up and down the main card
  - Perform environmental management tasks such as voltage monitoring, voltage margining and temperature monitoring
  - Initialize and manage the Sapir ASICs, the FGID tables
- Next generation of SPB
- Using Freescale P1013 processor
- DDR3 memory
- GbE switch replaces FE switch on SPB



# Inbar to Sapir in Enhanced Chassis

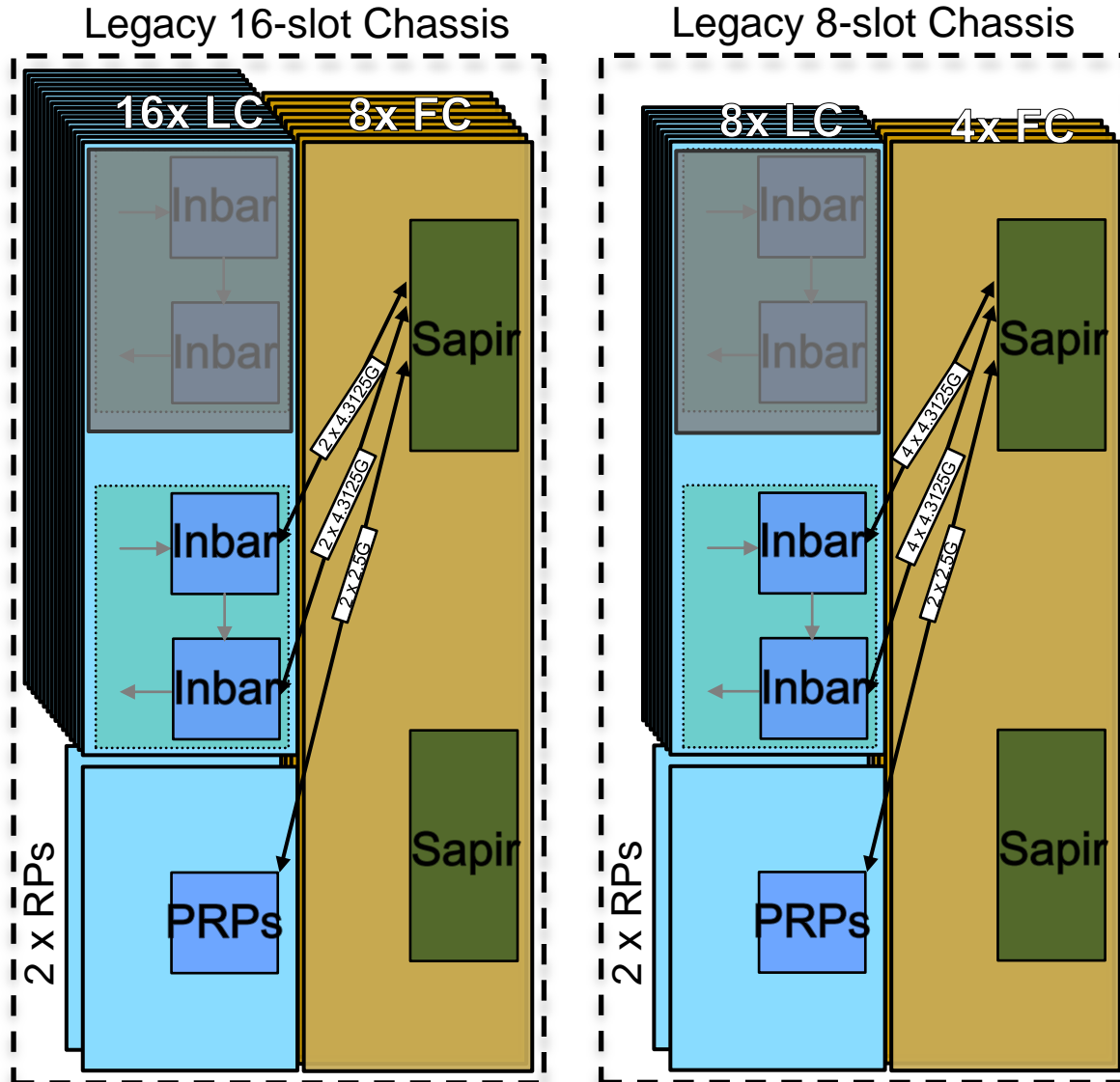
68 Rx/Tx links from each Sapir

Green Hornet / 16-slot Chassis Yellow Jacket / 8-slot Chassis



# Inbar to Sapir in Legacy Chassis

68 Rx/Tx links from each Sapir

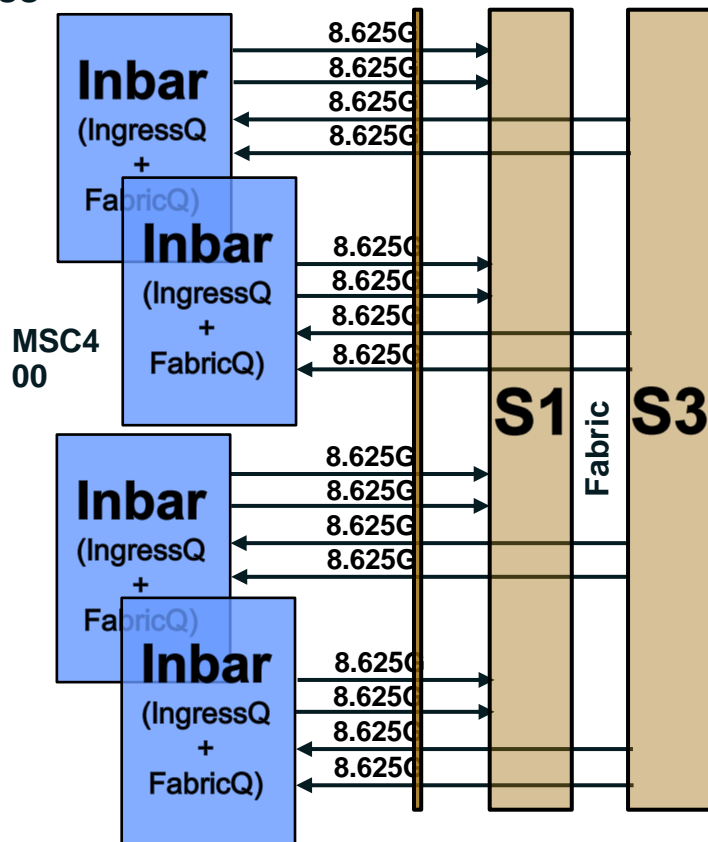


# Fabric Bandwidth Comparison

CRS-X vs CRS-3 (Taiko)

## CRS-X

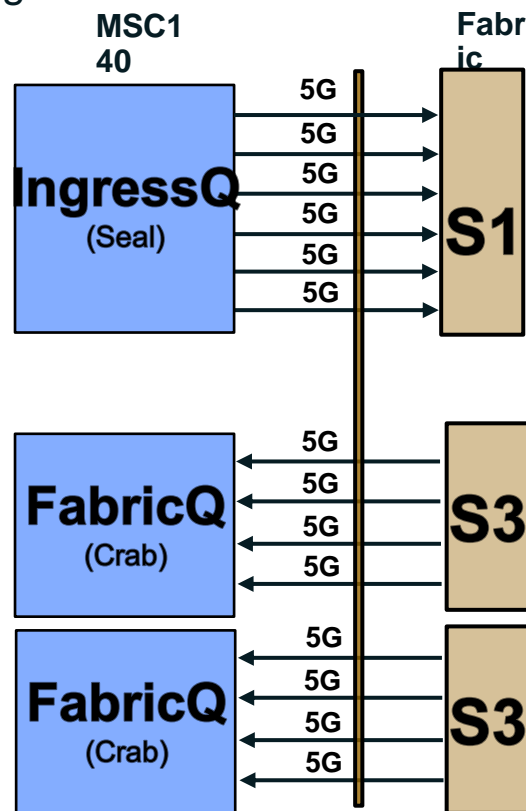
$8.625\text{G} \times 8 = 69\text{Gbps}$  ingress (per plane)  
 $69\text{G} \times 8 \text{ planes} \times 60 / 64 = \underline{517.5\text{Gbps}}$  total ingress



$8.625\text{G} * 8 = 69\text{Gbps}$  egress (per plane)  
 $69\text{G} * 8 \text{ planes} = \underline{552\text{Gbps}}$  total egress

## CRS-3

$5\text{G} \times 6 = 30\text{Gbps}$  ingress (per plane)  
 $30\text{G} \times 8 \text{ planes} \times 5 / 6 = \underline{200\text{Gbps}}$  total ingress



$5\text{G} * 8 = 40\text{Gbps}$  egress (per plane)  
 $40\text{G} * 8 \text{ planes} = \underline{320\text{Gbps}}$  total egress

# Fabric Bandwidth Comparison

CRS-X vs CRS-3 (Taiko)

## CRS-3

Doesn't use CRC32 but rely on encoding 8B/10B  
Header + FEC + encoding → 71% utilization

## CRS-X

Uses 4 extra bytes for CRC32 every two cells and no longer uses encoding 8B/10B  
Header + FEC + CRC32 → 87% utilization

Fabric Link	Raw BW (Gbps)	Cell Payload	Cell Header	FEC	CRC32	Cell Data	Cell total with 8B/10B	Utilization
CRS-1	2.5	120B	12B	4B	0B	136B	170B	71%
CRS-3	5							
CRS-X 4.3125	4.3125				2B	138B	138B	87%
CRS-X 8.625	8.625							



# Fabric Bandwidth Comparison

## CRS-X vs CRS-3 (Taiko)

### CRS-X

Inbar(IngressQ) → S1 // S3 → Inbar (FabricQ)

Ingress 517.5 (raw bw) (scrambling → no encoding tax)  
x 120 / 138 (cell overhead) = **450Gbps per LC slot**

Egress 552 (raw bw) (scrambling → no encoding tax)  
x 120 / 138 (cell overhead) = **480Gbps per LC slot**

Impact of losing 1 plane →  $480 \times 7 / 8 = 420\text{Gbps per slot}$

Impact of losing 2 planes →  $480 \times 6 / 8 = 360\text{Gbps per slot}$

### CRS-3

IngressQ → S1

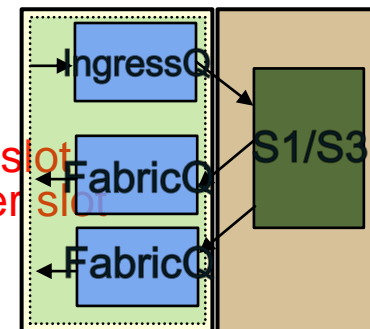
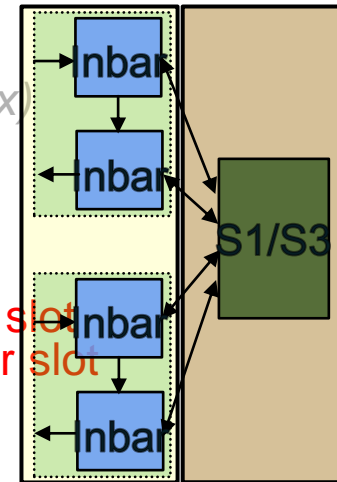
200Gbps (raw bw) x 8b/10b (encoding)  
x 120/136 (cell overhead) = **141Gbps per LC slot**

