State of the art Routers Architecture (Juniper, Cisco, Huawei)

October, 15th 2013

Maurizio Ipsale
JNCI, CCSI, HCSI

K labs

The ICT Knowledge factory

- High-quality training courses
- Real-life experience
- Theoretical knowledge
- Plenty of practical hands-on
- Best learning experience

K Labs is a company specialized in **Technical Trainings** for **Telco** and **ICT** market, particularly in the areas of **IP** data communications, **Networking**, **Optical Transport**, **Wireless**, **2G**, **3G** and **LTE** Access Network**, DSL**, **Radio Planning**, **Traffic Engineering**, **Security** and **Testing**

CISCO Learning Partner

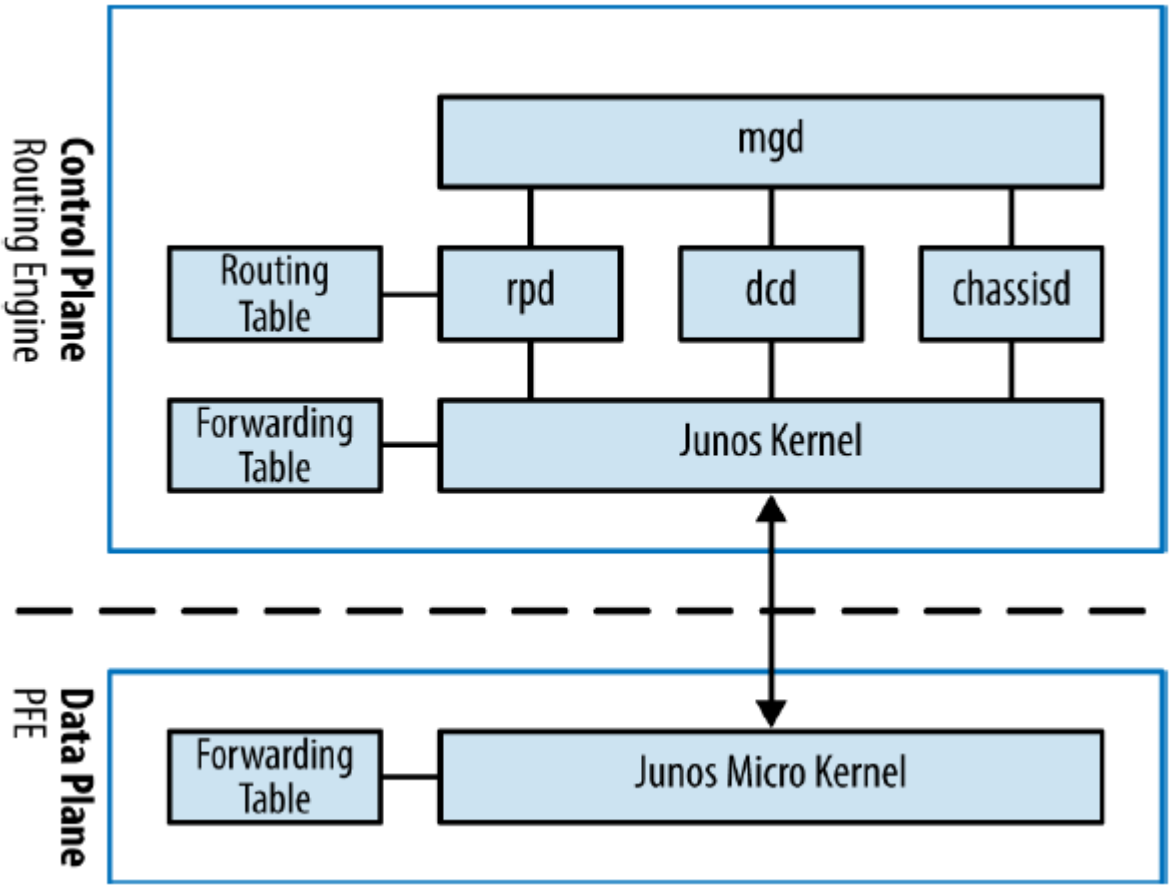
JUNIPER NETWORKS Authorized Education Partner

radware Authorized Training Center

HUAWEI Authorized Learning Partner

State of the art Routers Architecture (Juniper, Cisco, Huawei)

