

Architektura Carrier Routing System

T-SP4

David Jakl – Cisco

Program

- Overview
- Architecture
- Line Card
- Switch Fabric
- Multi-chassis
- Summary

CRS Overview

CRS-1

Carrier Routing System 1

May 2004

>8000 CRS Chassis
>400 Customers

Core, Peering, L3 (L2) Edge Router

- 40Gbps/slot Full Duplex
- Redundant 3-stage non-blocking Beneš Switch Fabric
- IOS XR - Highly modular, Microkernel-based
- SDR - Secure Domain Routers
- IPoDWDM 10GE/40G, Tunable, G.709, (E)FEC, 40G over 10G DWDM System (single lambda)

Control Plane:

Redundant Route Processors (RP → PRP)

More RPs for scale (DPRs) for "Process Placement" or SDR

PLIMs (Front):

Modular: SIP-800/6xSPA: 1/10GE, POS 155M->10G, E3, ATM

Fixed: 1/10GE, POS 2.5G->40G, 10GE/40G IPoDWDM

Services: CGSE = Carrier Grade Services Engine

Forwarding Line Cards (Back):

MSC40 – Core/Peering/Edge (4/8/16/MC)

2M routes, H-QoS 8000 queues/port, 2000 intfs

FP40 – Thin Core/Peering (4/8) +licenses

2M routes, 8 queues/port, 100 intfs



4-slot
320Gbps



8-slot
640Gbps



16-slot
1280Gbps



TODAY Multi-chassis MAX
→ 10.2 (8xLCC) → 92 Tbps (72xLCC)

CRS-3

Carrier Routing System 3

Powered by QuantumFlow Array

9/3/2010

FCS Q3/2010

140Gbps/slot Full Duplex

- 140G Switch Fabric (4/8/16/MC)

PLIMs (Front):

1x100GE Line Rate, CFP (L4)

14x10GE Line Rate, LAN/WAN PHY, XFP

20x10GE Oversubscribed 140G, LAN/WAN PHY, XFP

2012: 100GE OTN IPoDWDM PM-QPSK over 10G DWDM System (single lambda)

Forwarding Line Cards (Back):

MSC140 – **Edge** (4/8/16/MC)

4M routes, H-QoS **64000** queues/port, **12000** intfs

FP140 – **Core/Peering** (4/8/**16/MC**) +licenses

1/4M routes, 8 queues/port, **250** intfs

LSP140 – **MPLS Core** (limited IP) (4/8/16/MC) +licenses

HW Ready: E-OAM, Video monitoring, Time stamping, SyncE

CRS-1 → CRS-3

- **NO Chassis/Power/Cooling/RP upgrade**
- **Switch Fabric upgrade: Hitless = Zero packet loss**
- **40G & 140G Cards in the same chassis**



4-slot
1.12Tbps



8-slot
2.24Tbps



16-slot
4.48Tbps



TODAY Multi-chassis **MAX**
→ 35.8 (8xLCC) → 322 Tbps (72xLCC)

CRS-3 PLIMs

CRS-3	#100GE LR	#10GE LR	#10GE OS
4-slot	4	56	80
8-slot	8	112	160
16-slot	16	224	320
MC 8 LCC	128	1 792	2 560
MC 72 LCC	1 152	16 128	23 040



1x 100GE



CFP



14x 10GE



20x 10GE

CRS in 2012 and 2013

Subject to change

2012

- CRS-3 16-slot B2B (2+0) Multichassis
- 100GE SR10 CFP
- 100GE IPoDWDM PLIM
- 4x40GE OTN PLIM – CFP: SR4, LR4, FR
- Tunable DWDM XFP
- FlexPLIM – 6x10GE LAN/WAN/OTN + 4xSPA

2013

- CRS-3 8-slot B2B (2+0) Multichassis
- CGSE+ 80Gbps
- 400G/slot – Switch Fabrics, Line Cards, PLIMs

CRS Architecture

CRS Architecture

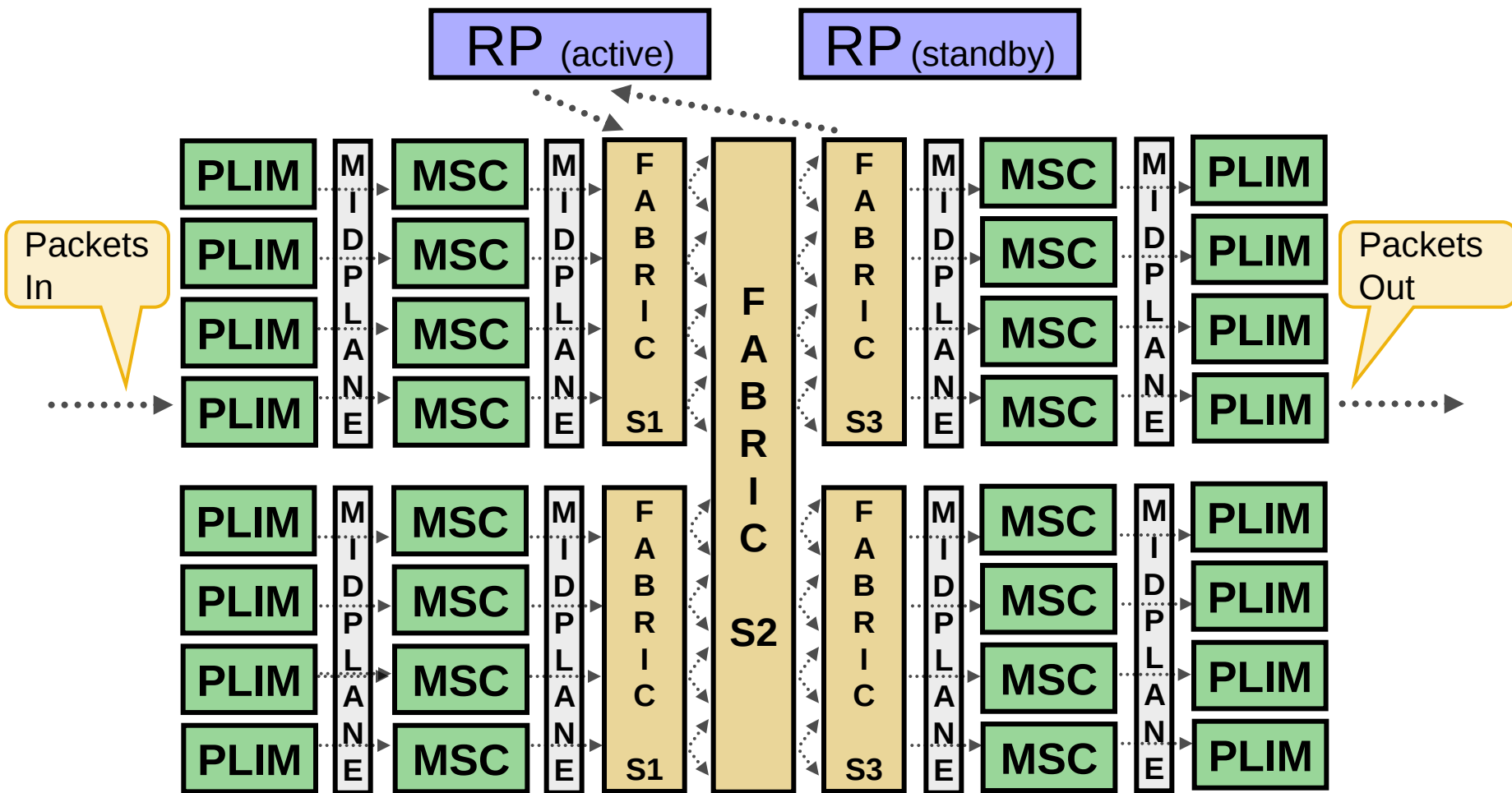
PLIM – Physical Layer Interface Module

MSC – Modular Service Card

RP – Route Processor

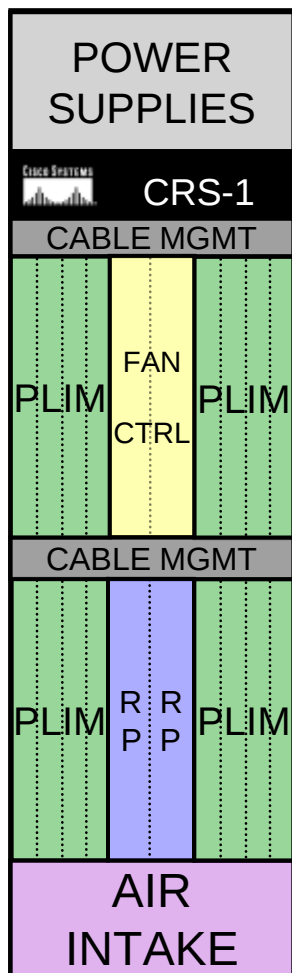
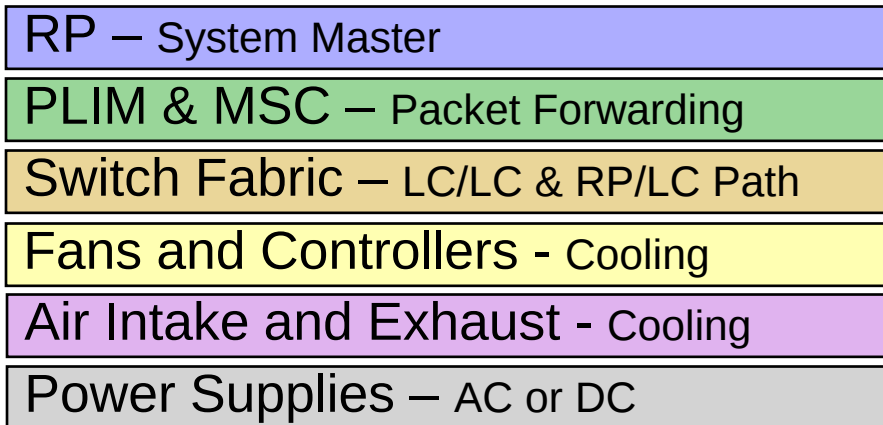
Switch Fabric

- 4 or 8 redundant planes
- Multi-Chassis option



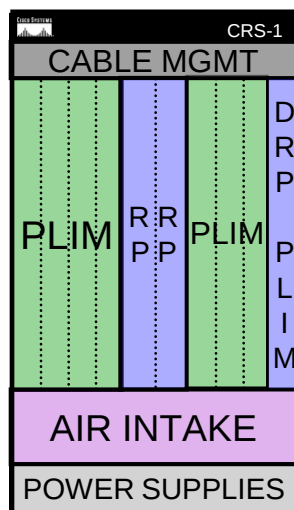
CRS High Level Physical Architecture

Line Card Chassis (LCC)