S3 → FabricQ

320Gbps (raw bw) x 8b/10b (encoding)  
x 120/136 (cell overhead) = **225Gbps per LC slot**  
(or 113Gbps per FabricQ)

Impact of losing 1 plane →  $141 \times 7 / 8 = 123\text{Gbps per slot}$

Impact of losing 2 planes →  $141 \times 6 / 8 = 105\text{Gbps per slot}$



# Minimum Links Needed to Keep Fabric Plane UP

	Taiko S2		Topaz S2	
<b>Metro S13</b>	S1→S2	1 of 36	S1→S2	1 of 36
	S2→S3	49 of 72	S2→S3	37 of 54
<b>Taiko S13</b>	S1→S2	1 of 128	S1→S2	1 of 36
	S2→S3	49 of 256	S2→S3	36 of 54
<b>Topaz S13</b>	N/A	N/A	S1→S2	1 of 54
	N/A	N/A	S2→S3	37 of 54

# Verification

## show commands

- Let try these show commands to understand the traffic received on this interface.
- On both SUT and Router1:

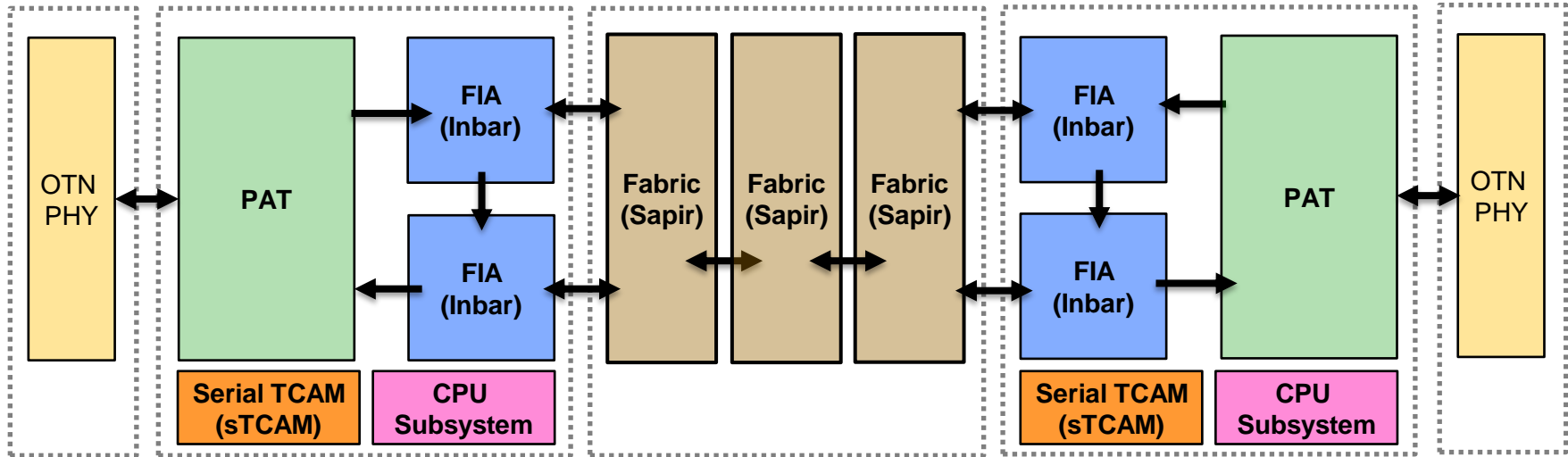
```
(admin)#show contr fabric plane all statistics det  
(admin)#show controllers fabric link port fabricqrx all  
statistics brief
```

# Life of a packet

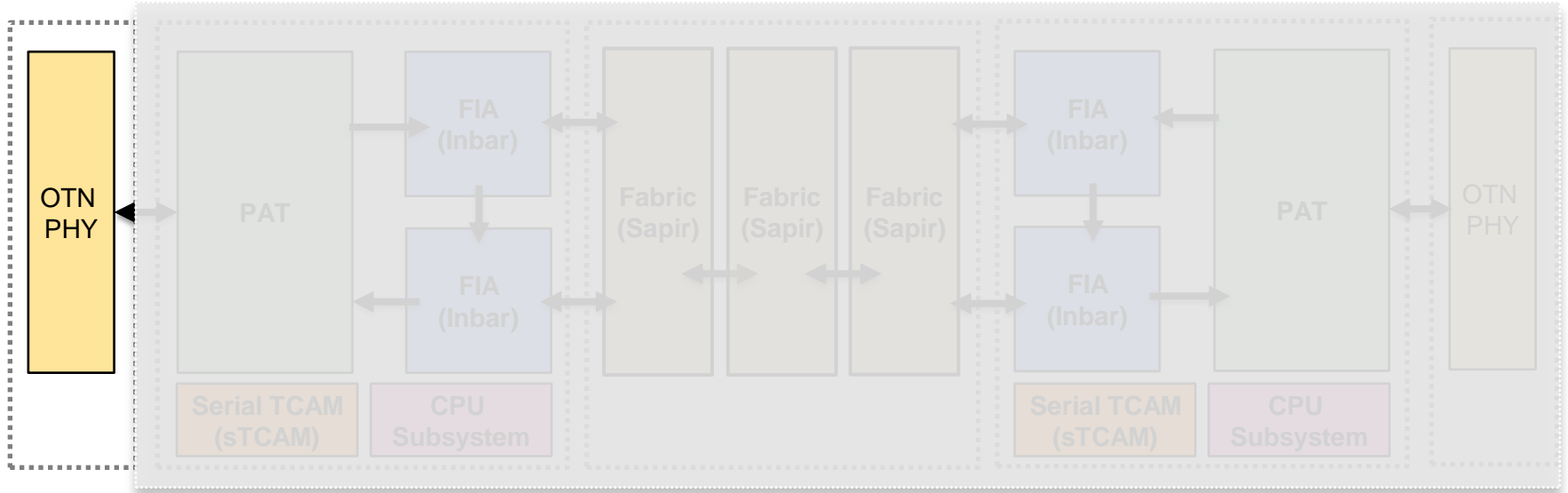


# CRS-X Life of a Packet

- Major packet forwarding components remains unchanged with respect to CRS-3



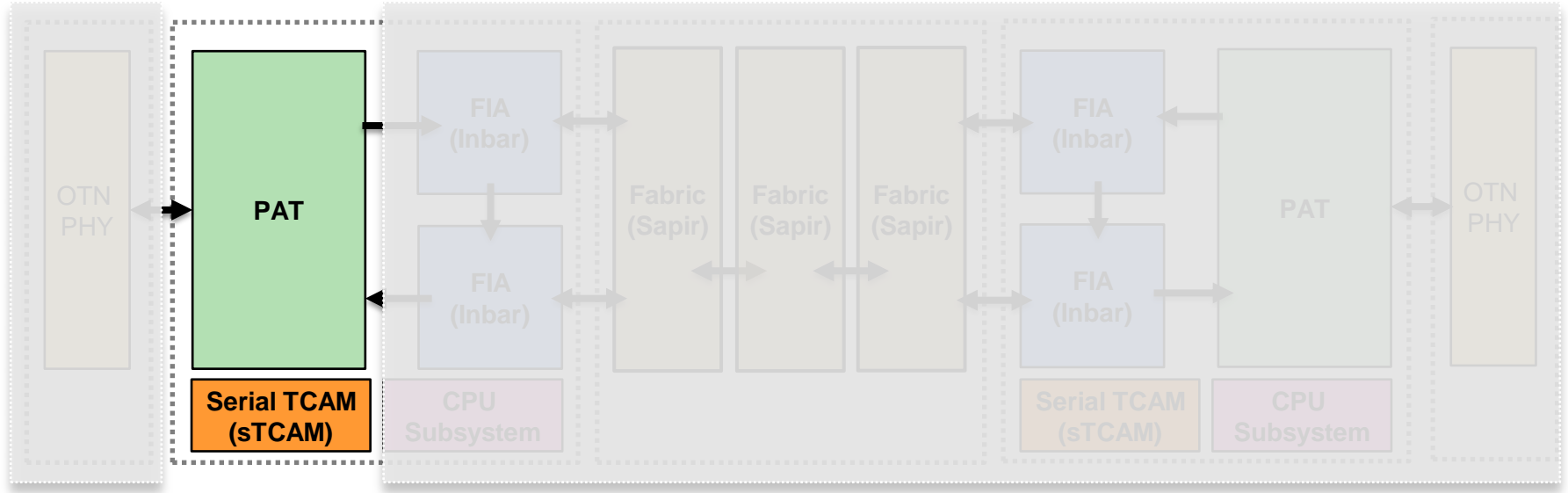
# CRS-X Life of a Packet



- The packets enter the linecards via Pluggable optics (which is not shown in the diagram).
- If G.709 encapsulation is used, OTN framer calculates the FEC and also performs de-capsulation of G.709 framing

```
show controllers hun0/2/0/1 phy
show controllers hun0/2/0/1 int
show controllers hun0/2/0/1 stats
show controllers plim asic statistics interface
hundredGigE 0/2/0/1
show controllers plim asic ether queues interface
HundredGigE0/2/0/0
```