Junos Software Architecture

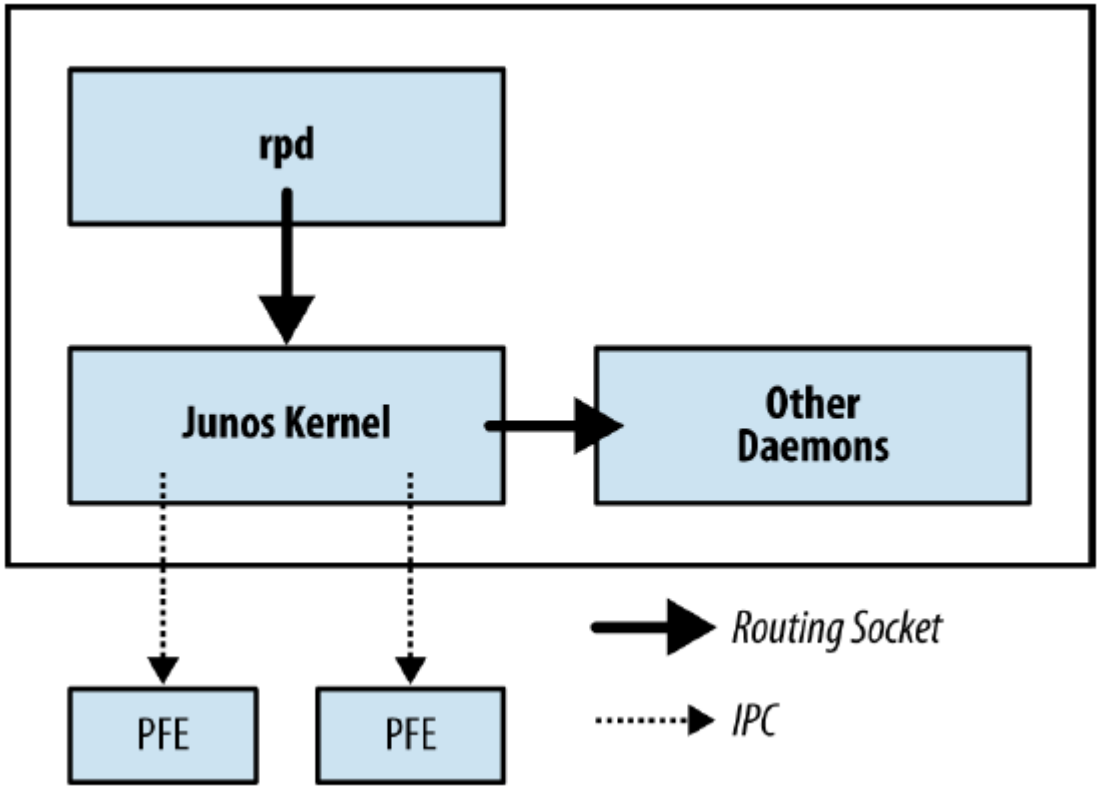


- Management daemon (mgd)
- Routing protocol daemon (rpd)
- Device control daemon (dcd)
- Chassis daemon (chassisd)

References:
1. Douglas Richard Hanks Jr., Harry Reynolds, Juniper MX Series, O'Reilly Media, Oct 2012
2. Antonio Sánchez-Monge, This Week: A Packet Walkthrough on the M, MX, AND T Series, Juniper Networks Books, Jan 2013