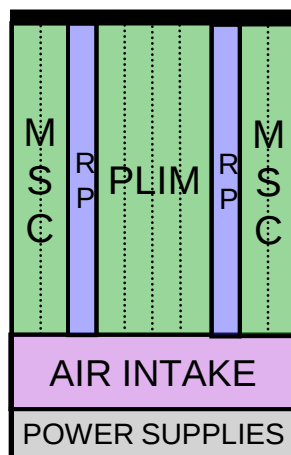


16 Slot

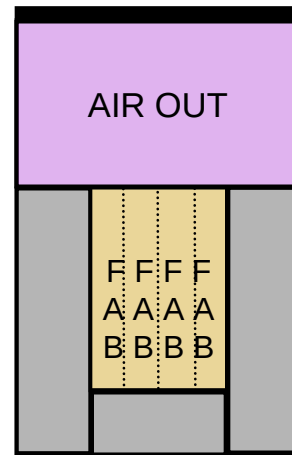
Front View



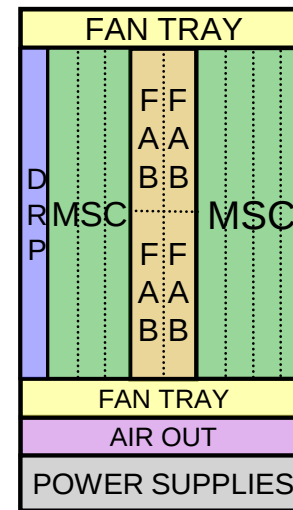
8 Slot



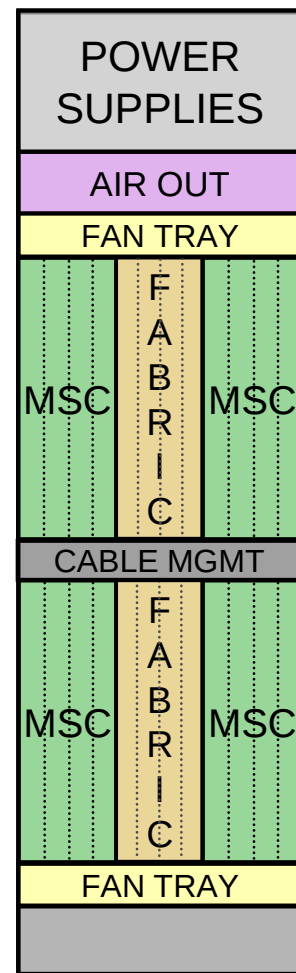
4 Slot



4 Slot



8 Slot



16 Slot

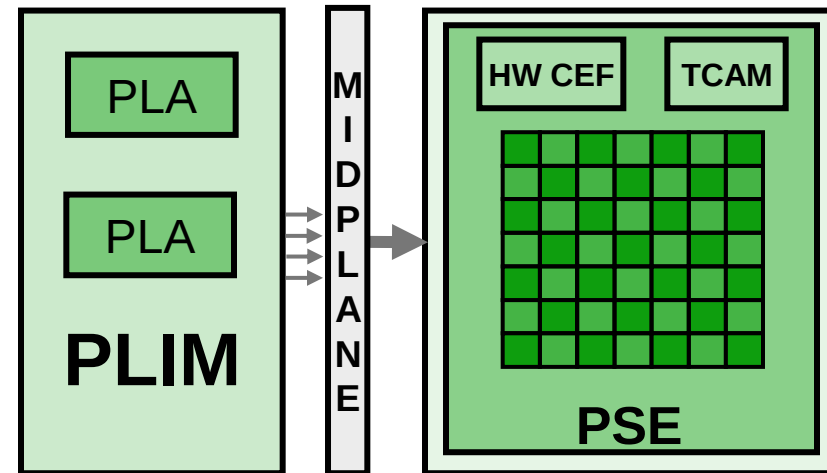
Rear View

CRS Line Card

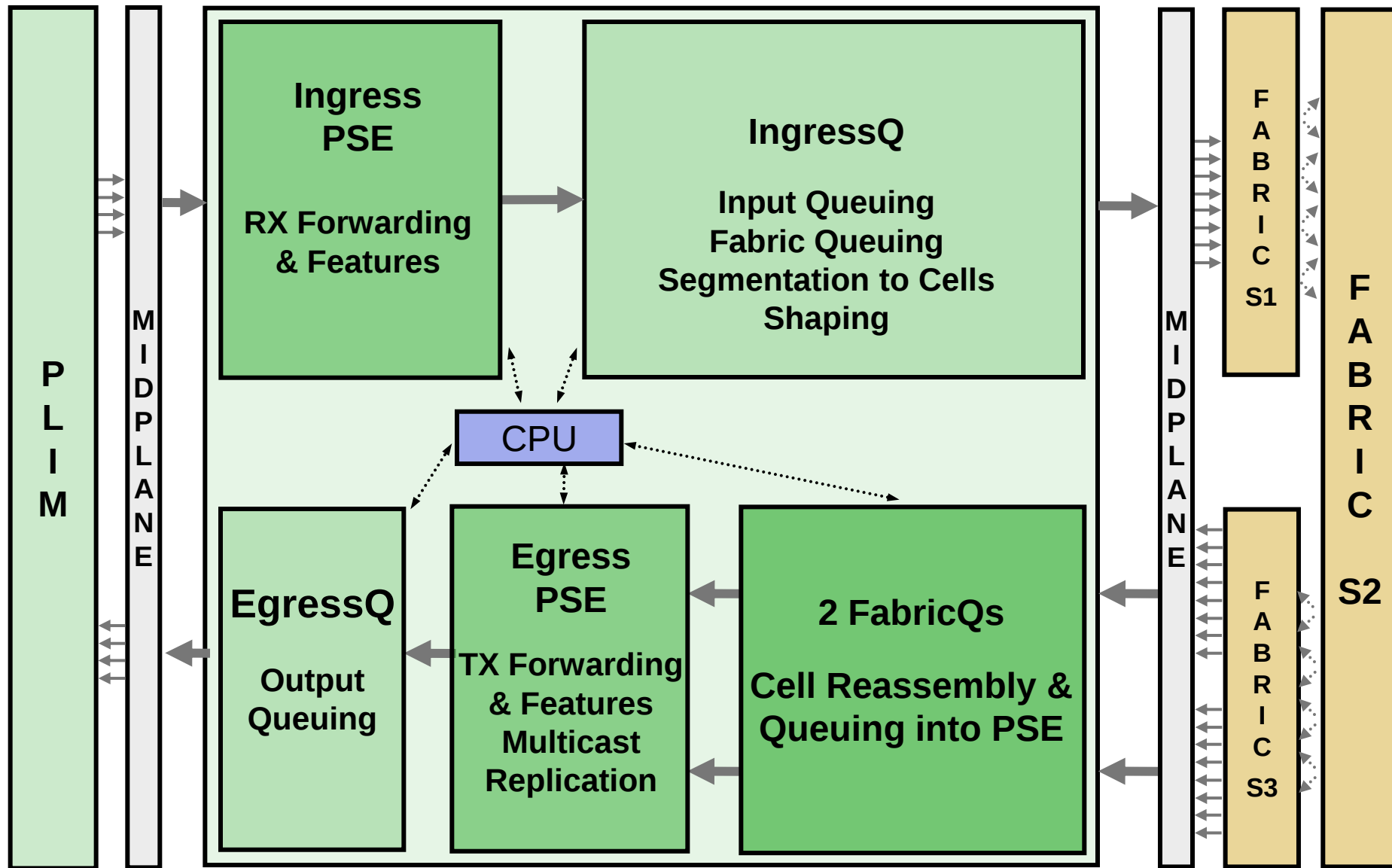
CRS-1 PLIM

- 1 → 4 PLIM ASICs (PLAs) depending on card type
- Nominally 40 Gbps for CRS-1
- Oversubscription allowed when all Ethernet
- 96 Gbps aggregate bandwidth into PSE = no bandwidth bottleneck even when oversubscribed

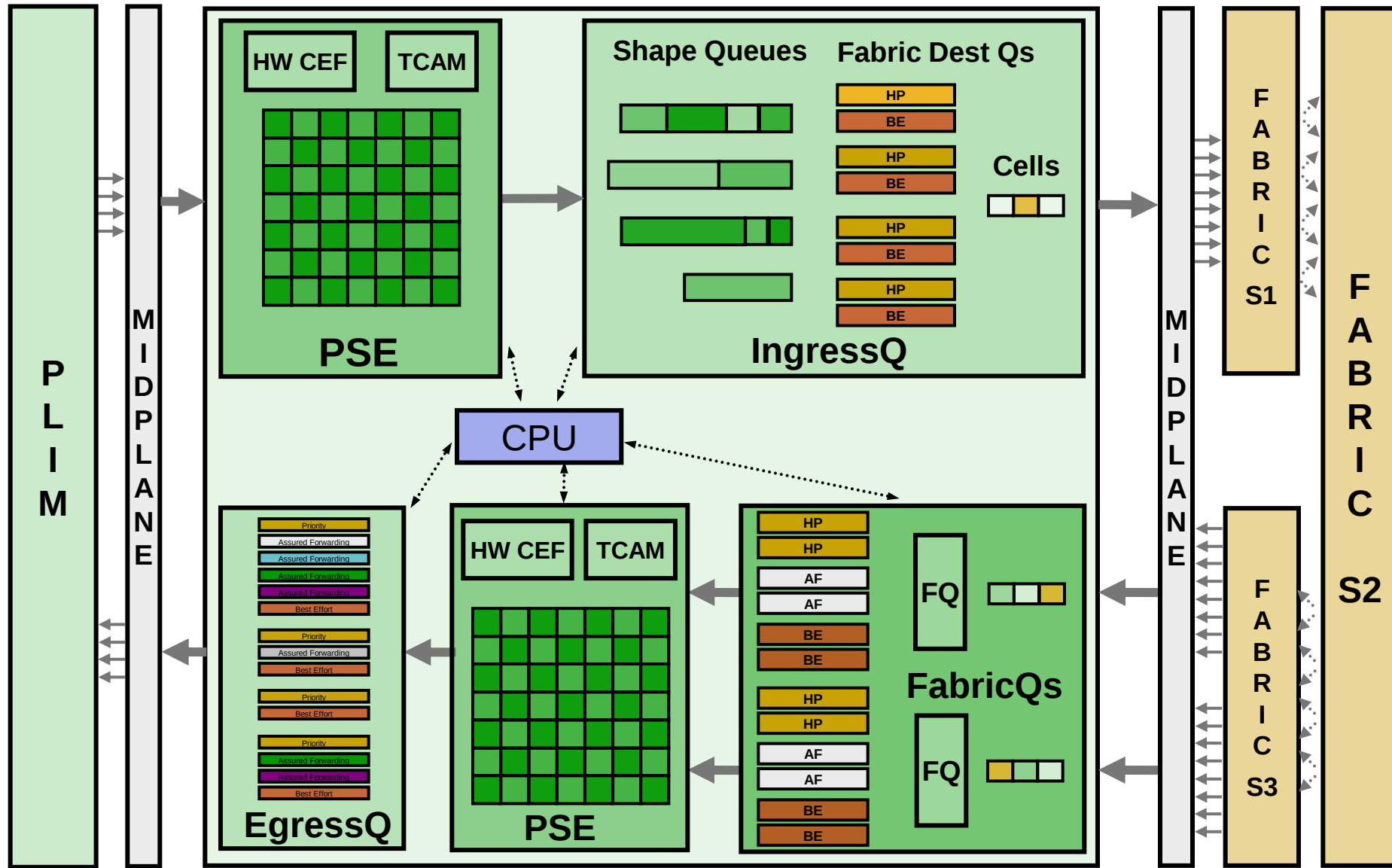
PLIM	PLAs	BW to PSE (per PLA)
4xOC192 POS	2	48 Gbps
16xOC48 POS	4	24 Gbps
1xOC768 POS	1	96 Gbps
4/8xTenGE	2	48 Gbps
SIP-800	2	48 Gbps



CRS Line Card Architecture

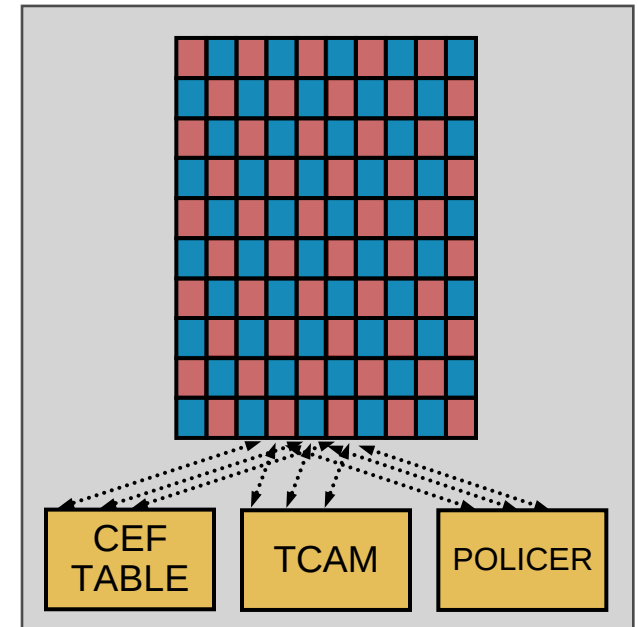


CRS Line Card Architecture - Details

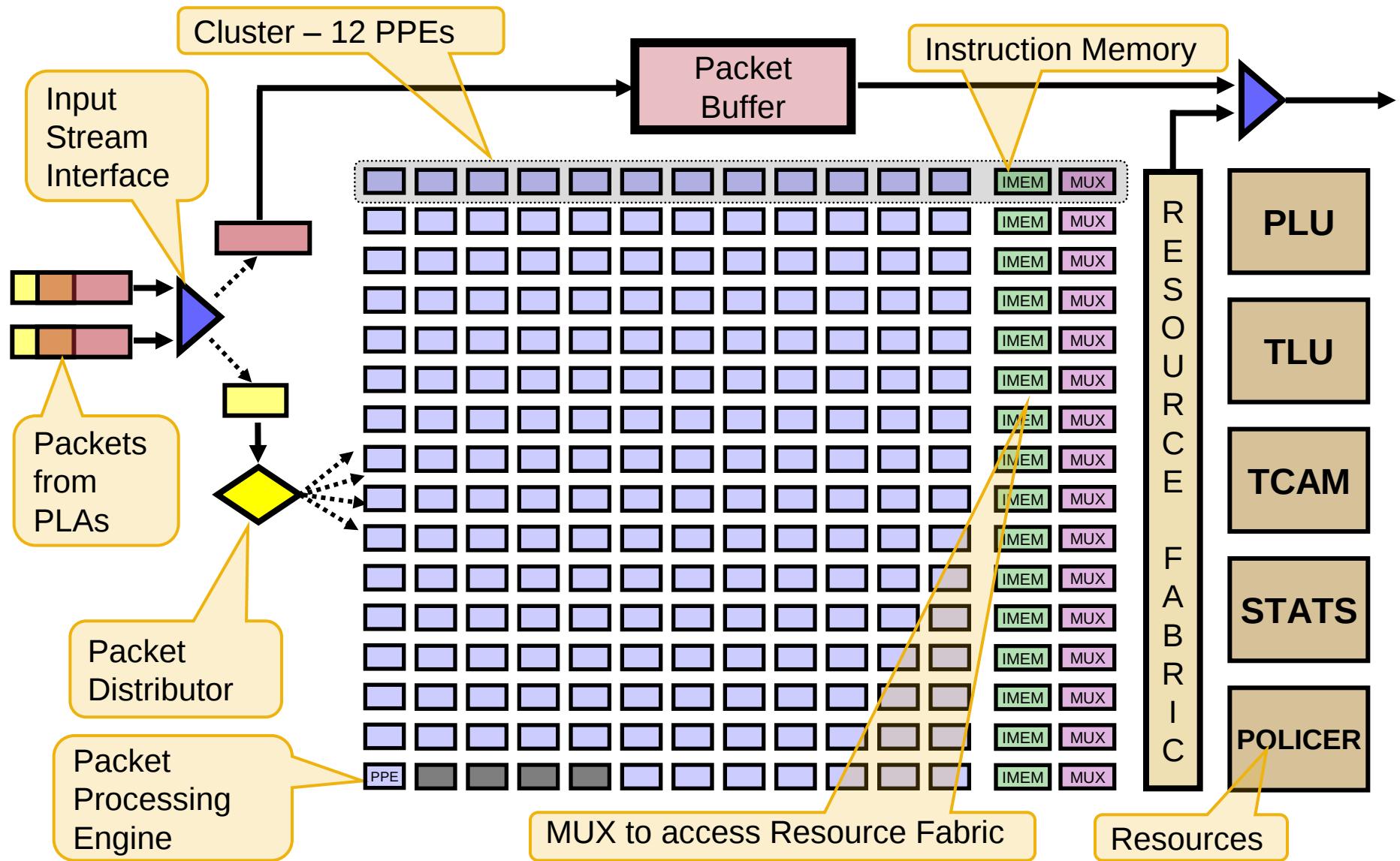


CRS PSE = Packet Switching Engine ASIC

- 188 Parallel Processing Engines (PPE)
 - Independent operation, not pipelined
 - All PFEs can access forwarding resources
- Micro-coded for Service Flexibility
- Performs per packet operations
 - IPv4, IPv6, MPLS, and Multicast lookups
 - Statistics
 - ACLs and Netflow accounting
 - Policing, Marking and WRED
- Access to memories and TCAMs
- Adds 8-byte buffer header