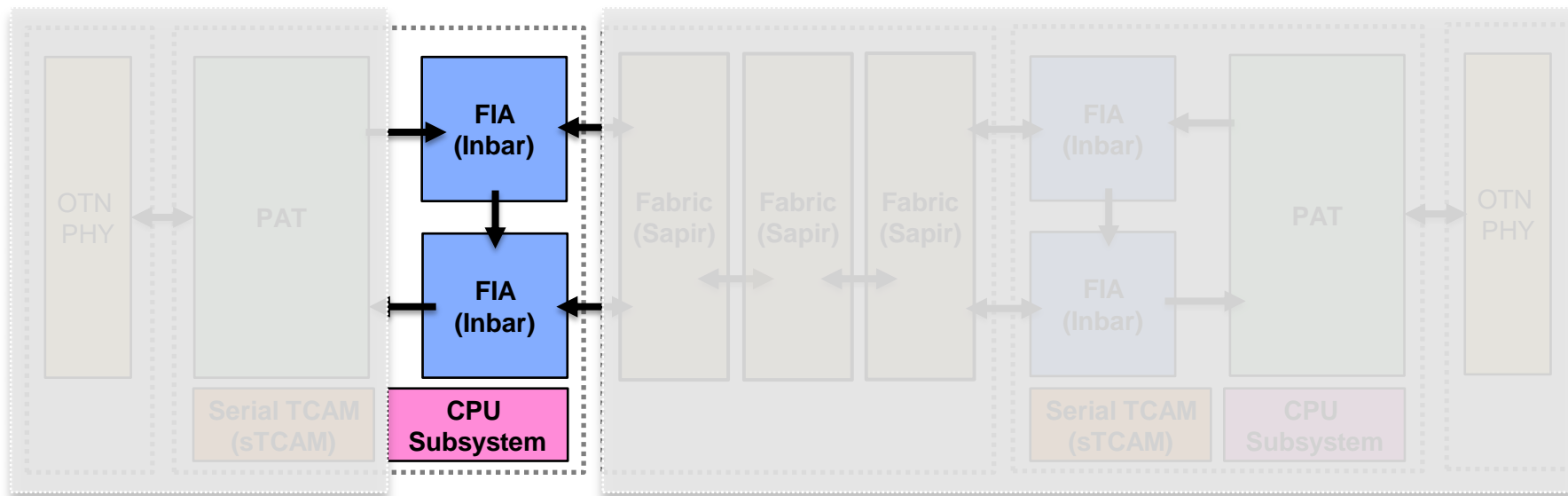
# CRS-X Life of a Packet



- The ingress packet is then passed along to PAT which
  - performs MAC processing
  - applies ingress features (QoS policing, ACL, NFv9, accounting...)
  - performs ingress lookup (IPv4/IPv6/MPLS)

```
sh controllers pse pat statistics summary instance 0
location 0/2/CPU0
sh controllers pse pat summary instance 0 location
0/2/CPU0
```

# CRS-X Life of a Packet

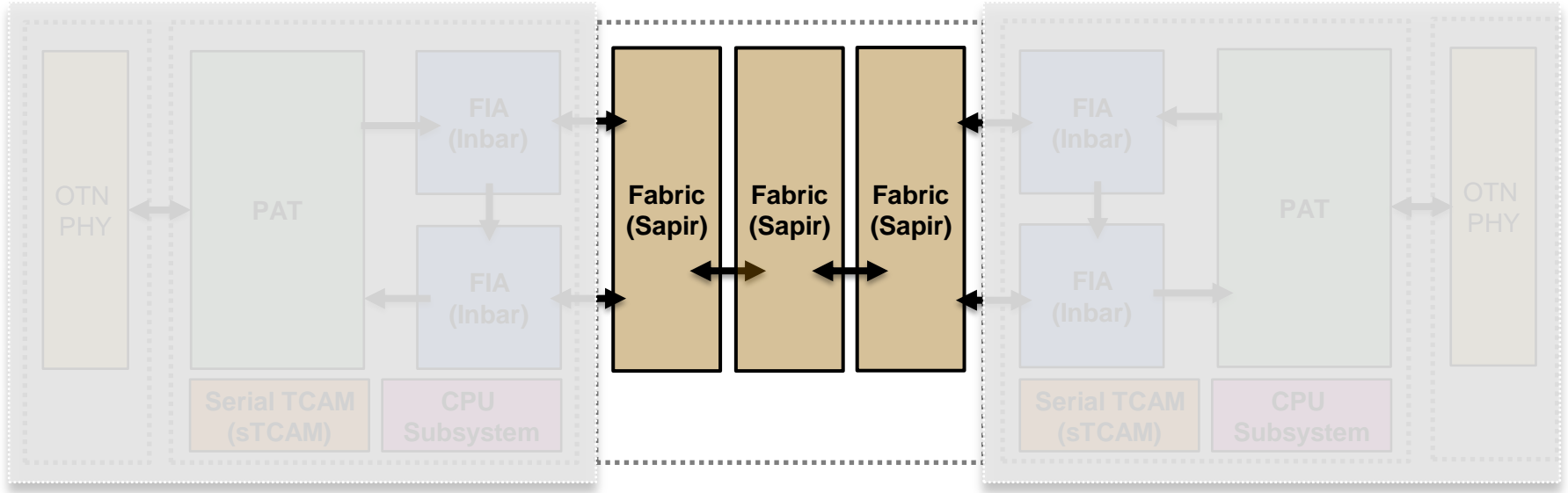


- The Inbar FIA provides for fabric QoS and cell segmentation
- It is here that traffic is segmented from packets into cells and queued to the various fabric destinations and sent towards the fabric

```
sh controllers ingressq statistics location 0/2/CPU0  
sh controllers ingressq fabric links loc 0/2/CPU0
```



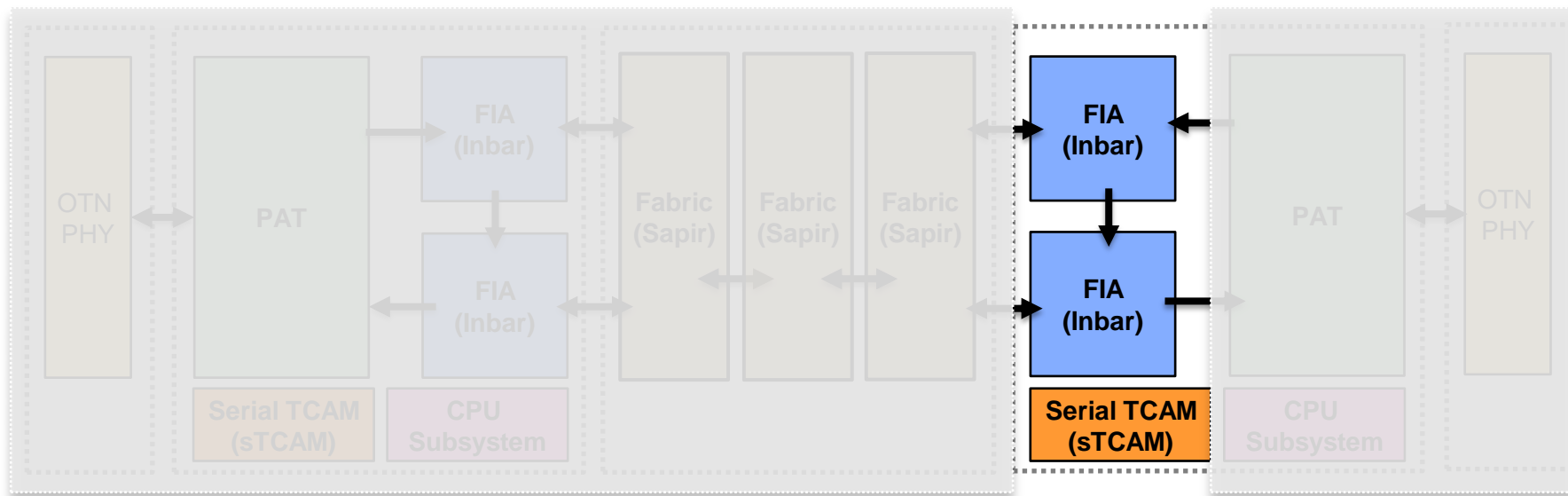
# CRS-X Life of a Packet



- The Sapir ASIC forms the fabric interfaces and performs cell switching as well as multicast replication
- In a single chassis, the same ASIC will act as S1 or S3 for the FIA
- In a multi chassis, FCC will contain only S2 cards

```
sh controllers fabricq statistics loc 0/2/CPU0
sh controllers fabricq stat detail loc 0/2/CPU0
sh controllers fabricq link-info all loc 0/2/CPU0
```