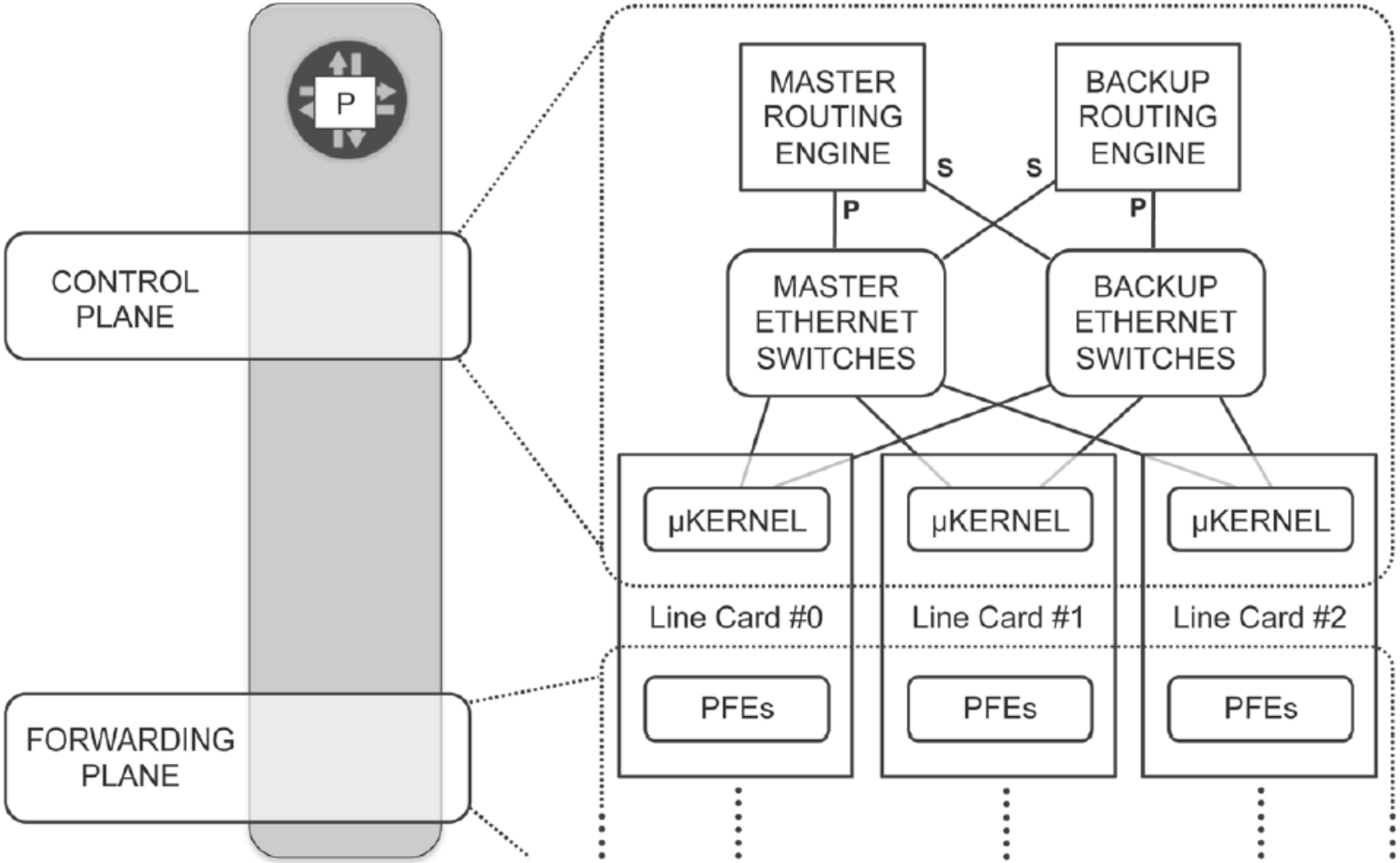
Courtesy of Juniper Networks Education Services

Routing Sockets



Routing sockets are a UNIX mechanism for controlling the routing table. The Junos kernel takes this same mechanism and extends it to include additional information to support additional attributes to create a carrier-class network operating system.