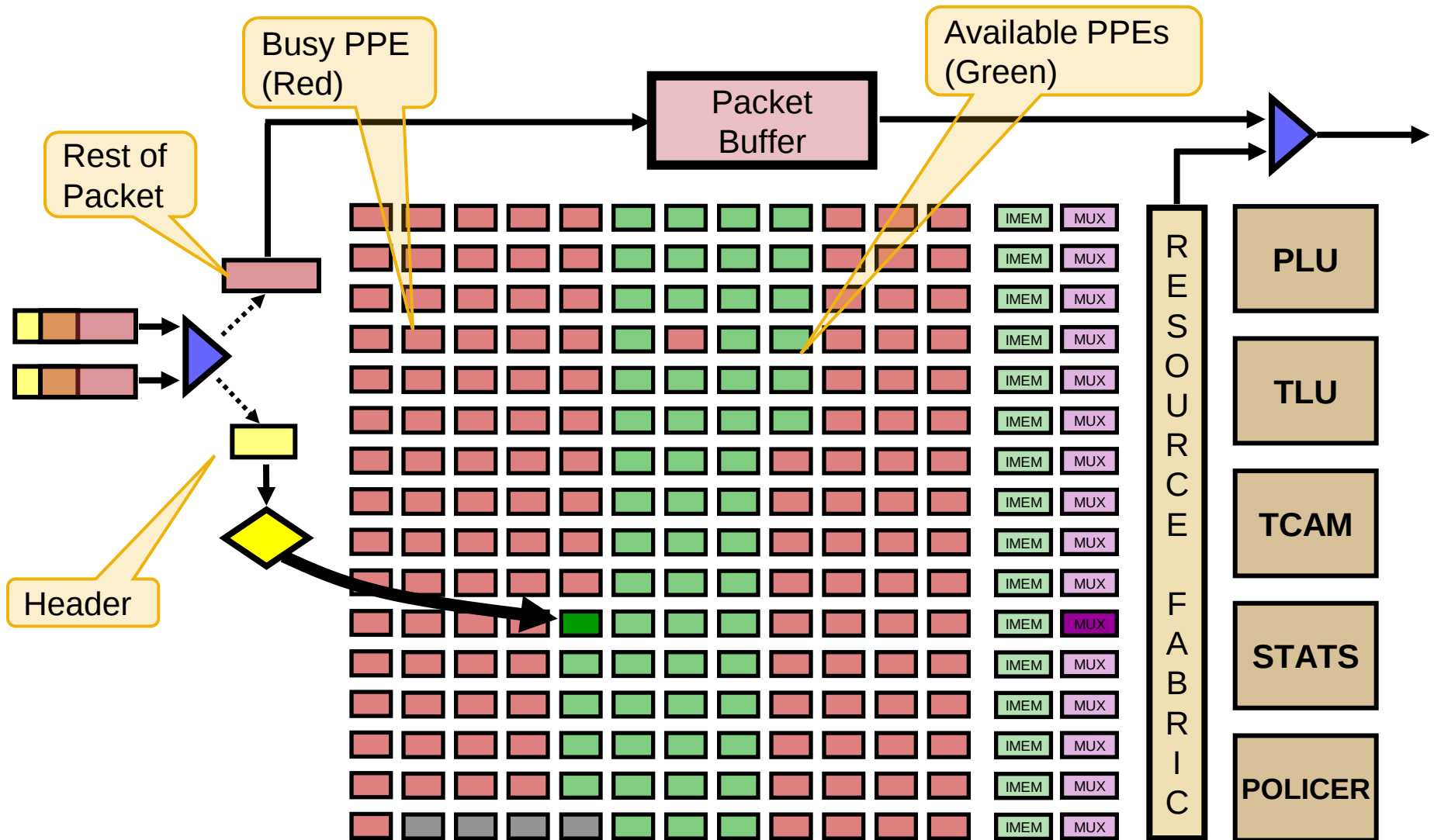


CRS-1 PSE Forwarding ASIC Architecture



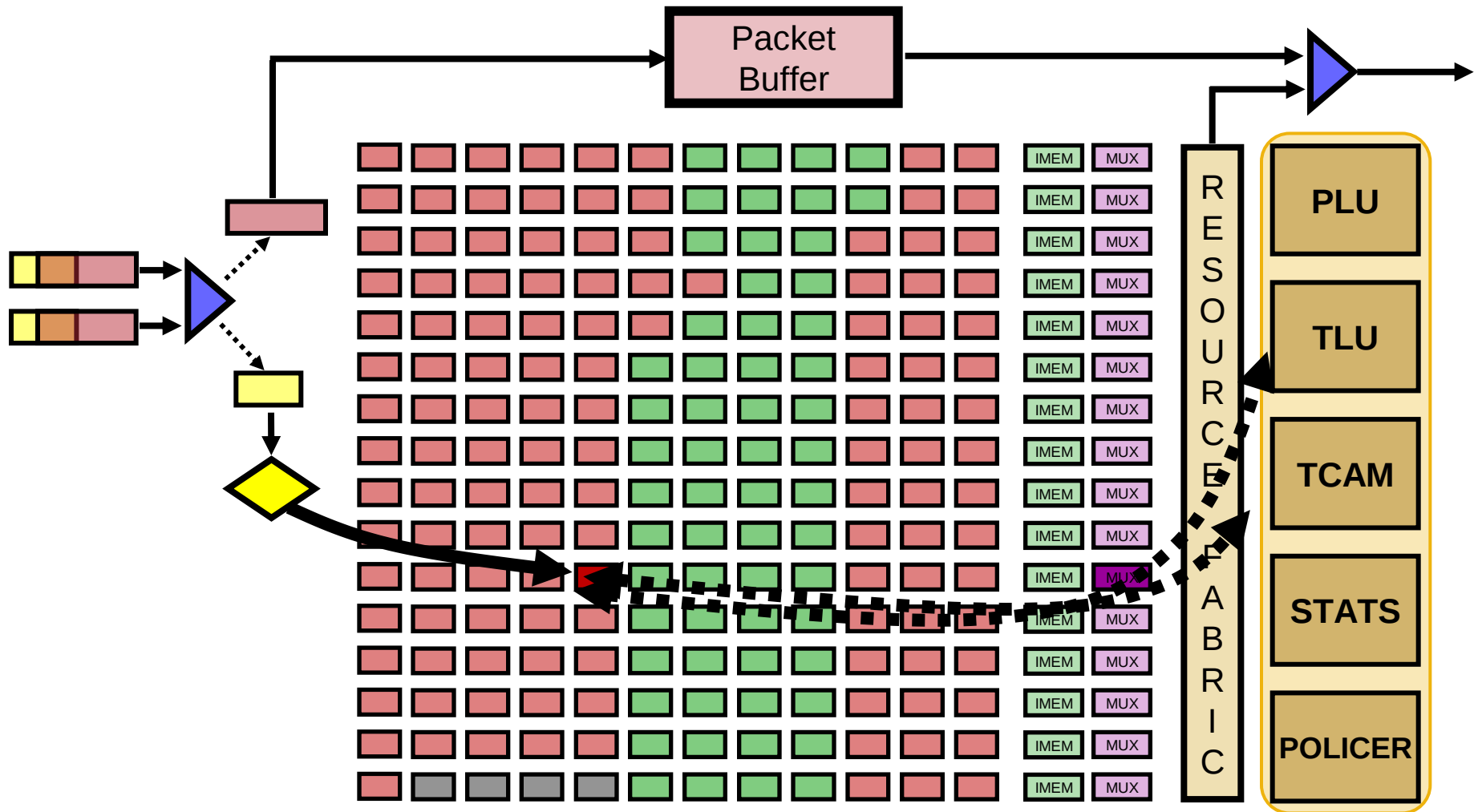
Packet Path within PSE

Assign Packet (Header) to a PPE



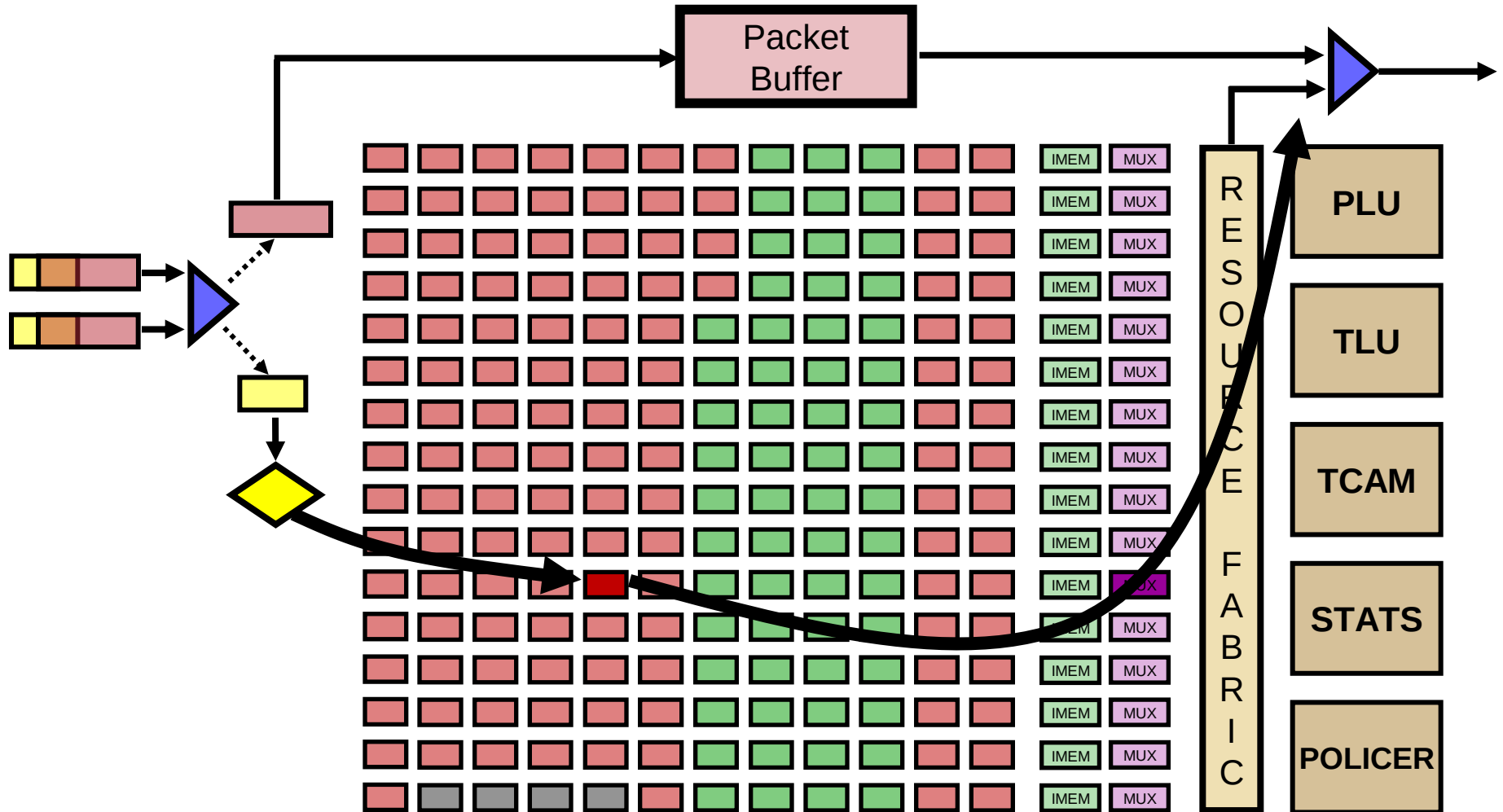
Packet Path within PSE

Perform lookup and features using Resources

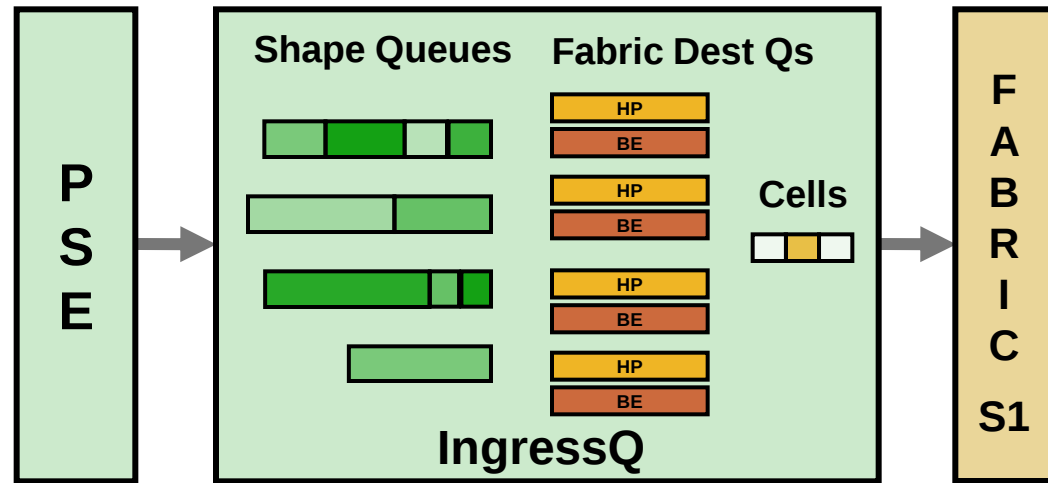


Packet Path within PSE

Recombine header and tail and send to IngressQ



IngressQ



- **Input Shaping Queues**

Per Interface

HP & LP per class if configured (max 8k queues)

- **Fabric Destination Queues**

HP & LP for every FabricQ in system

4 queues for every MSC in entire system

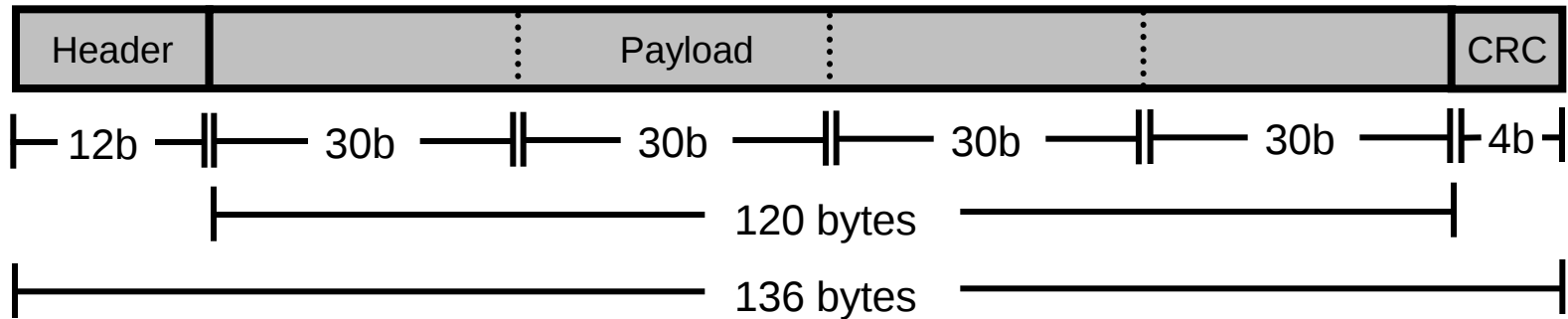
Queue determined by Ingress QoS or Fabric QoS

- **Segmentation of packets into cells**

- **45 Gbps limit between Shape Queues and Fabric Destination Queues (140 Gbps in CRS-3)**