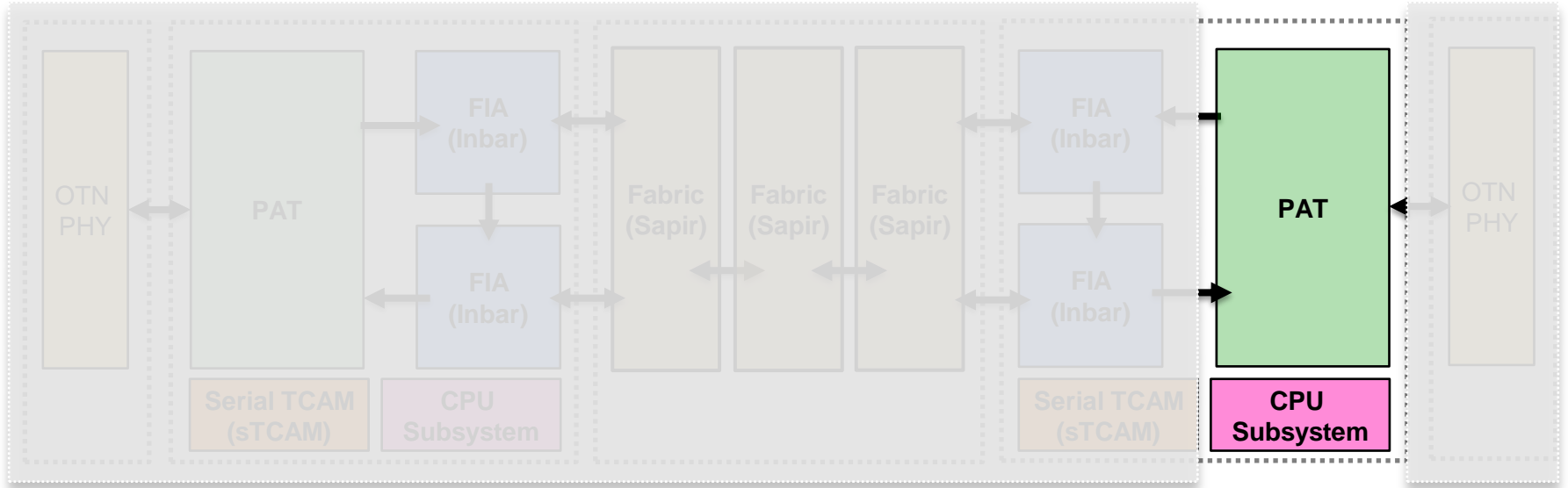
# CRS-X Life of a Packet



- In the egress path, the cells are reassembled by the Inbar in the egress path

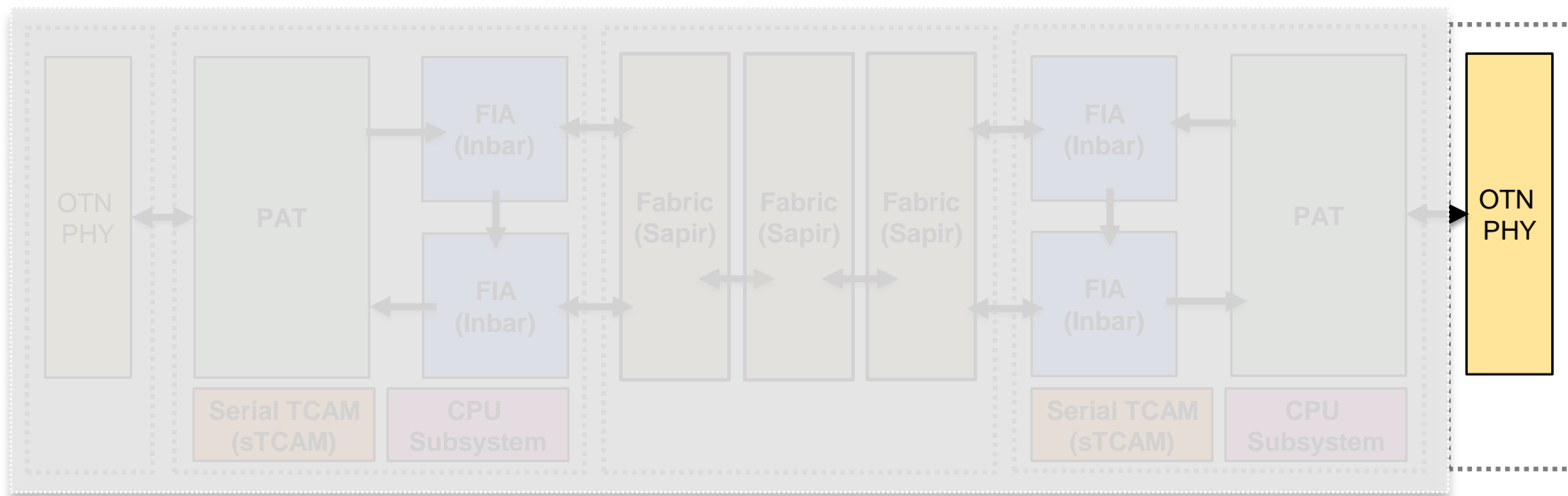
```
sh controllers fabricq statistics loc 0/2/CPU0
sh controllers fabricq stat detail loc 0/2/CPU0
sh controllers fabricq link-info all loc 0/2/CPU0
```

# CRS-X Life of a Packet



- The PAT NPU provides for egress MAC processing as well as lookup (IPv4/IPv6/MPLS) alongside with feature processing
- If egress queuing is configured, it will be processed within the PAT NPU's traffic manager block prior to being sent to the OTN PHY which handles OTN framing and FEC calculation

# CRS-X Life of a Packet



- If G.709 encapsulation is used, OTN framer calculates the FEC and also performs encapsulation of G.709 framing
- Finally our packet leaves the line card through the use of pluggable optics

# CRS-X MultiChassis migration

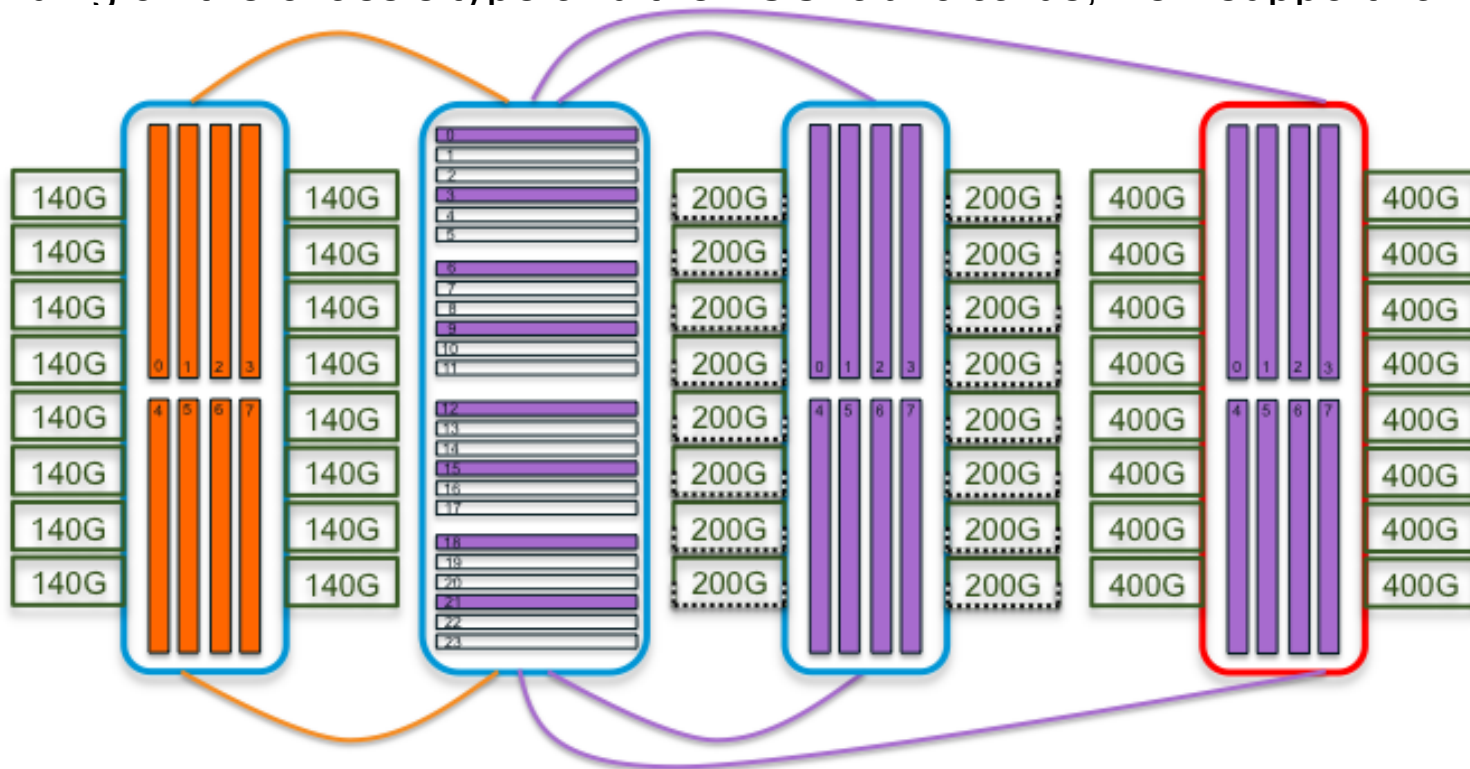


# CRS-X Program

## MC Support

BW/slot	Chassis	Fabric
40G	Legacy	40G
140G		
200G		
400G		
40G	Enhanced	400G
140G		
200G		
400G		

- Multichassis up to 4+1 support will be introduced in 5.1.3, 8+2 support coming in 5.1.4
- MultiChassis: S2 cards must be 400G, but S13 cards can be CRS-1 (40G), CRS-3 (140G) or CRS-X (400G)
- Depending on the chassis type and the LCC fabric cards, we'll support various mode



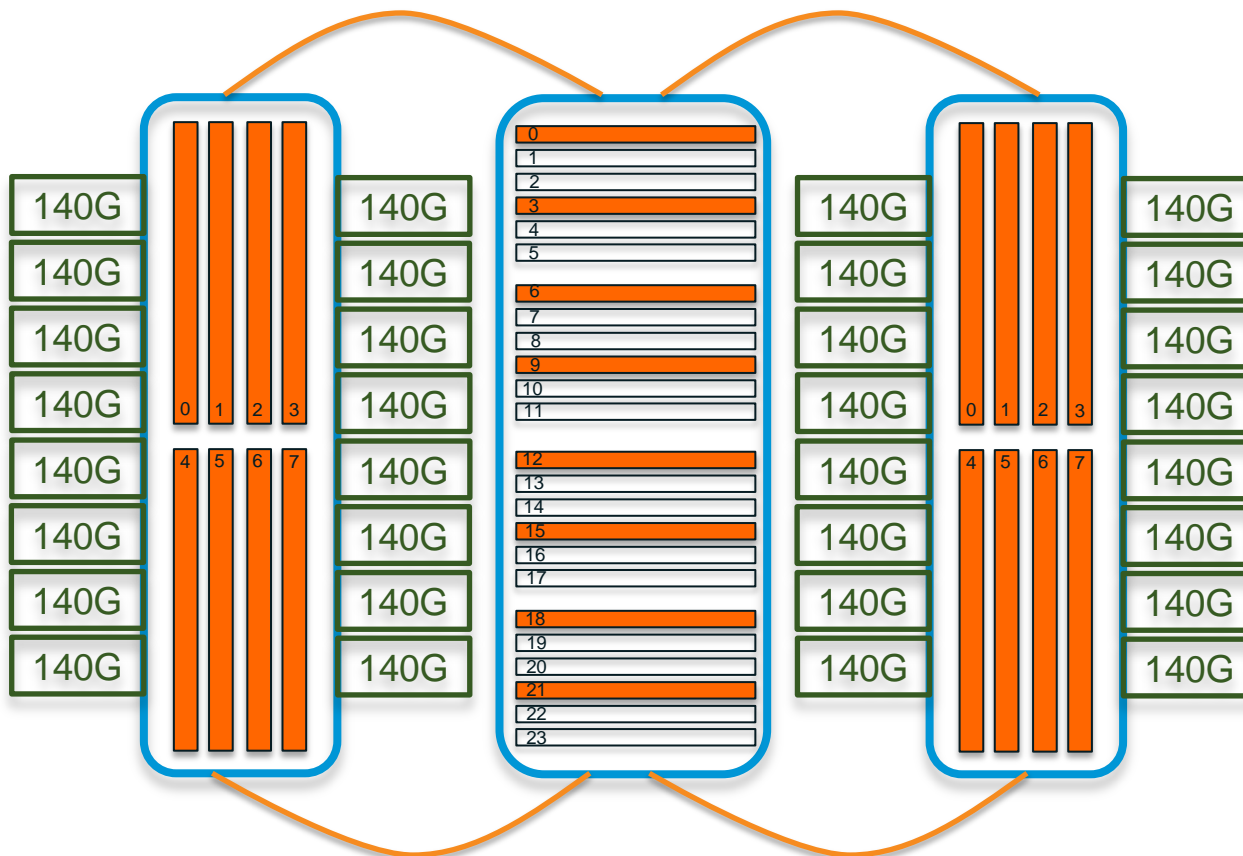
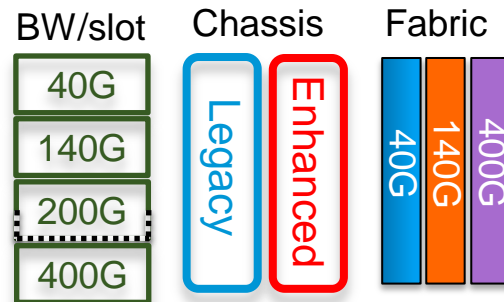
Interchassis fibers are the same for 40G/140G/400G but different between B2

# Migration Path to CRS-X

Upgrading Legacy MultiChassis to 400G (1/3)

Legacy LCC in MultiChassis 2+1: customers who can't go through the complex process of replacing chassis...