Control Plane



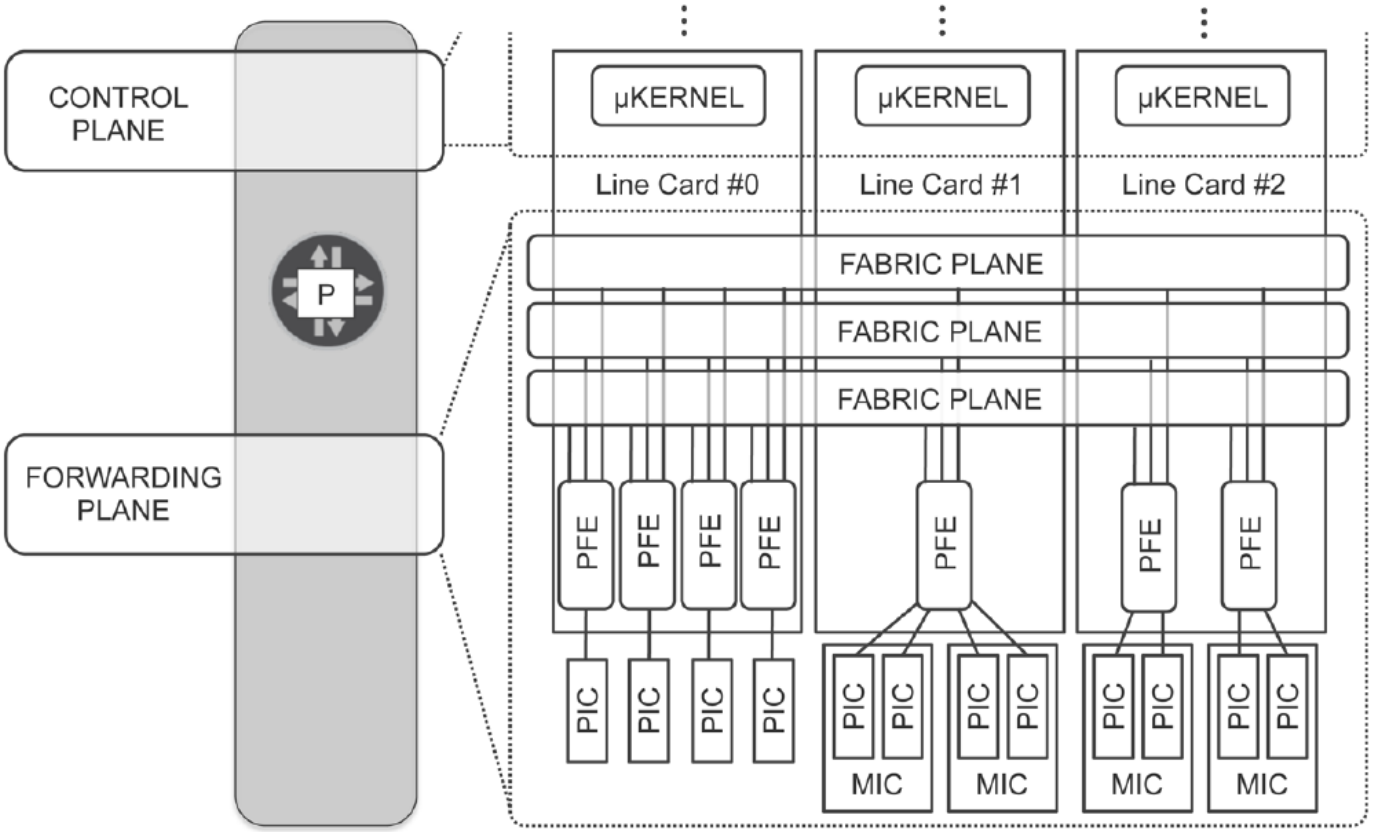
Courtesy of Juniper Networks Education Services