- **Discard bitmap**

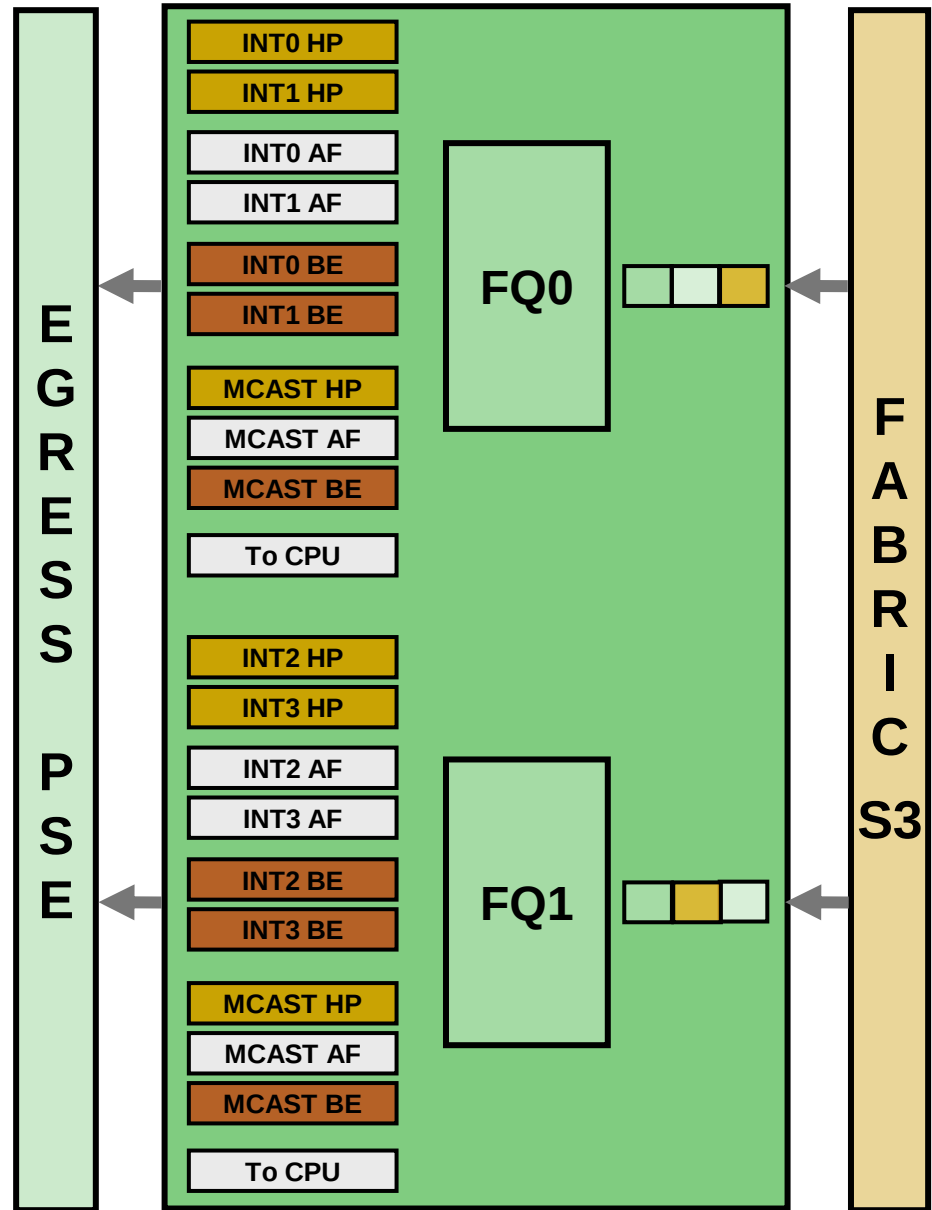
CRS Fabric Cells



- 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 8 fabric planes (4 for 4 slot CRS)

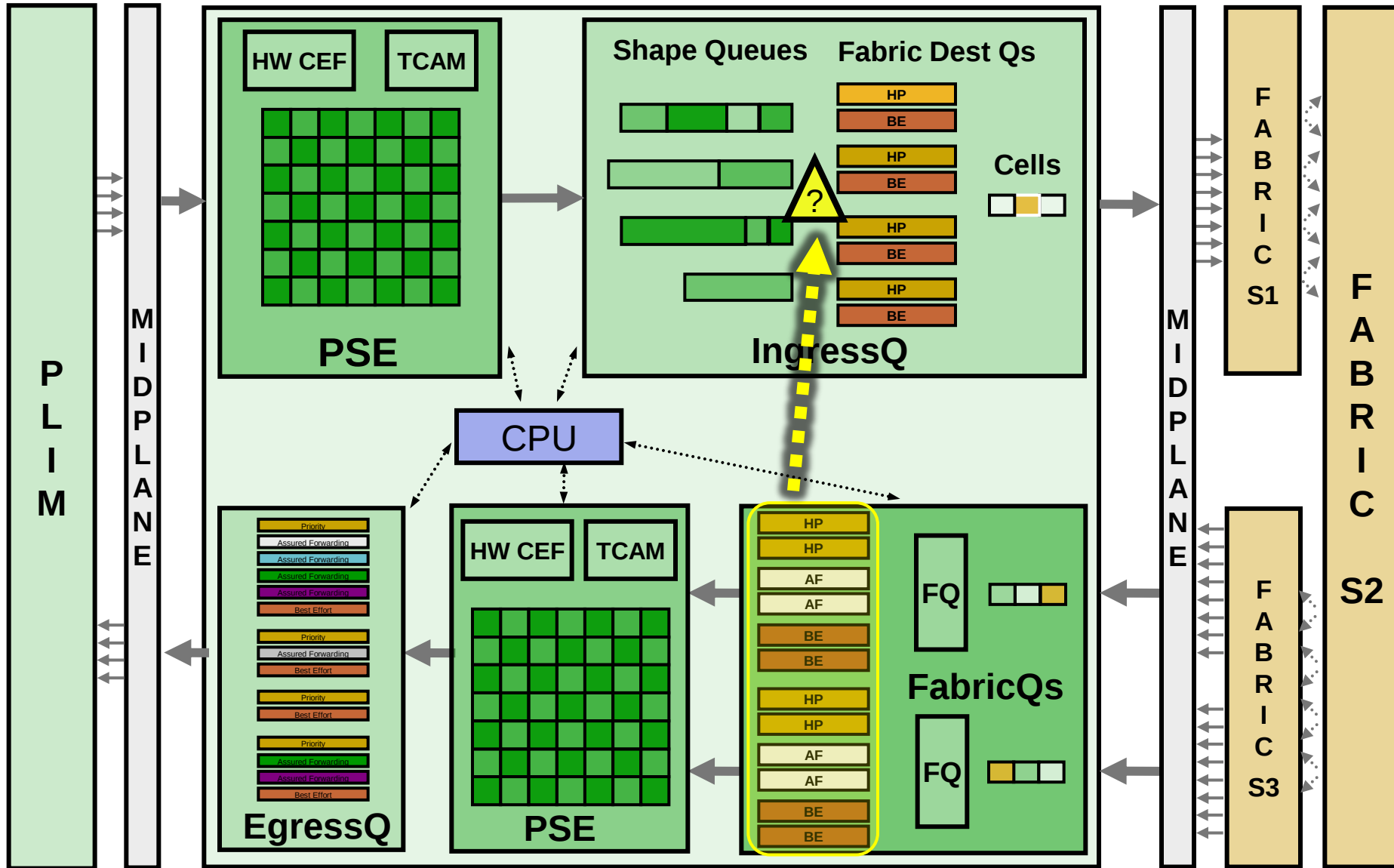
FabricQ Queues

- 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

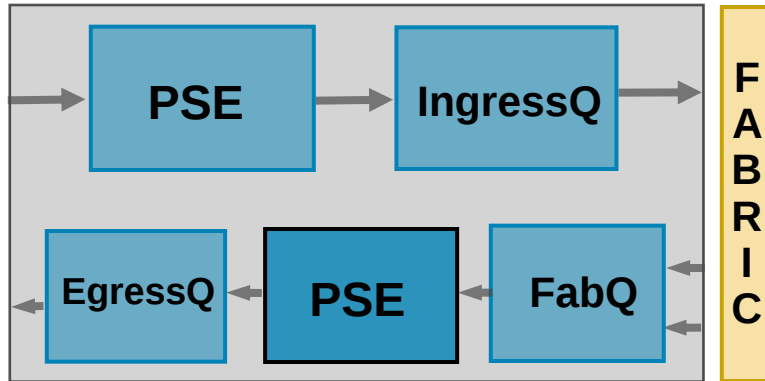


Discard Bitmap Concept

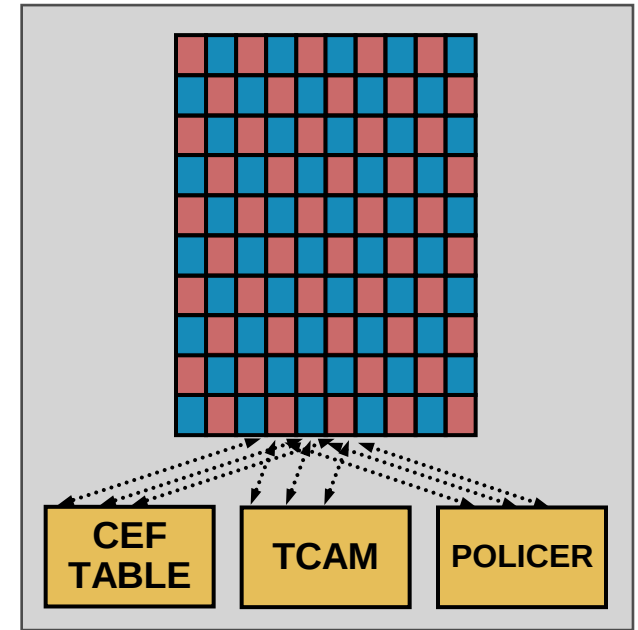
Drop on Ingress when a FabricQ queue is congested



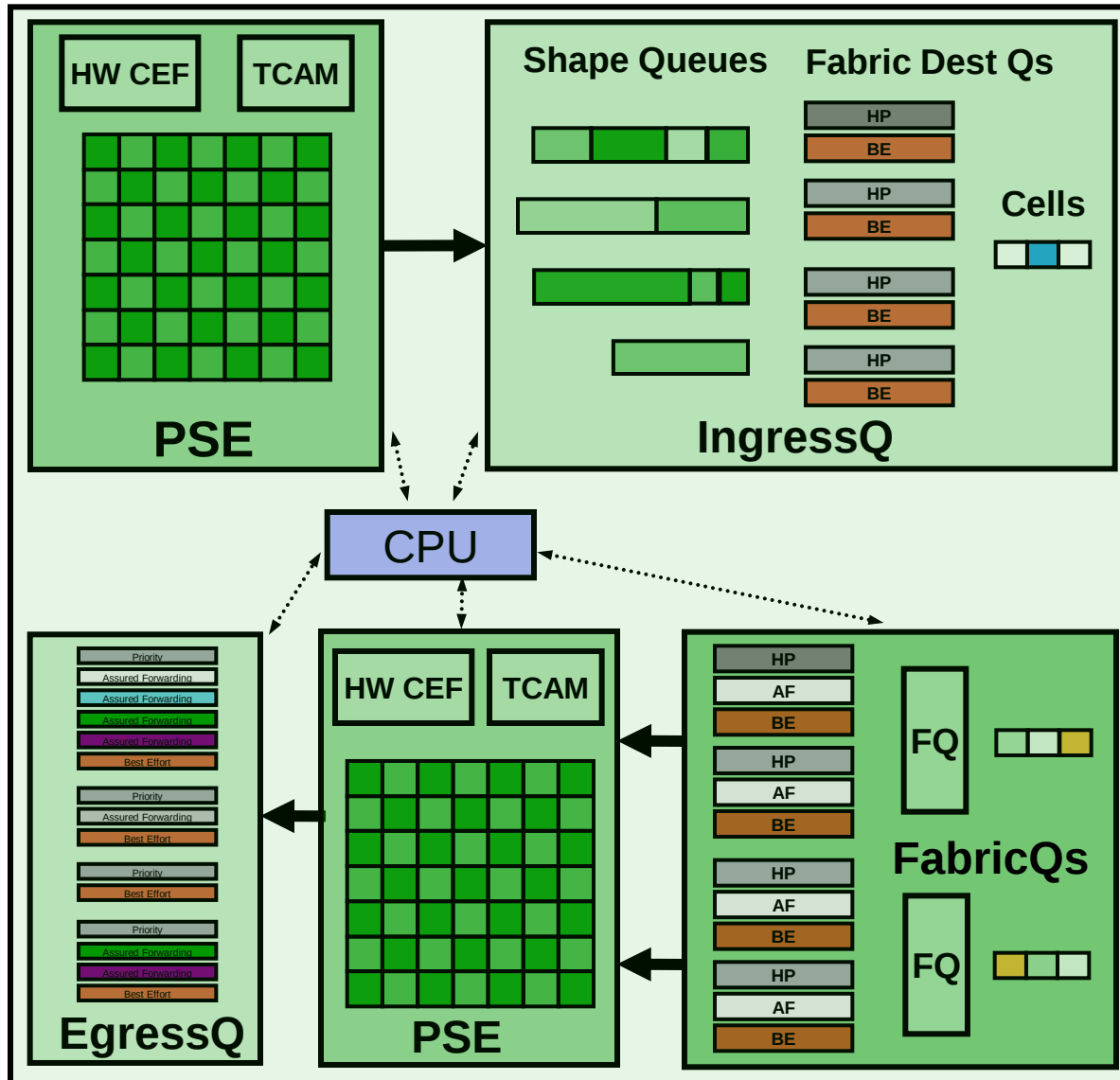
Egress PSE



- Identical hardware to Ingress PSE
- Performs lookup for output adjacency
- Performs output features: Policing, Marking, WRED, Netflow, ACLs, Statistics, ...
- HW multicast replication for multiple ports on same line card



CRS 2-Stage Lookup Improves Scaling



Ingress PSE selects

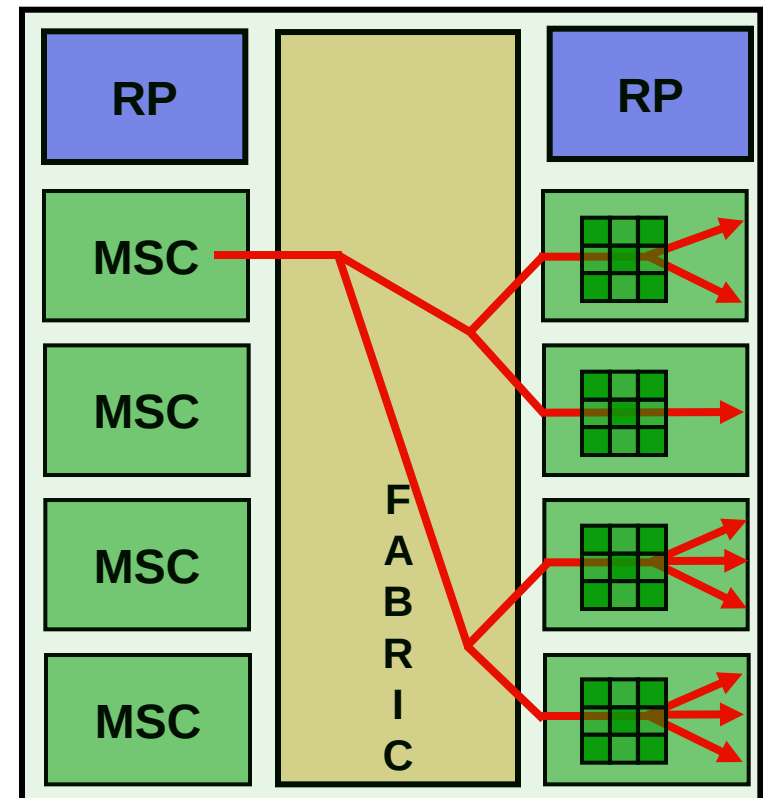
- Output Line Card
- Output Interface
- Fabric Dest Q (HP/LP)
- FabricQ (FQ0 or FQ1)
- IngressQ Queue
 - HP, AF, or BE