Note: we don't plan to support 8-slot in MultiChassis.

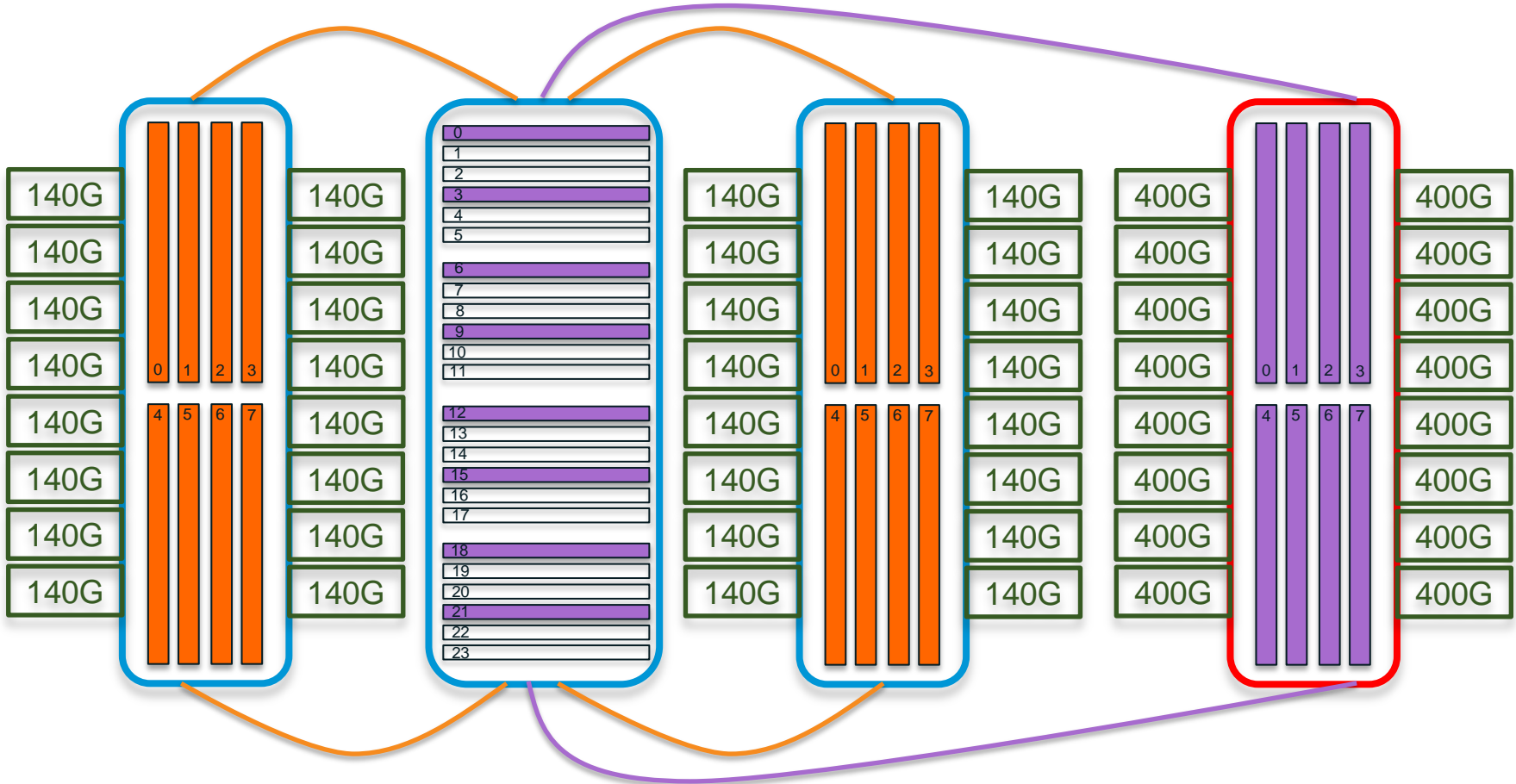


# Migration Path to CRS-X

Upgrading Legacy MultiChassis to 400G (2/3)

Legacy LCC in MultiChassis 2+1: customers who can't go through the complex process of replacing chassis, can still consider growth by upgrading FCC and adding a CRS-X LCC2

BW/slot	Chassis	Fabric
40G	Legacy	40G
140G		
200G		
400G		
	Enhanced	400G
		140G
		40G



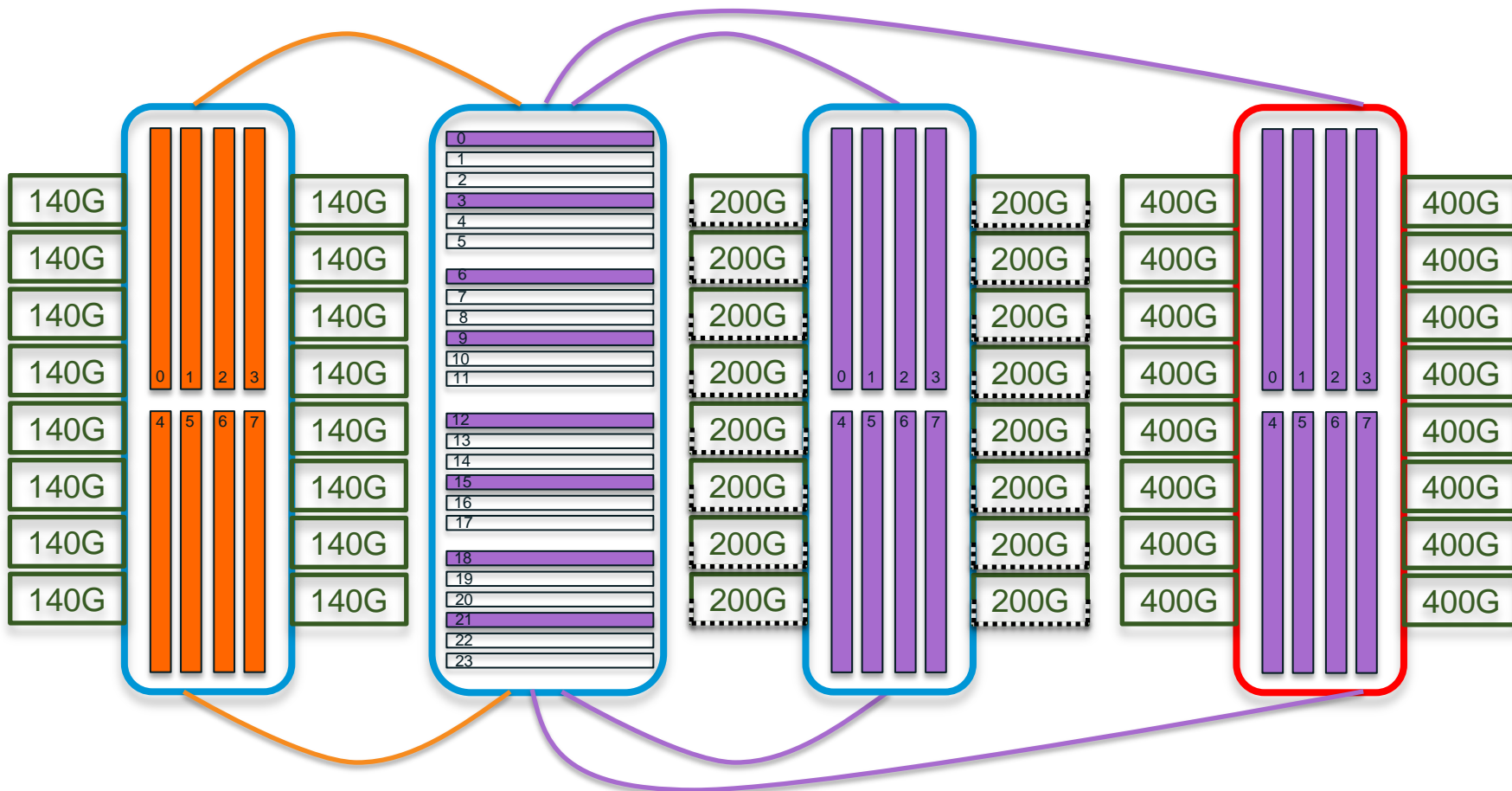


# Migration Path to CRS-X

Upgrading Legacy MultiChassis to 400G (3/3)

...can still consider growth by upgrading FCC and adding a CRS-X LCC2, or additionally migration some LLC to the CRS-X Hybrid mode:

BW/slot	Chassis	Fabric
40G	Legacy	40G
140G		
200G		
400G		
40G	Enhanced	40G
140G		
200G		
400G		