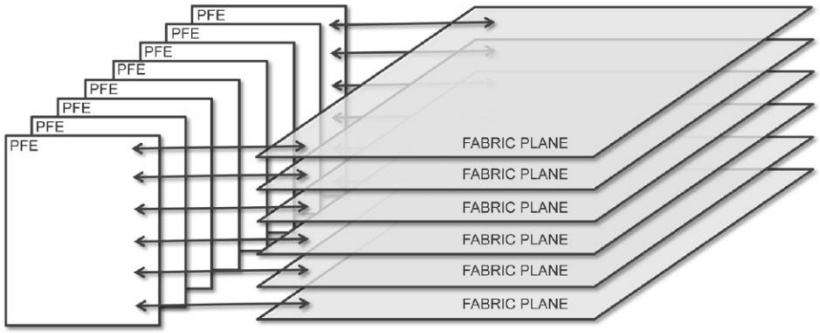
Forwarding Plane

Architecture Details

of the Forwarding Plane

in Multi-PFE Platforms



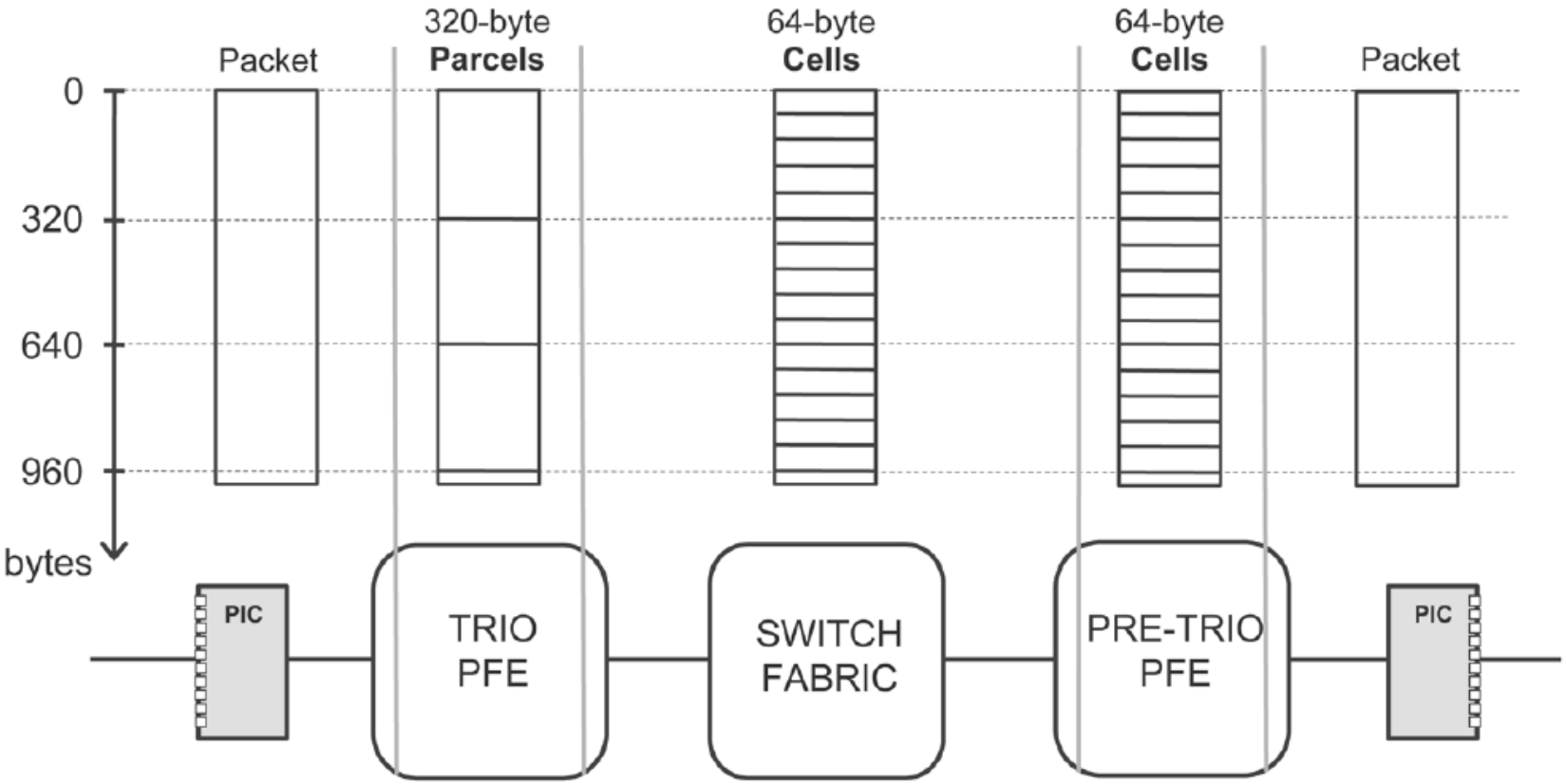