Egress PSE selects

- Output Interface
- Output Queue
- Adjacency
- Dest MAC address

Hardware Multicast Replication

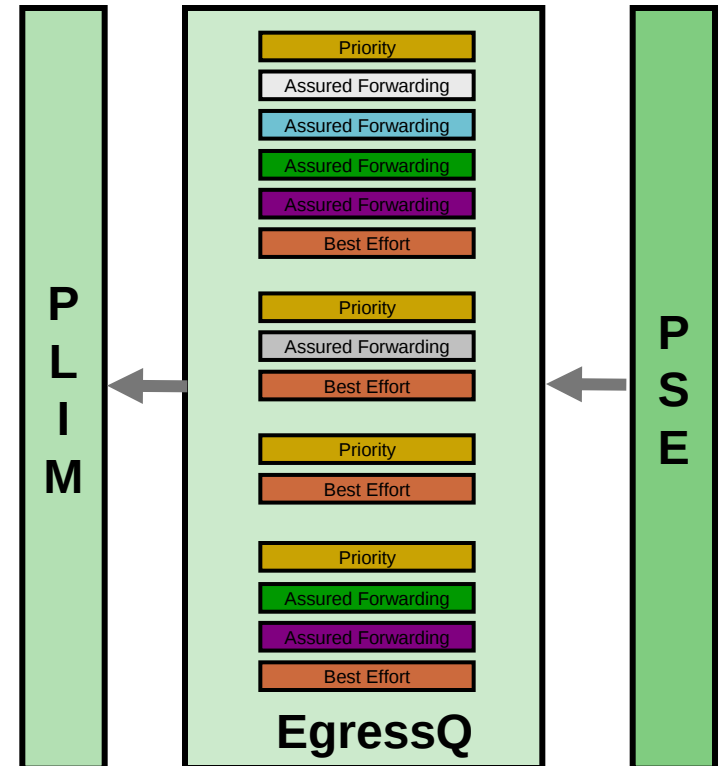
- HW Replication within fabric planes
 - For cells going to multiple line cards
 - No performance impact for additional replications
- HW Replication on Egress PSE
 - For multiple ports on a line card
- Efficient scale for high fan-out
 - No increase in load on MSC



1 plane of 8 (4) shown

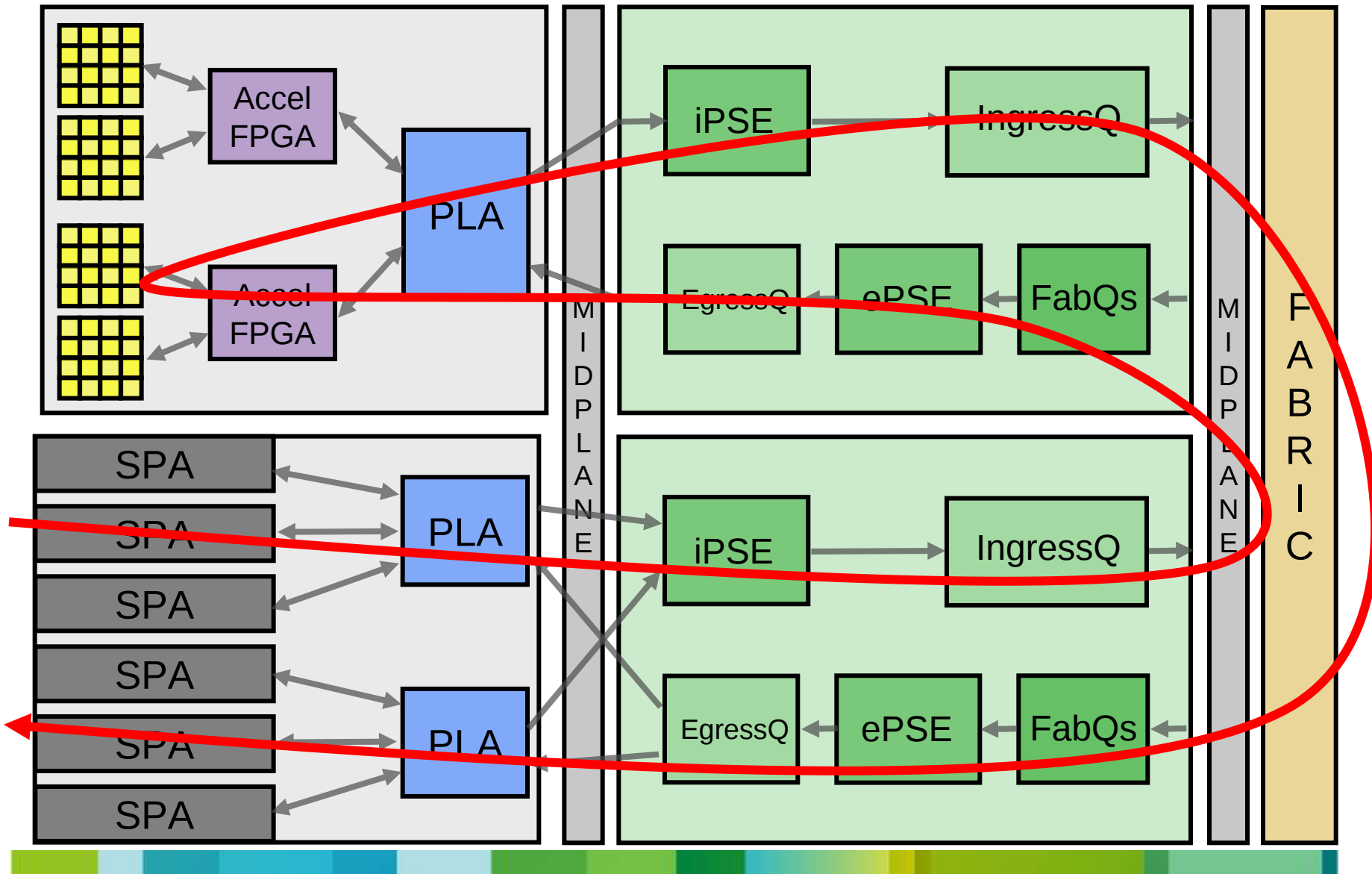
EgressQ

- Per interface/sub-interface queuing
- 8k queues
- 1GB packet buffer
- P2PMDRR
- MQC configuration
 - Strict HP queue
 - Bandwidth guarantees
 - Shaping
 - Bandwidth remaining
- 3 Level Hierarchy
 - Port – Highest Level Queuing Engine
 - Group – Middle Level Queuing Engine
 - Queue – Lowest Level Queuing Engine