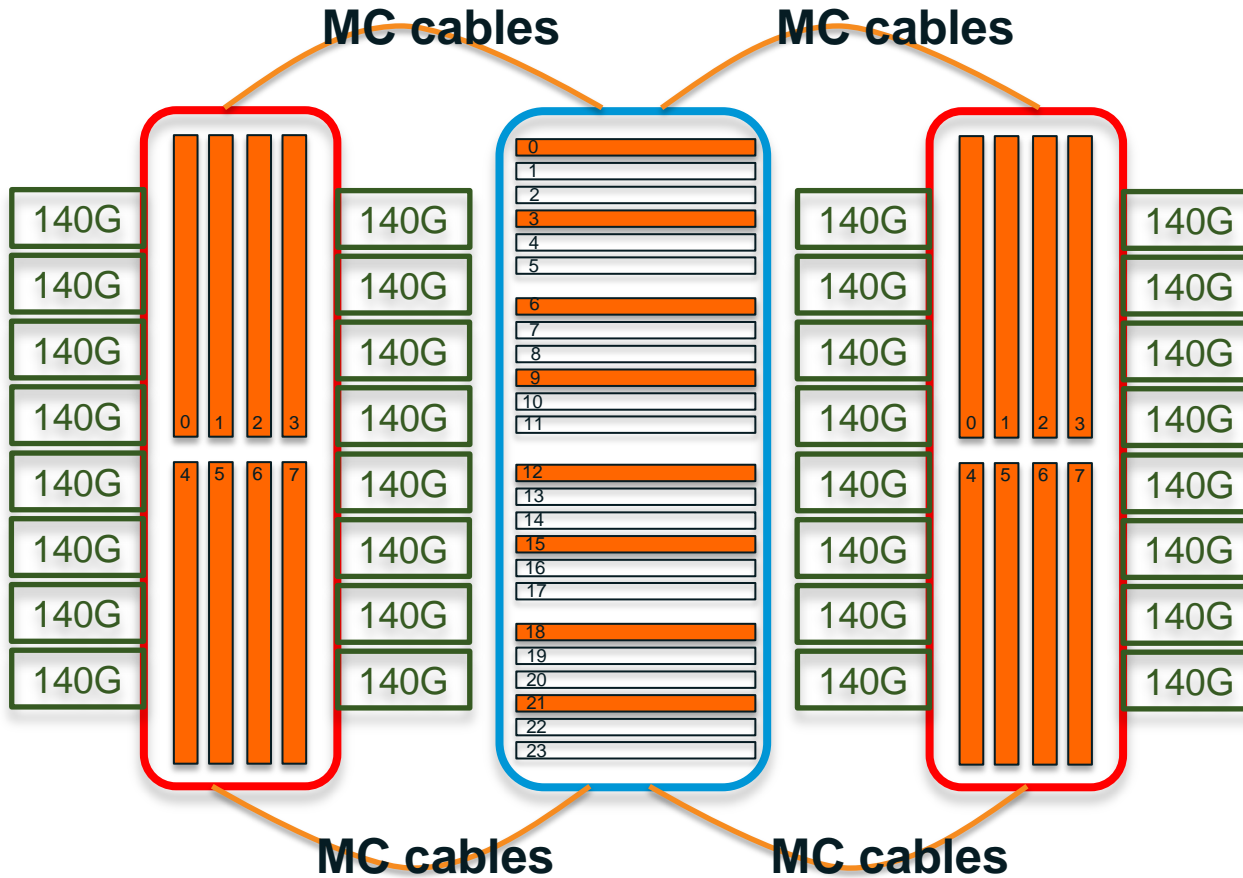


# Migration Path to CRS-X

Migrating MC CRS-3 to B2B CRS-X (1/3)

Some customers asked to migrate their MC CRS-3 2+1 or 3+1 to B2B CRS-X.

BW/slot	Chassis	Fabric
40G	Legacy	40G
140G		
200G		
400G		
40G	Enhanced	40G
140G		140G
200G		400G
400G		

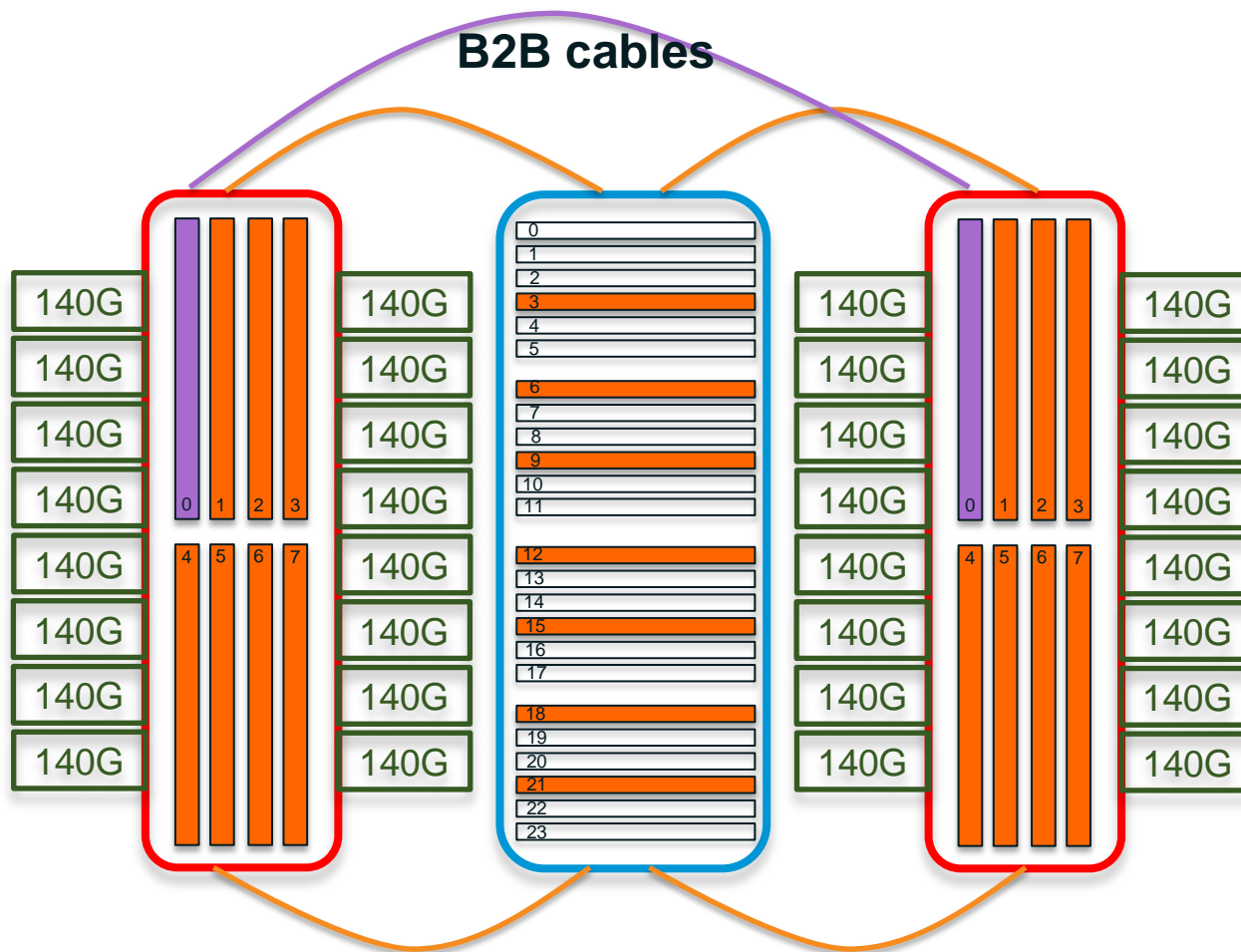


# Migration Path to CRS-X

## Migrating MC CRS-3 to B2B CRS-X (2/3)

Migration of Plane 0, disconnecting bundles to FCC SM0, swapping FC140/M with FC400/M and interconnecting them with B2B bundles

BW/slot	Chassis	Fabric
40G	Legacy	40G
140G		
200G		
400G		
40G	Enhanced	40G
140G		140G
200G		400G
400G		



Repeat operation plane by plane

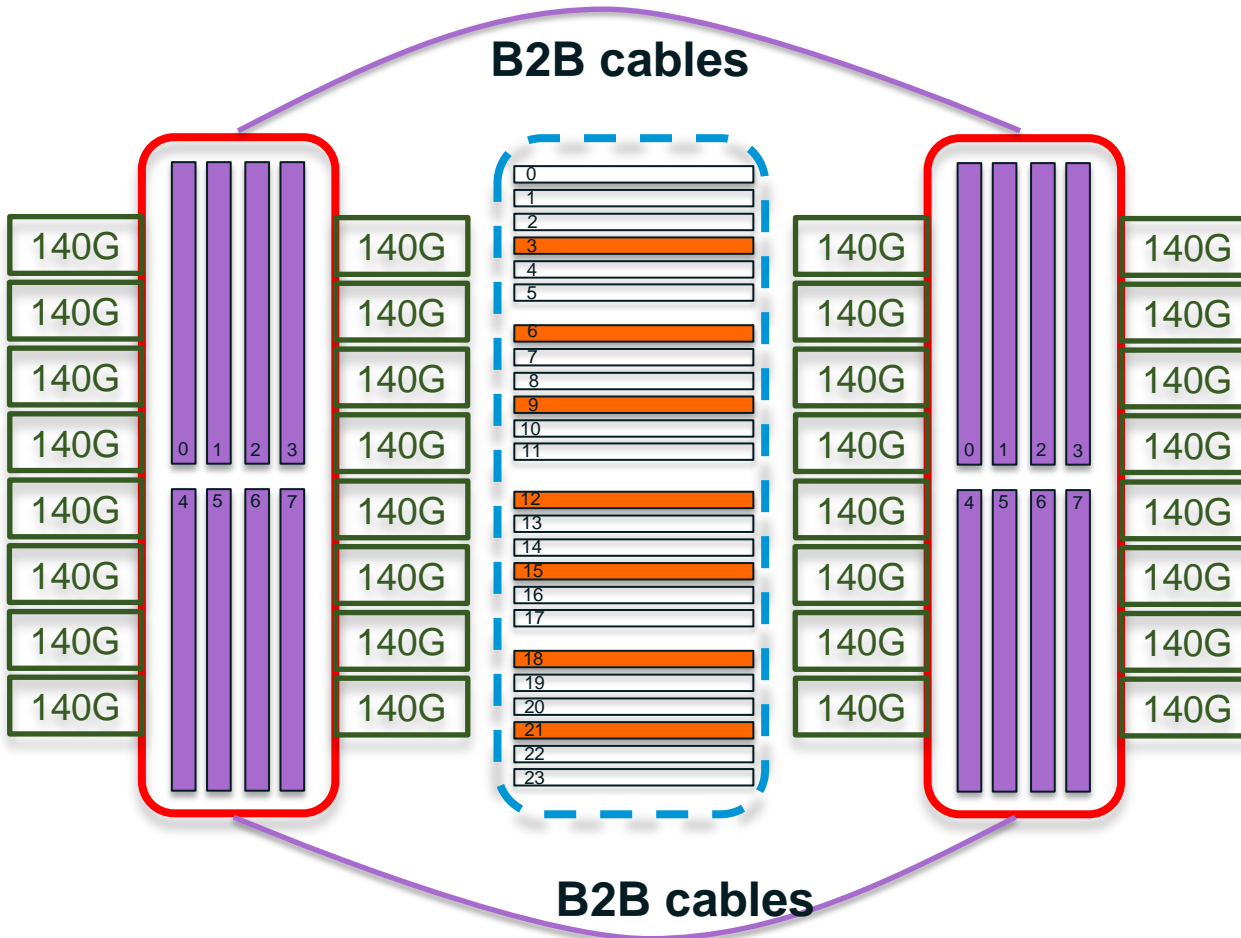


# Migration Path to CRS-X

Migrating MC CRS-3 to B2B CRS-X (3/3)

We don't need to care about fabric addressing modification in this scenario.