Model of PFE Interconnection

with Fabric Planes

Courtesy of Juniper Networks Education Services

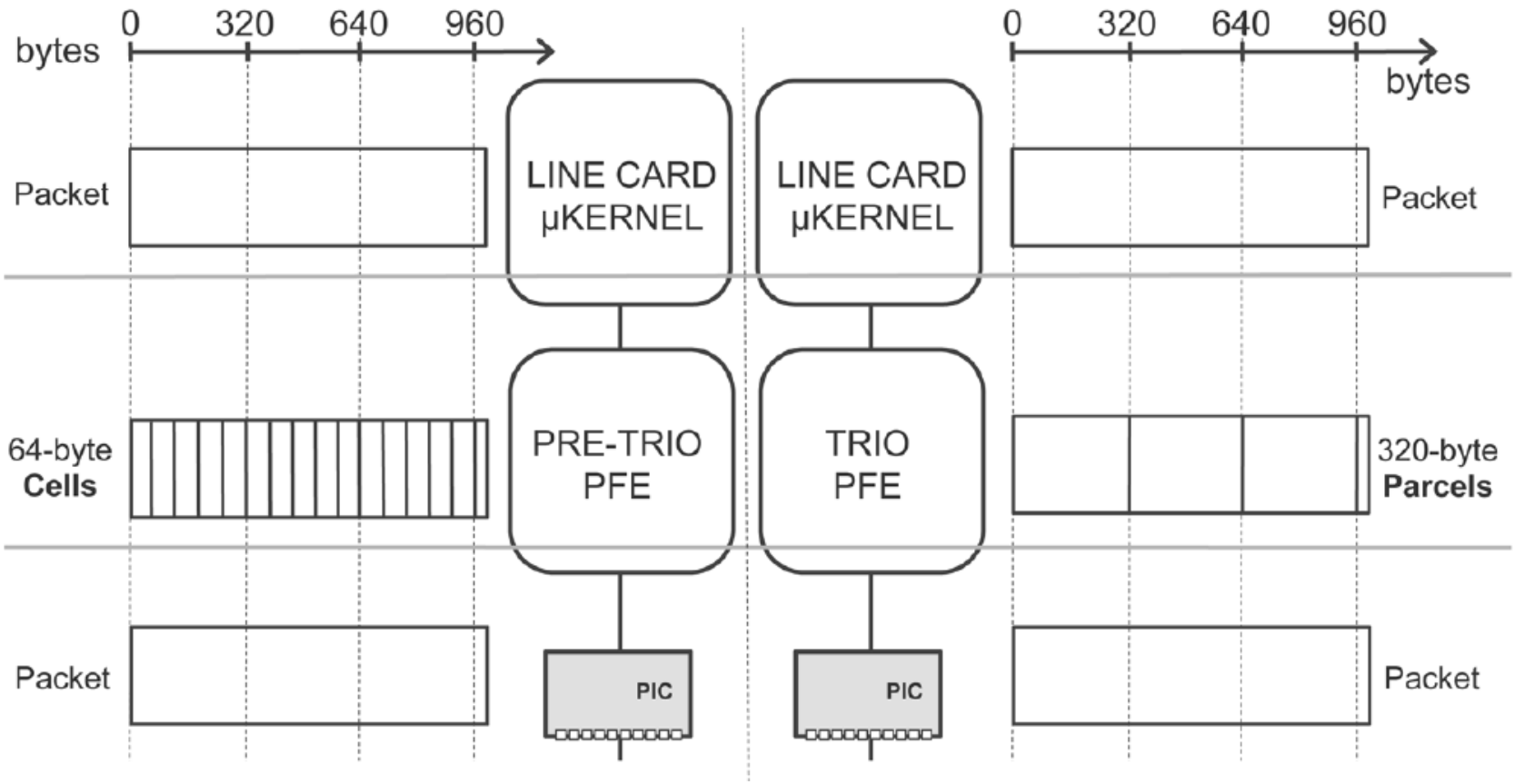
Internal Processing of Transit Packets