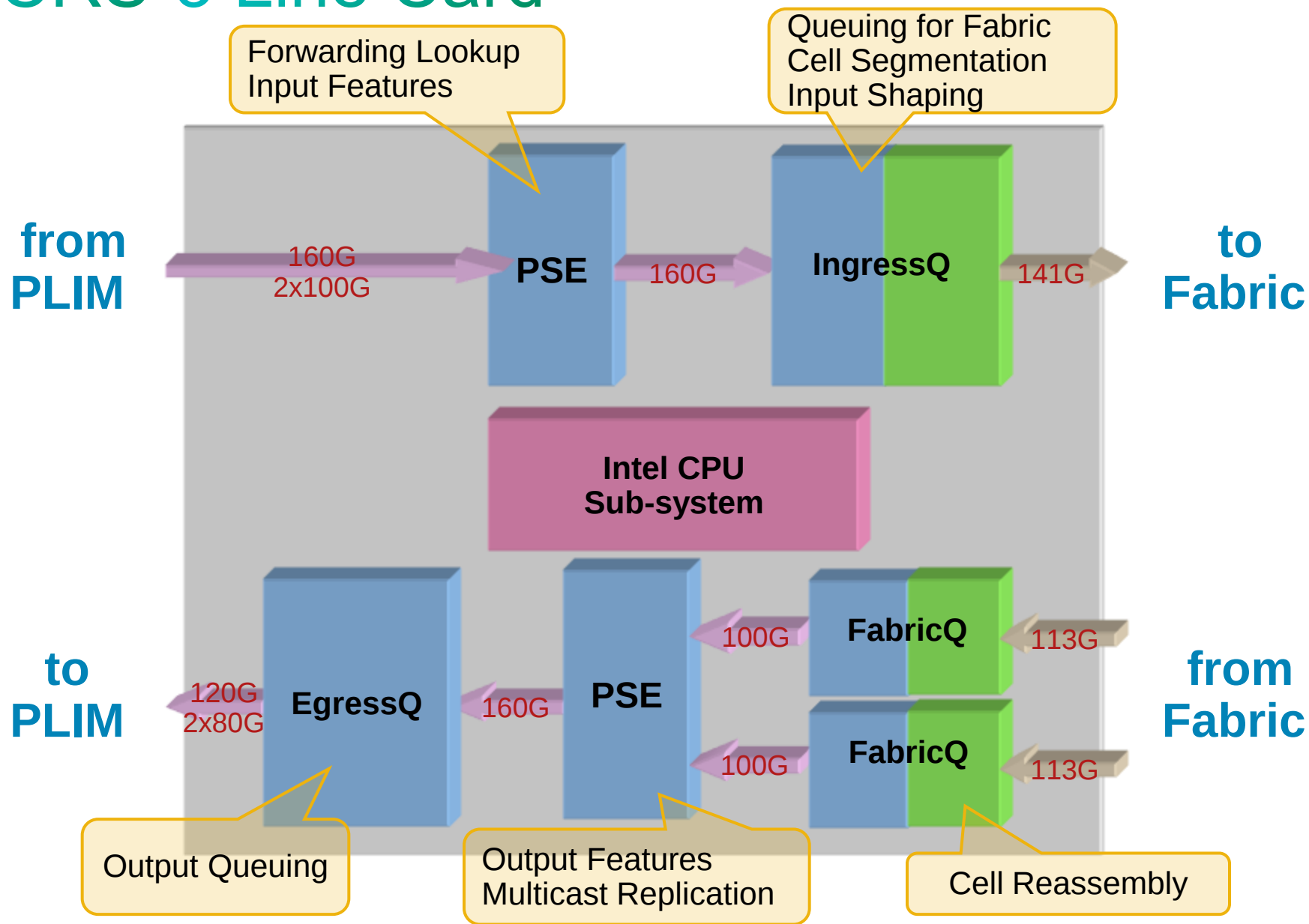




CGSE Packet Flow

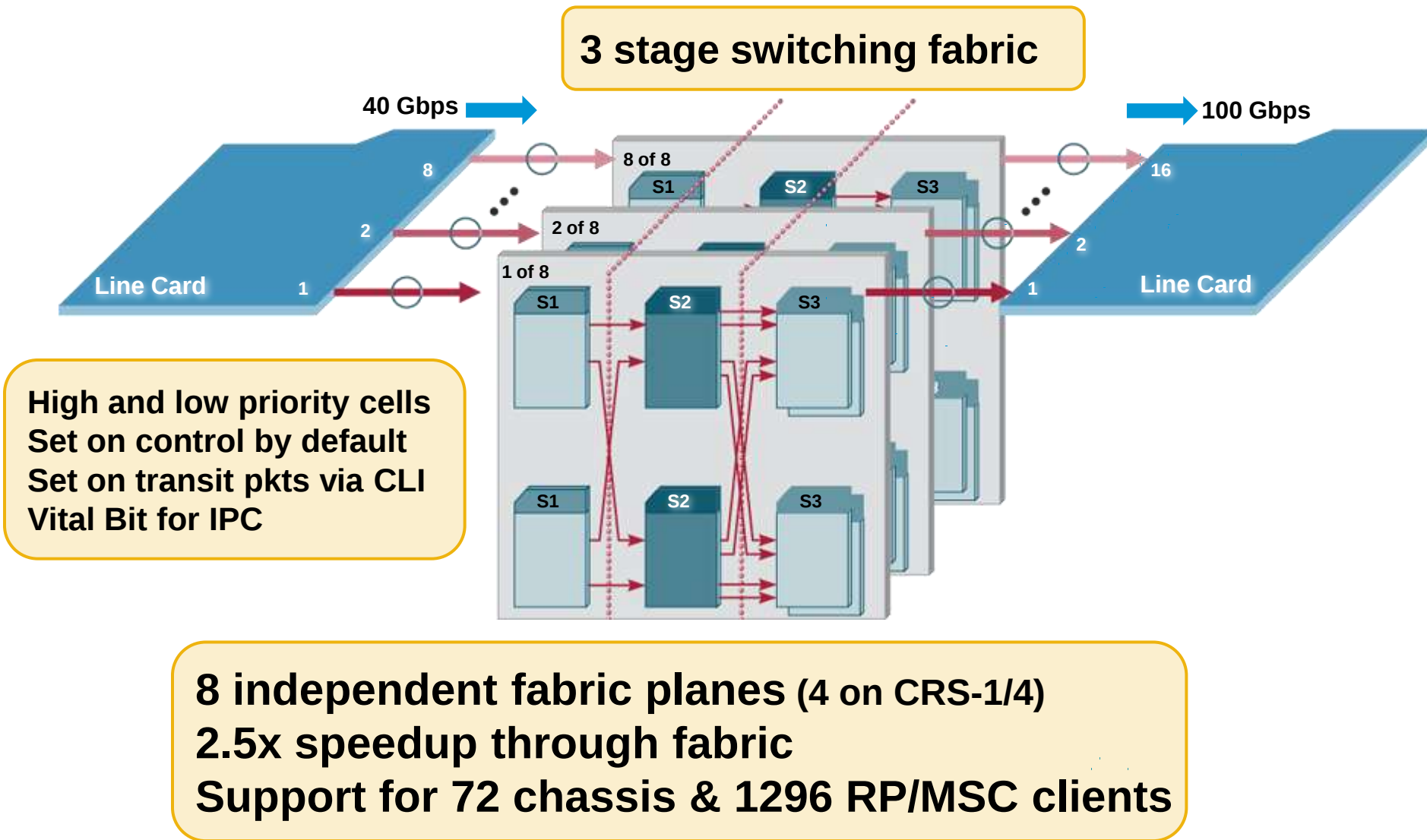


CRS-3 Line Card



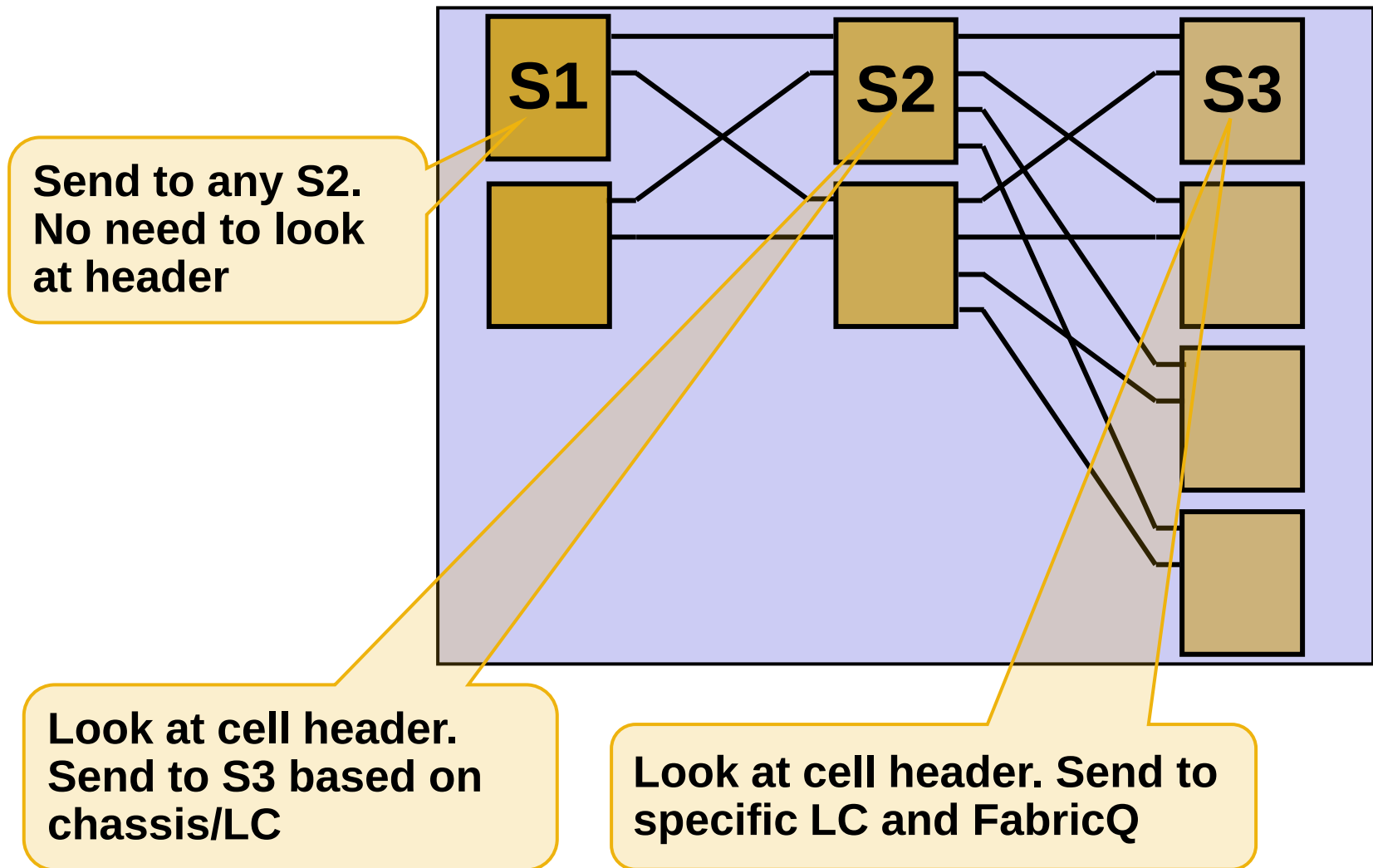
CRS Switch Fabric

CRS-1 Switch Fabric Overview



Decisions at Each Stage in 16 slot

For each of the 8 planes

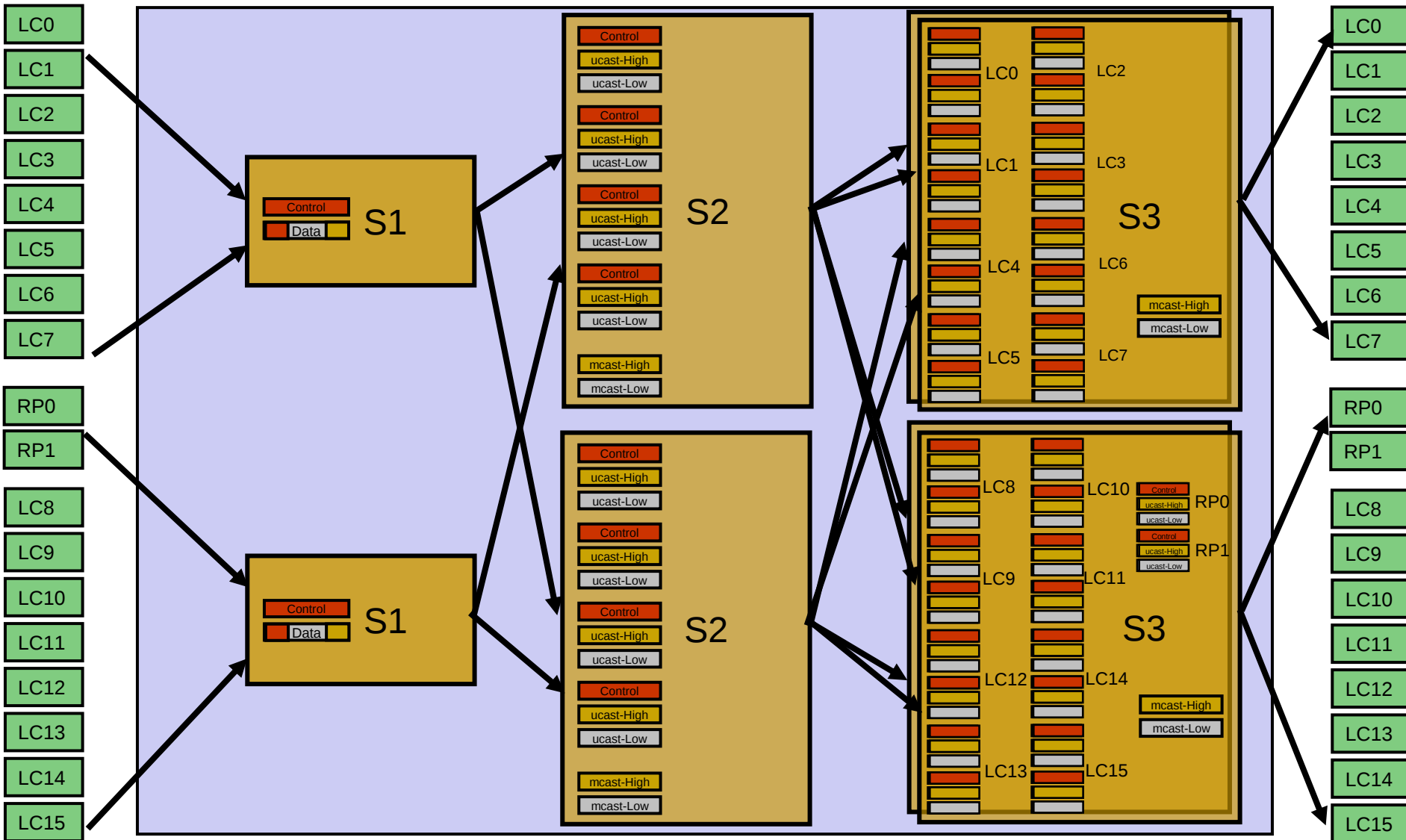


Full View of 1 Plane



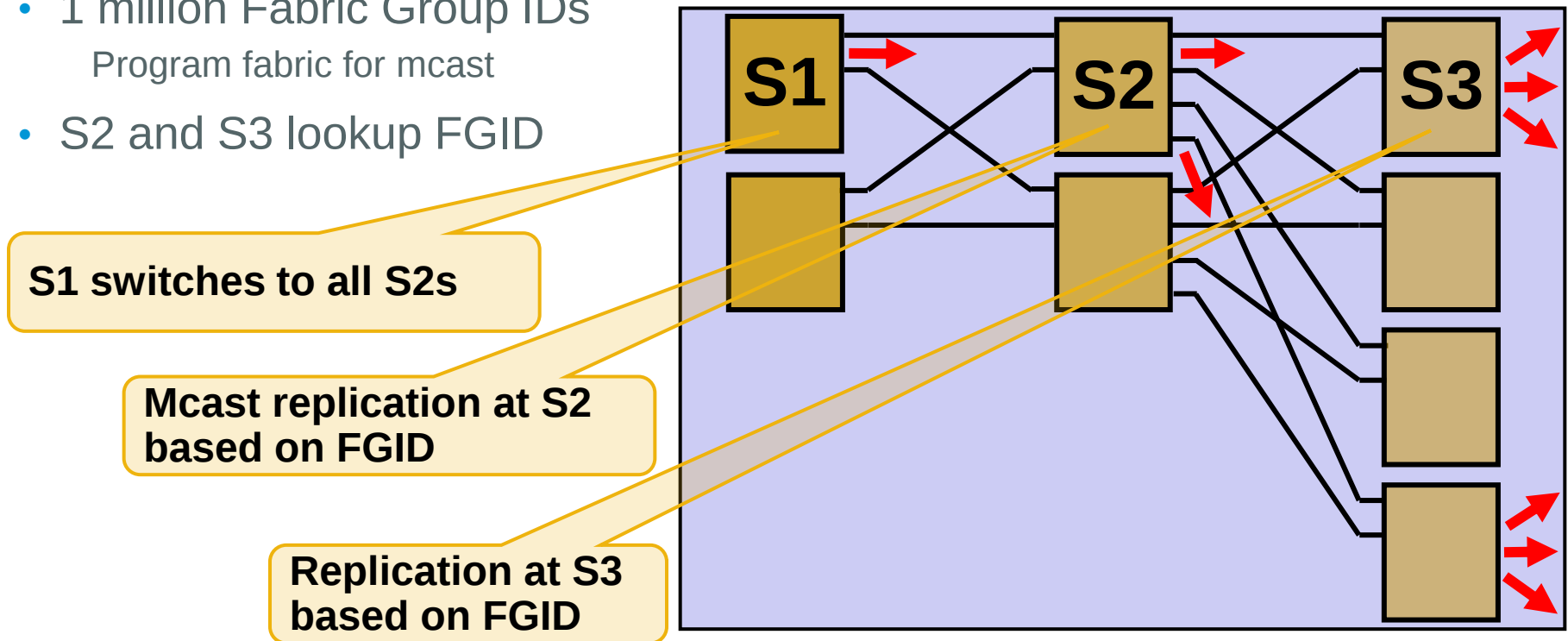
Queuing Inside the Fabric

Each destination and HP/LP queued separately



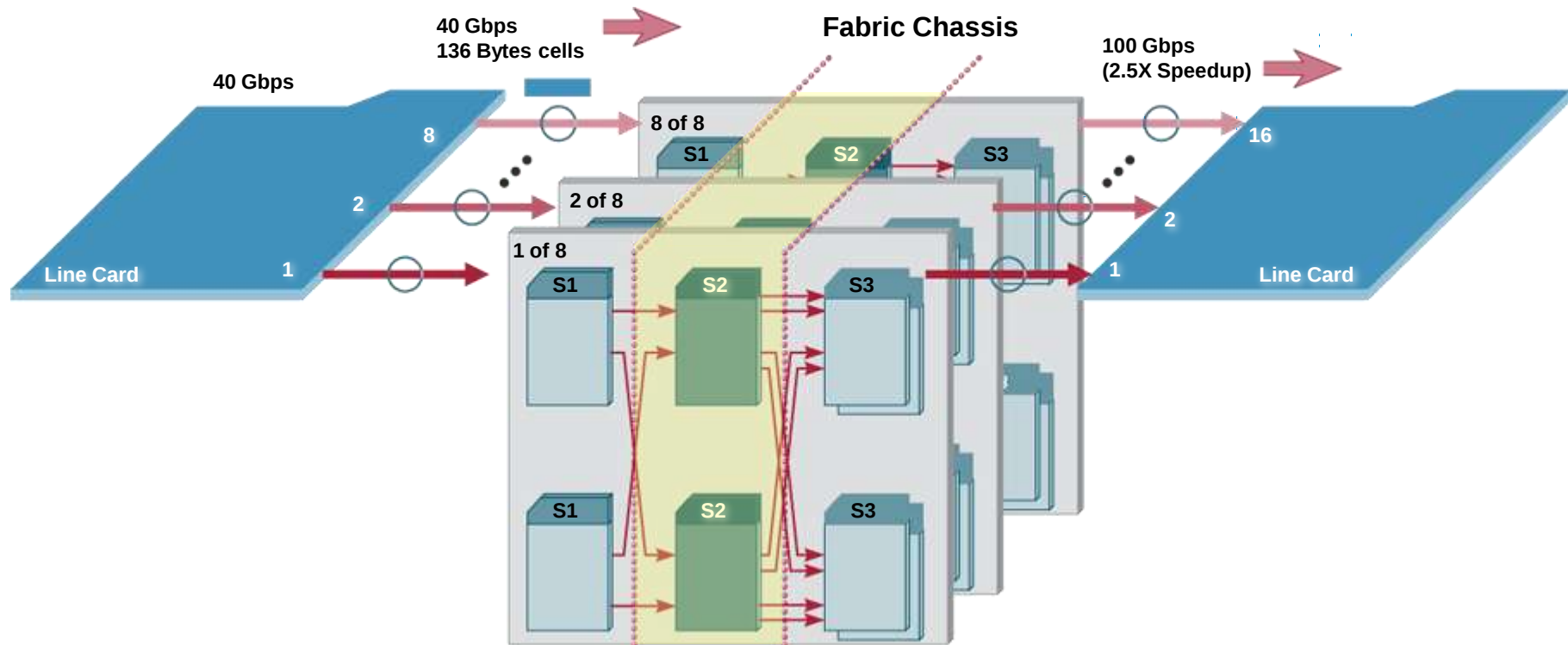
Switch Fabric Multicast Replication

- CRS-1 provides efficient multicast replication via 3 operations
 - S2 can replicate cells to registered (via FGID) S3 SEAs
 - S3 can replicate cells to registered (via FGID) FabricQs
 - Egress PSE can replicate packets for each output port
- 1 million Fabric Group IDs
 - Program fabric for mcast
- S2 and S3 lookup FGID