BW/slot	Chassis	Fabric
40G	Legacy	40G
140G		
200G		
400G		
40G	Enhanced	40G
140G		140G
200G		400G
400G		



Once the fabric migration is completed, we can migrate the control-ethernet links. And shut down FCC



# IOS-XR Software



# Software Install Terminology

## Software Maintenance Upgrade

- Provides timely temporary point fixes for urgent issues for a given package version
- Fix integrated into the subsequent IOS XR maintenance release.
- Implementation changes only. No interface changes (no changes to CLI, APIs, IPC etc.) or new feature content
- Ideally not traffic impacting (Hitless, non traffic impacting)
- SMU is named by release and bugid - Examples - hfr-rout-3.2.2.CSCei63263.pie



**PIE?**



**Mini?**

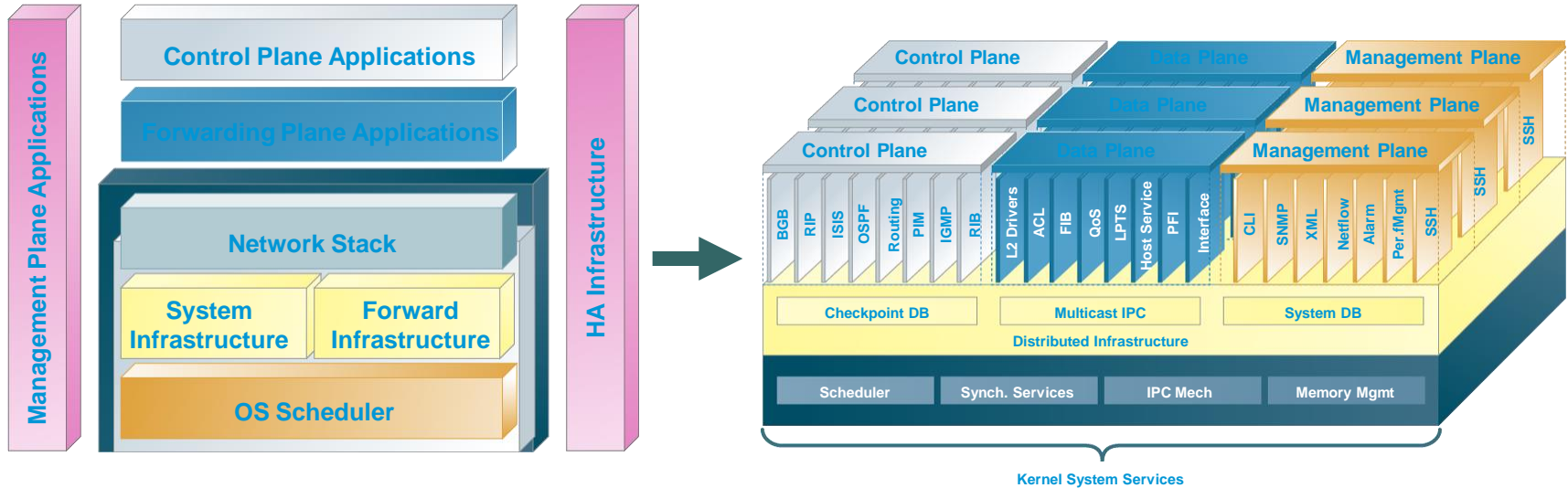


**Package?**



**SMU?**

# Router OS Evolution

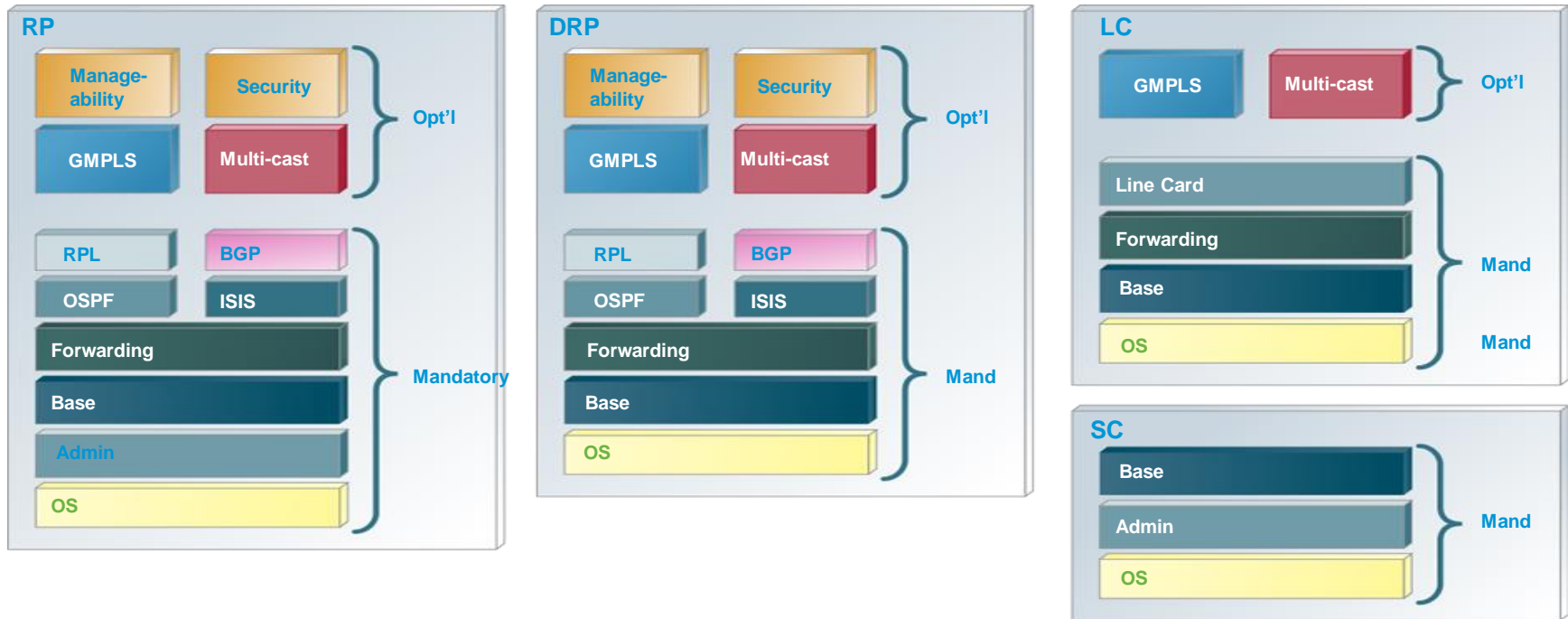


- Monolithic Kernel
- Centralized Infrastructure
- Integrated Network stack
- Centralized applications



- Micro Kernel
- Distributed Infrastructure
- Independent Network stack
- Distributed applications

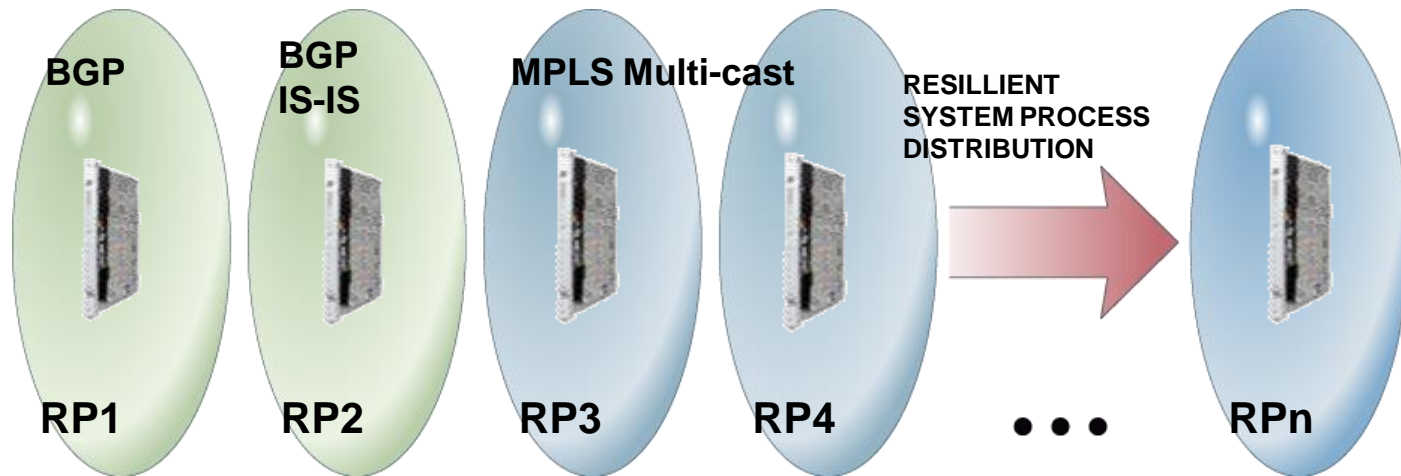
# IOS XR Modular Packaged Software



- Upgrade specific packages/Composites
  - Across Entire system
    - Useful once a feature is qualified and you want to roll it without lot of cmd
  - Targeted Install to specific cards
    - Useful while a feature is being qualified
    - Reduces churn in the system to card boundary
- Point Fix for software faults