INTERNAL OVERHEAD ADDED TO THE CELLS/PACKETS NOT SHOWN FOR SIMPLICITY

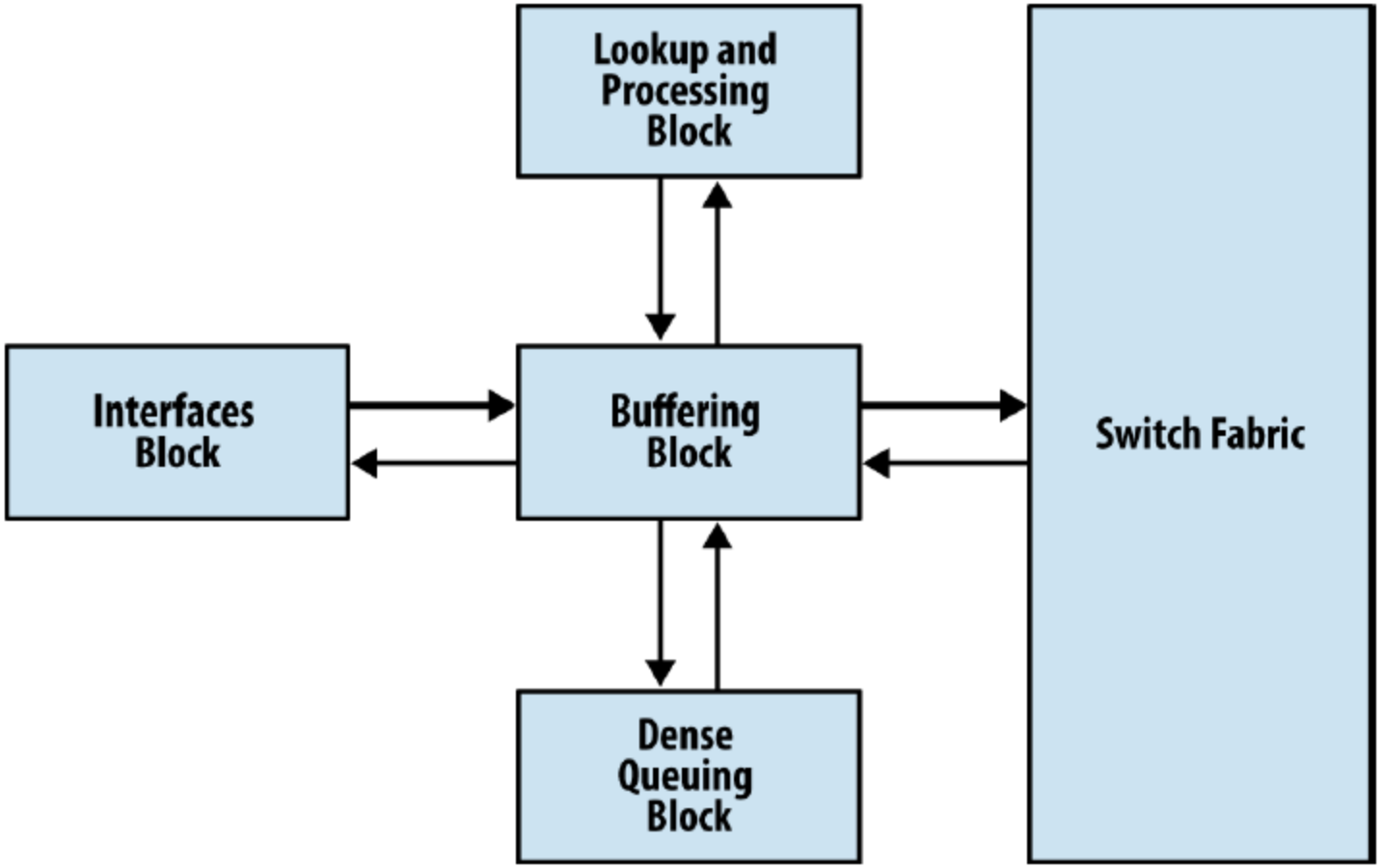
Courtesy of Juniper Networks Education Services