CRS Multi-chassis

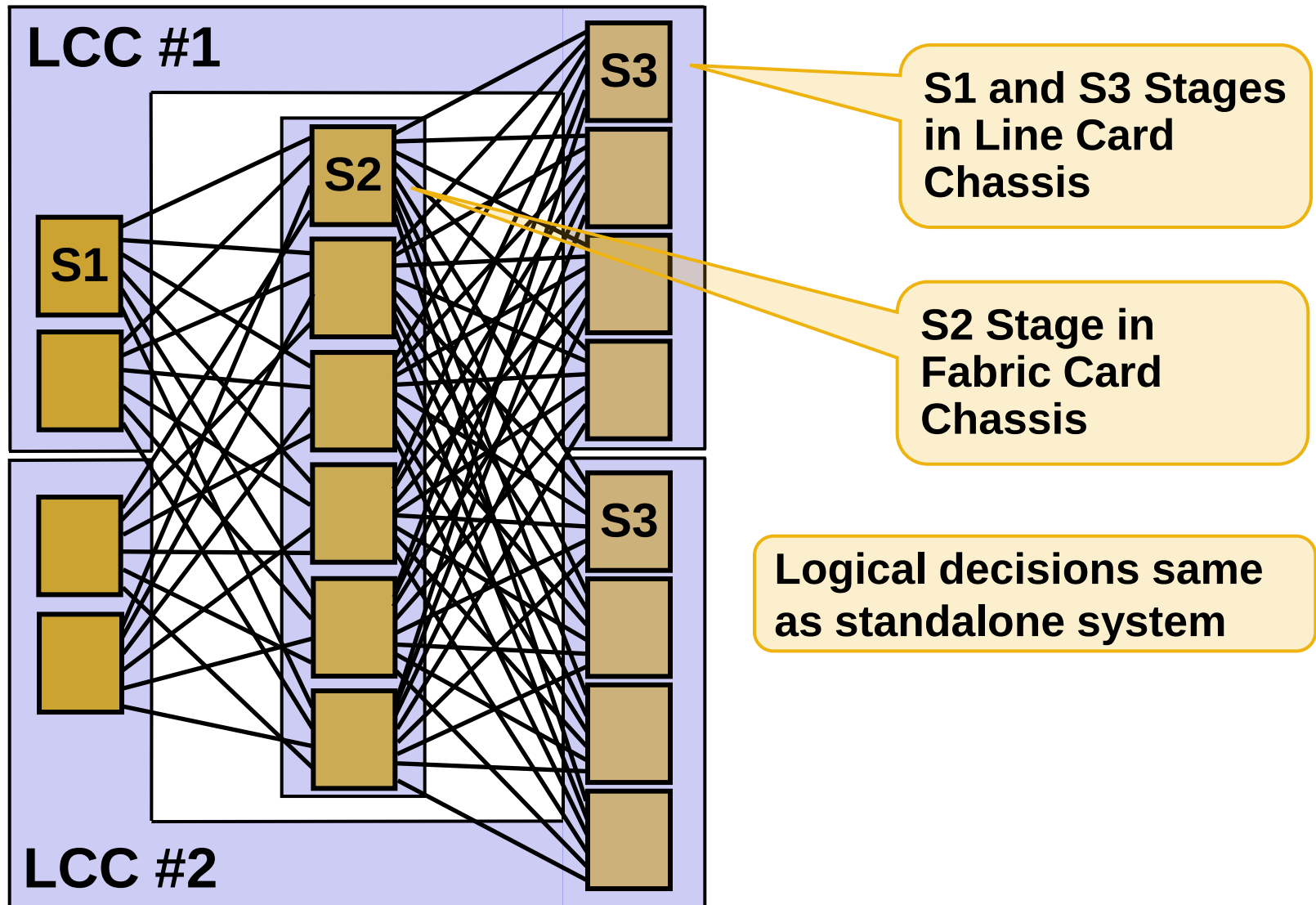
CRS-1 Multi-chassis



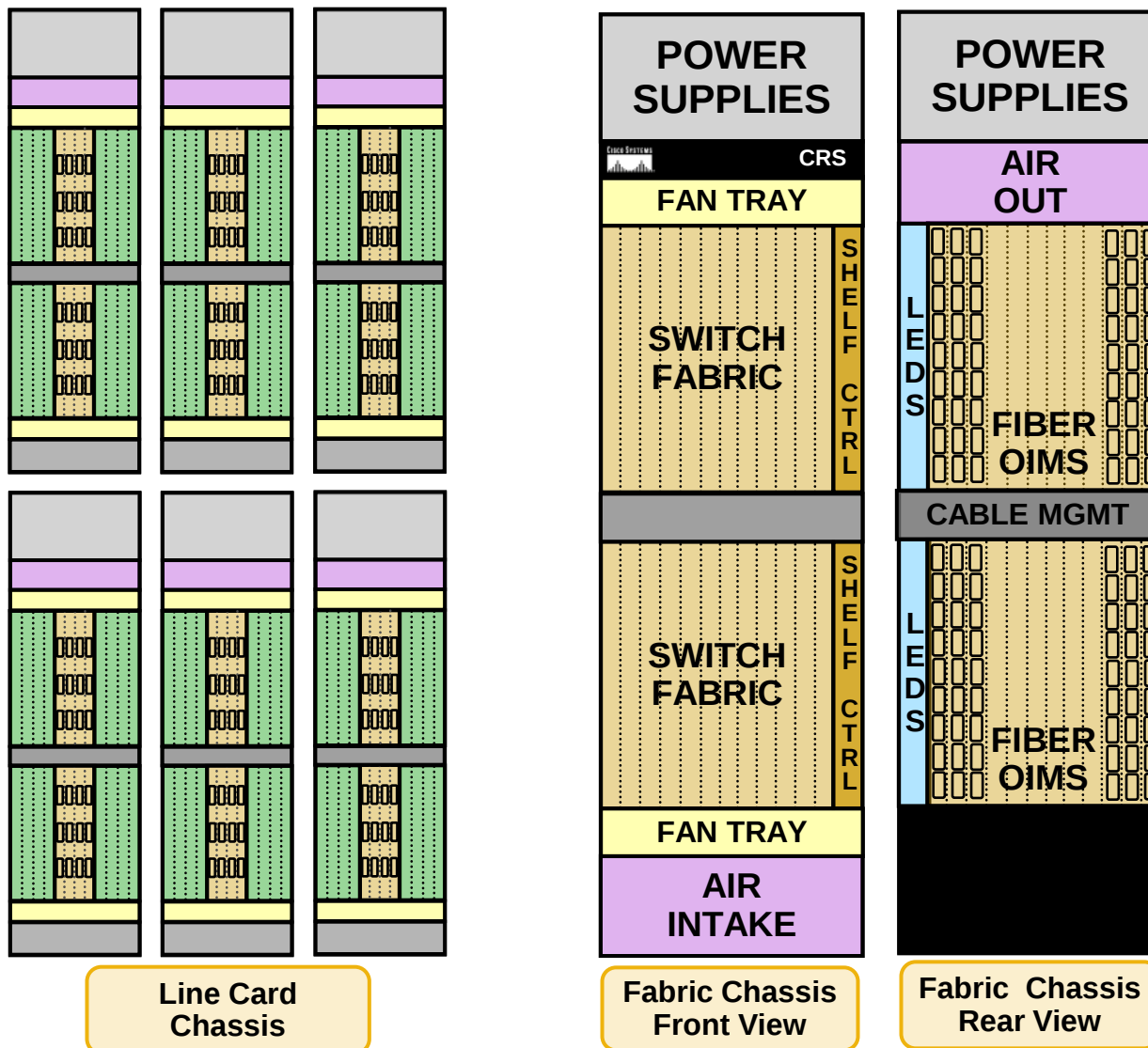
S2 stage moves into Fabric Card Chassis (FCC)
Logical operation of fabric remains the same

Decisions at Each Stage in 2+1 Multi-chassis

For 1 the 8 planes



CRS Multi-chassis Components



CRS Multi-chassis

- Single to Multi-chassis upgrade w/o packet loss

Moves fabric stage 2 to separate chassis

Upgrade LCC fabric cards

Add fabric card chassis (FCC)

Connect control GE

Connect LCC and FCC with fiber bundles

FCC Optical Connections



Optical Array Cable (100m)



72 Fiber Bundle

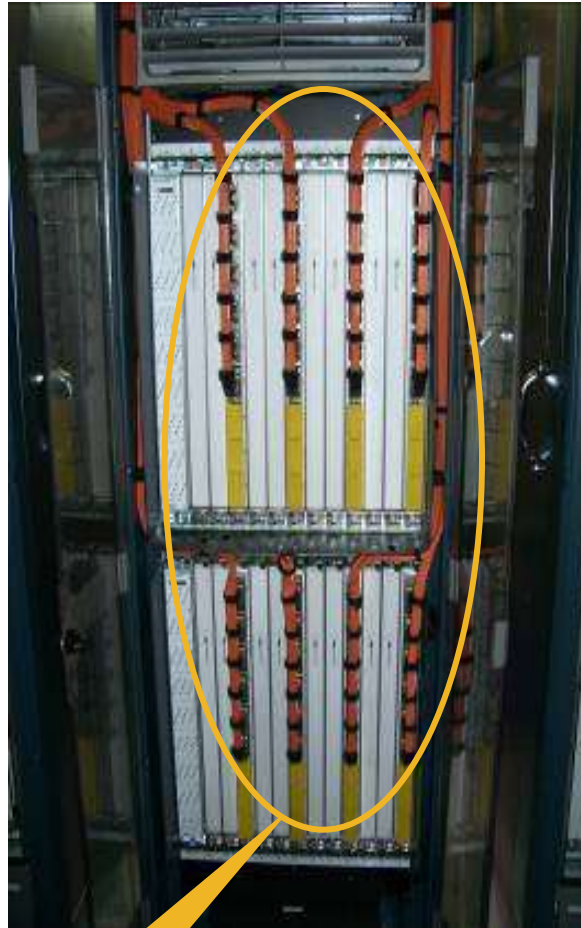


CRS Multi-chassis 2+1



LCC 0

8 S13
boards



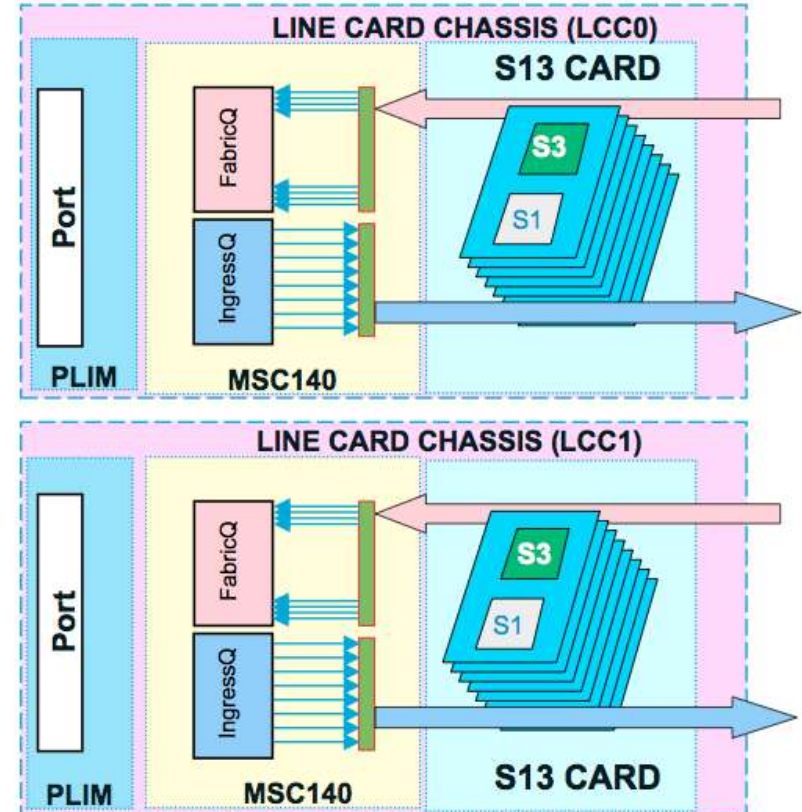
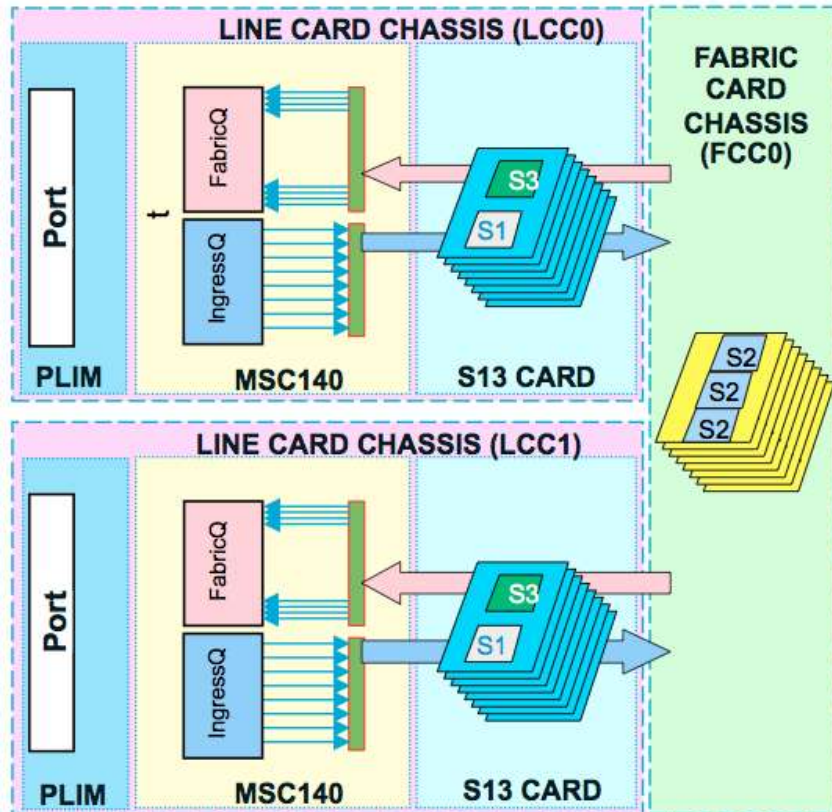
8 S2
boards
& OIMs