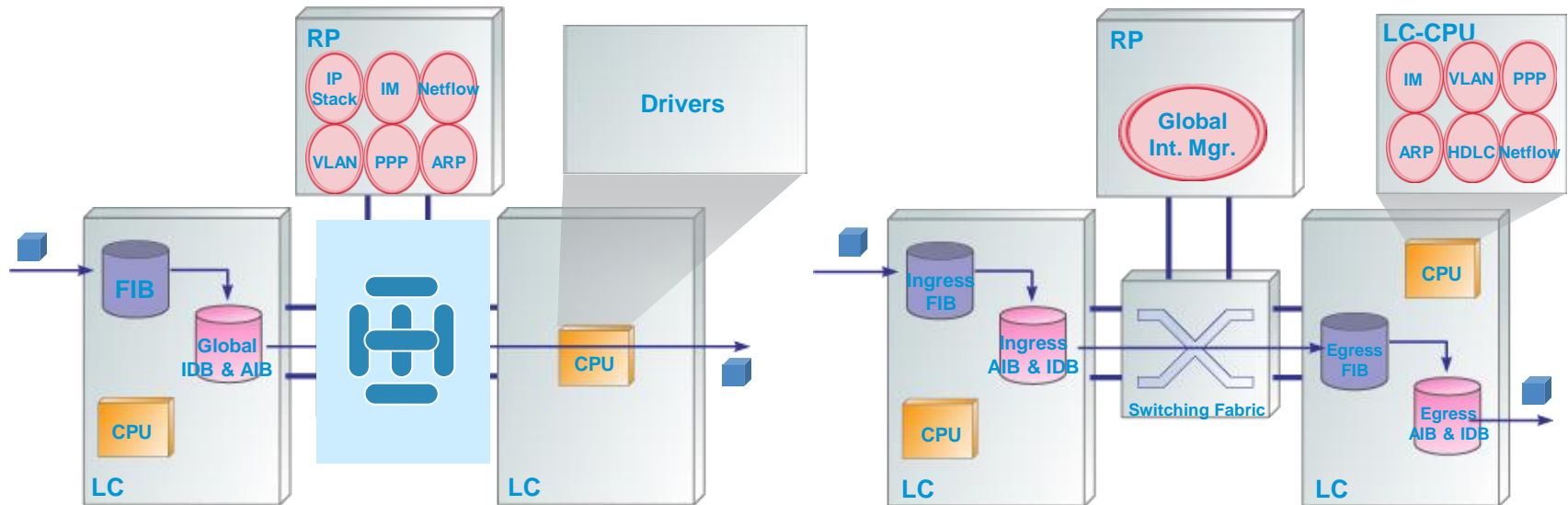


# Distributed Control Plane



- Routing protocols and signaling protocols can run in one or more (D)RP
- Each (D)RP can have redundancy support with standby (D)RP
- Out of resources handling for proactive planning

# Distributed Forwarding Infrastructure



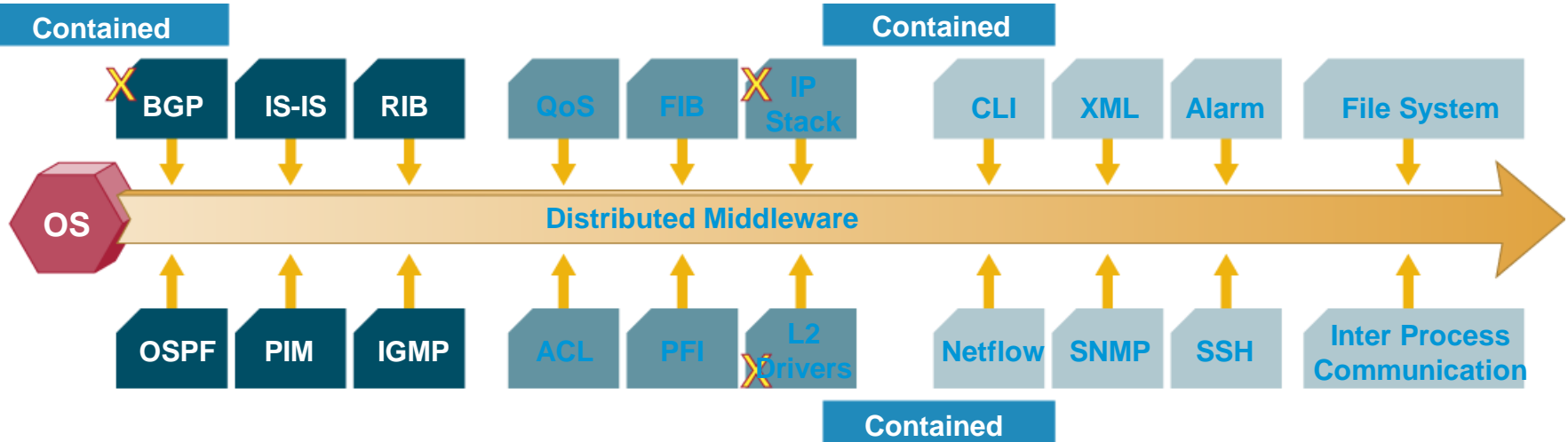
## Single stage forwarding

- Single global Adjacency Information Base (AIB) distributed to all line cards
- Single global Interface Management DB distributed to all line cards
- Only Ingress FIB – forces forwarding features to be run in RP

## Two stage forwarding

- Each line card has independent AIB only for local interfaces
- Each line card has independent Interface DB for local interfaces
- Both Ingress and Egress FIB – allows forwarding features to be independently run in LCs

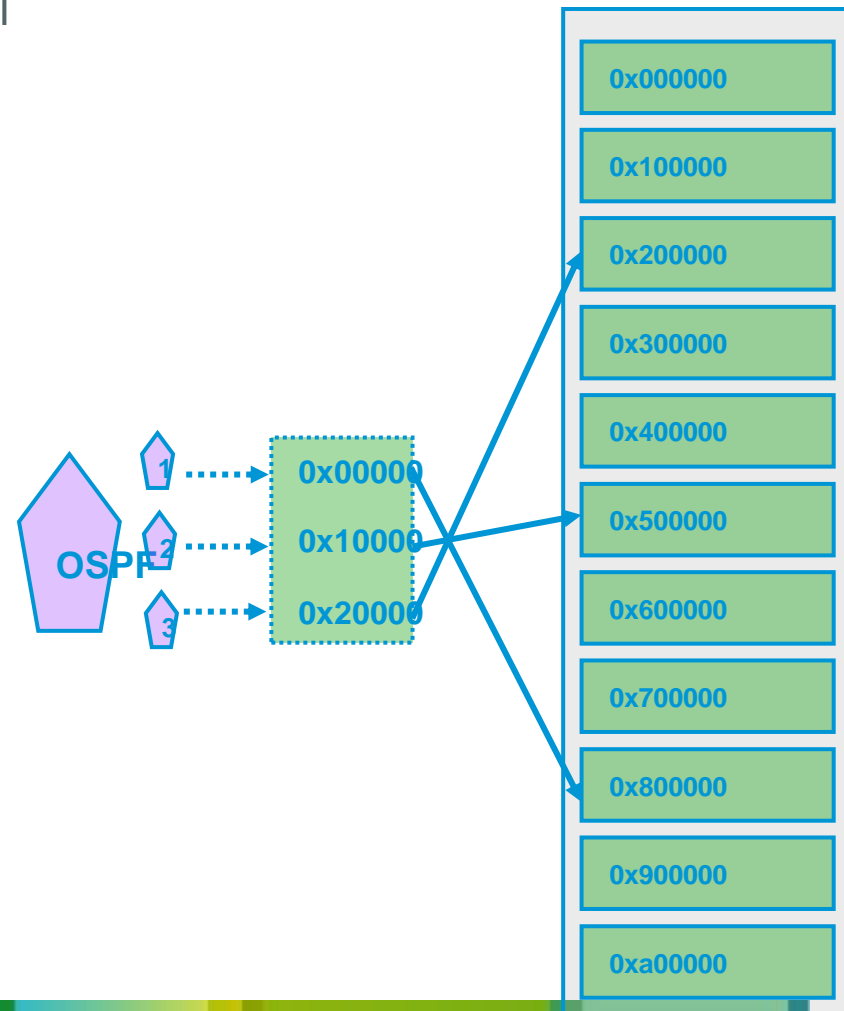
# Process Restartability



- Used for small/contained faults (individual or small groups of process failures)
- Processes support restarting with dynamic state recovery
  - Mirrored State via checkpoint or synchronization with peer
- First line of defense- All Processes are restartable for fault recovery
- Certain processes are '**mandatory**' – must always be running. Failure of mandatory processes can cause RP failover
- Second line of defense - Card-level Redundancy is used when Process Restart fails-

# Protected Process Memory Space

- Each process has a virtual memory space
  - Kernel/MMU maps virtual address to physical address (at page level)
  - Threads share the memory space
- One process cannot corrupt another's memory
  - Process can only access virtual space
  - In IOS – all processes shared same virtual space
- Communication between processes via controlled APIs
- Limited use of shared memory



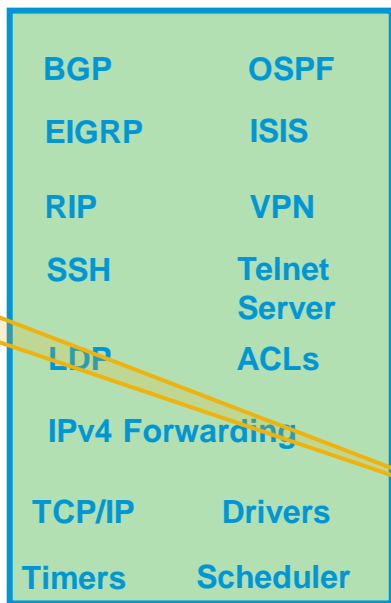
# Process Restart

## Microkernel Architecture Enables Restart of Most Processes

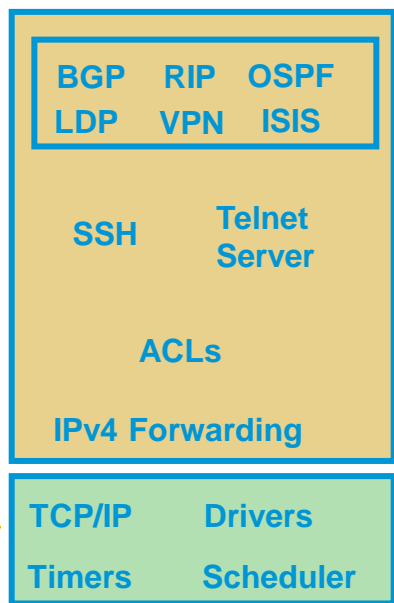
- Microkernel includes minimal functionality
- Non-kernel processes can be restarted
- Critical to HA, ISSU, and MDR functions
- Restarting many processes can be tricky
  - Dependent processes may also need to restart



Green areas cannot restart



**Monolithic**  
*IOS (7200, 12k (IOS))*



**Kernel**  
*BSD based routers*



**Microkernel**  
*IOS XR*

# IOS-XR and IOS Config Differences

- IOS-XR configuration is held in binary form which is quicker to parse and process - ‘show running-configuration’ is just an ASCII representation of the binary data extracted from all nodes in the system
- There is no concept of a startup configuration like in IOS
- If one copies the running config to startup, a backup config with the name “startup” is created
- Router config is based on two stage config model.
- “running” or “active” config can not be modified directly.
- Instead, user config first enters a staging area (first stage)
- Must be explicitly promoted to be part of active config (second stage).

IOS-XR	IOS
Configuration changes do NOT take place after <CR>	Configurations take place immediately after <CR>
Configuration changes must be ‘committed’ before they take effect	No commit
Allows you to verify your configuration before applying it	No verification required
Two stage configuration model	Not available
Configuration rollback	Not available
Provision to pre-configure	Not available
New config plane – Admin mode	Not available
Feature centric	Interface centric

Thank you.