FCC

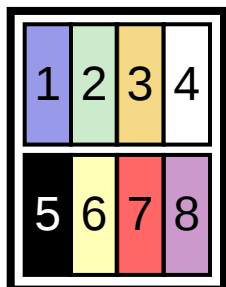


LCC 1

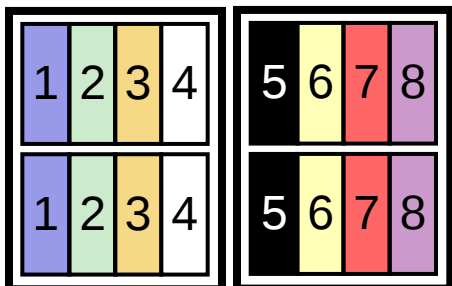
CRS 2+1 → 2+0 (B2B) Multi-chassis



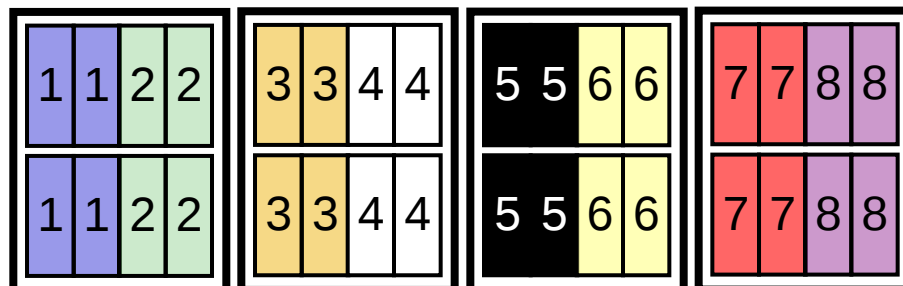
CRS Multi-chassis Fabric Plane Distribution



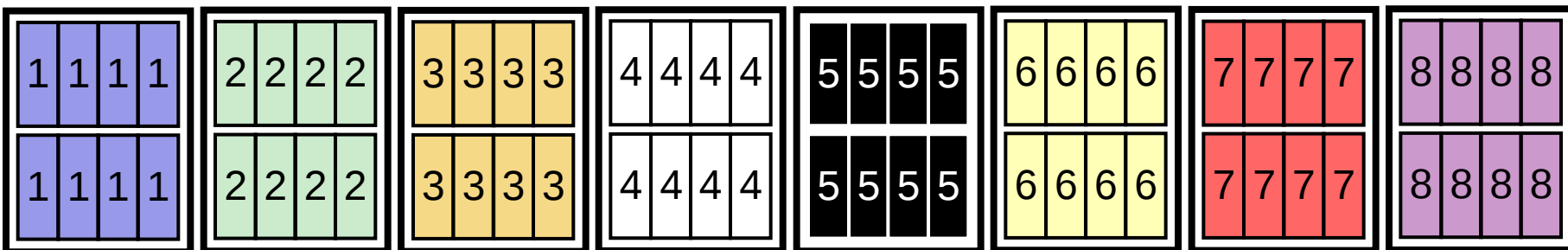
1 FCC
Max 9 LCC



2 FCC
Max 18 LCC



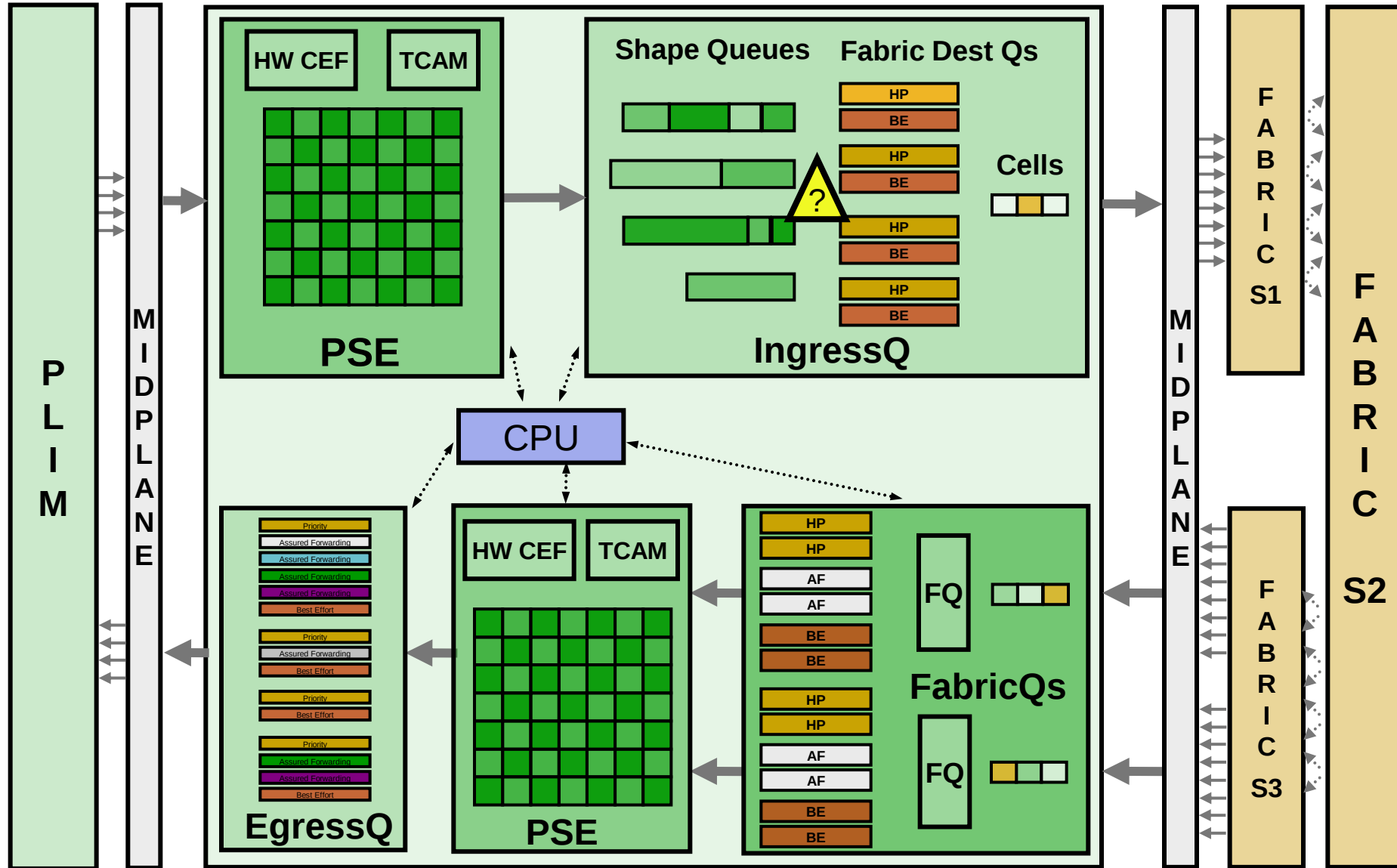
4 FCC
Max 36 LCC



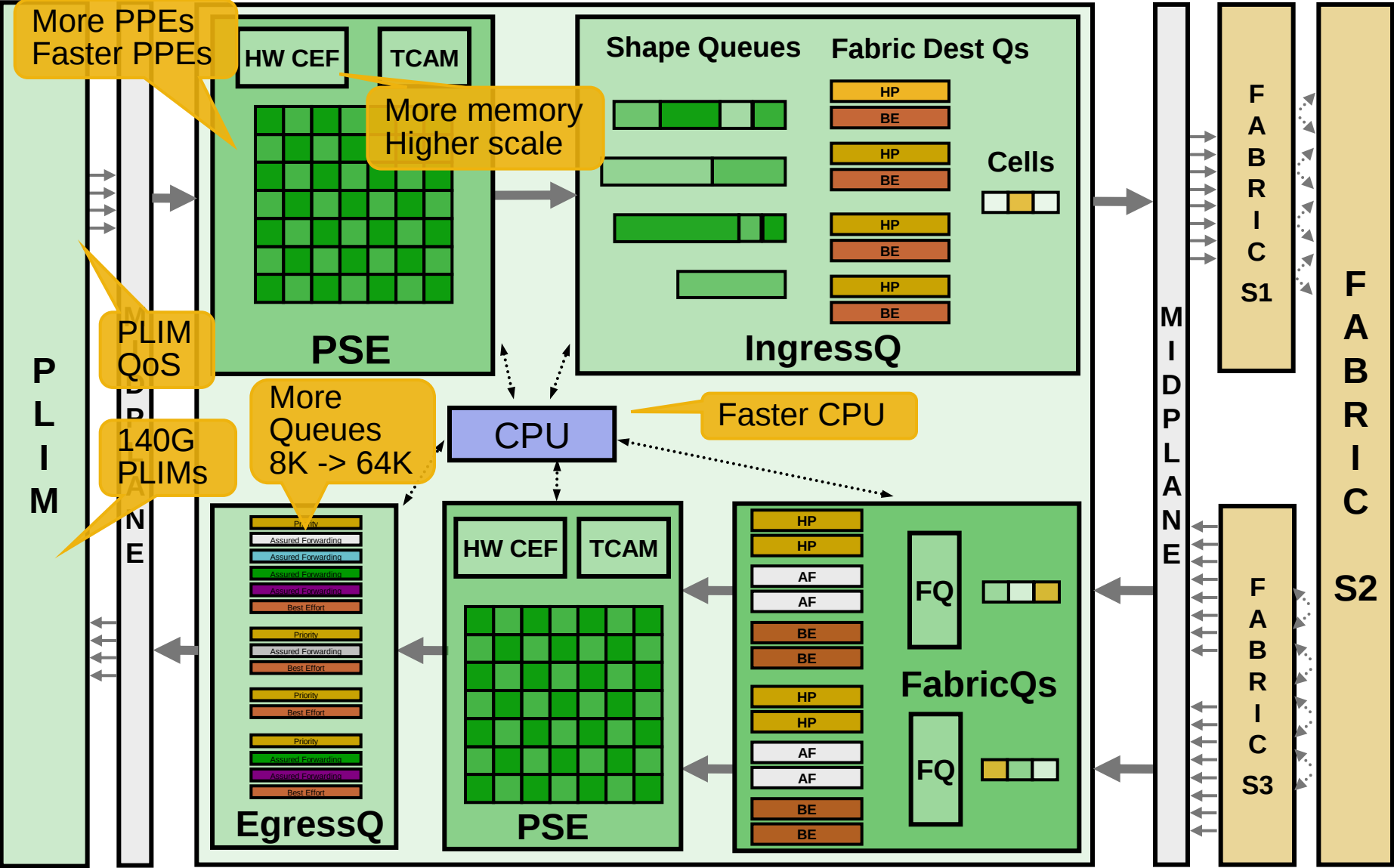
8 FCC
Max 72 LCC

CRS Summary

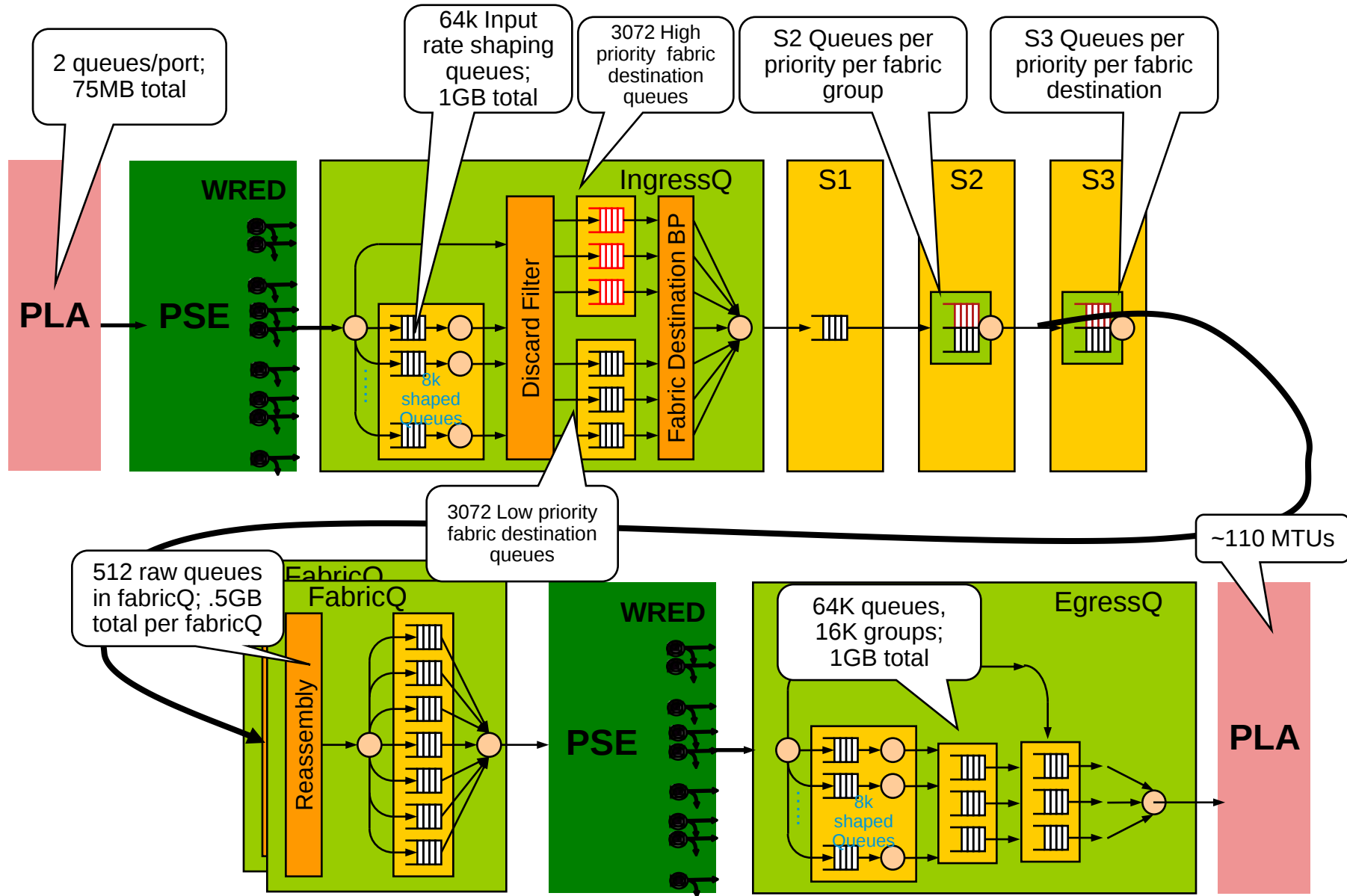
CRS Packet Path Review



CRS-3



CRS-3 Queuing



CRS Summary

#1 The Fastest Industry Leading Core Router

Carrier Class Architecture

- Fully Modular
- Scalable
- Reliable
- Predictable
- Backward compatible



Odkazy

www.cisco.com/go/crs

- **Data Sheets and Literature:**

http://www.cisco.com/en/US/products/ps5763/prod_literature.html

- **Support Documentation:**

http://www.cisco.com/en/US/products/ps5763/tsd_products_support_series_home.html