Internal Processing of Control and Exception Packets



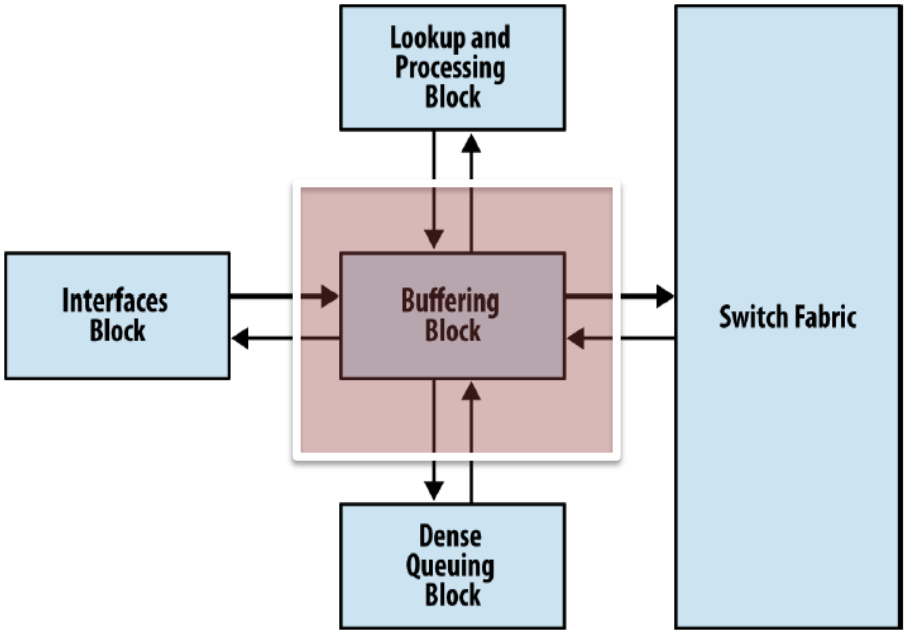
Courtesy of Juniper Networks Education Services

Trio Functional Block



Courtesy of Juniper Networks Education Services

Buffering Block

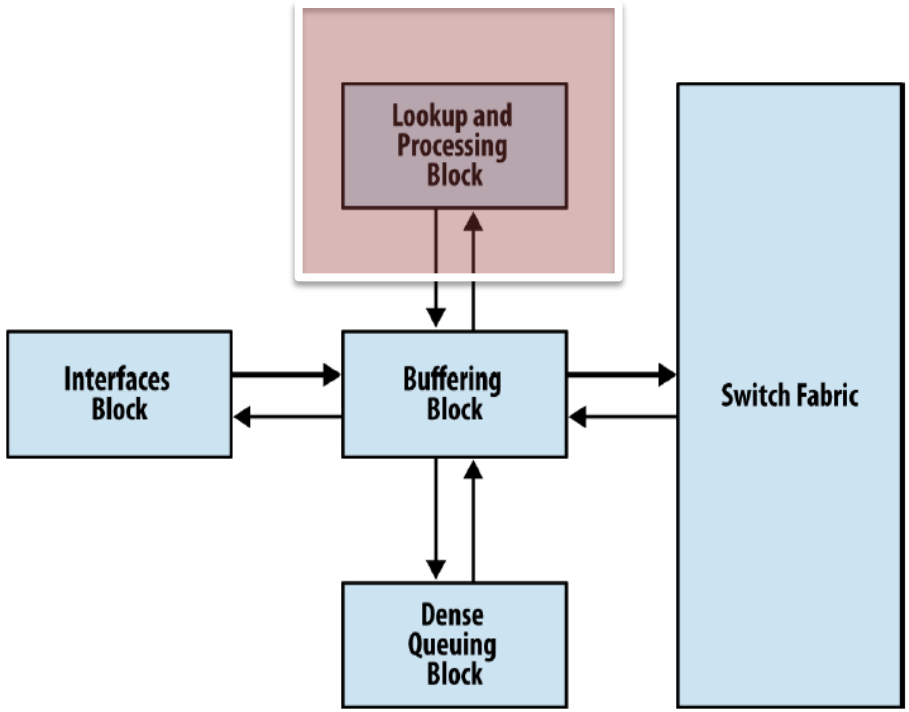


- Packet Data
- Fabric Queuing
- Revenue Port Queuing
- Delegate Responsibilities:
    - Process Oversubscription (<24x1GbE or <2x10GbE on a MIC)
    - Revenue Port Queuing (H-QoS)

Courtesy of Juniper Networks Education Services

Lookup Block

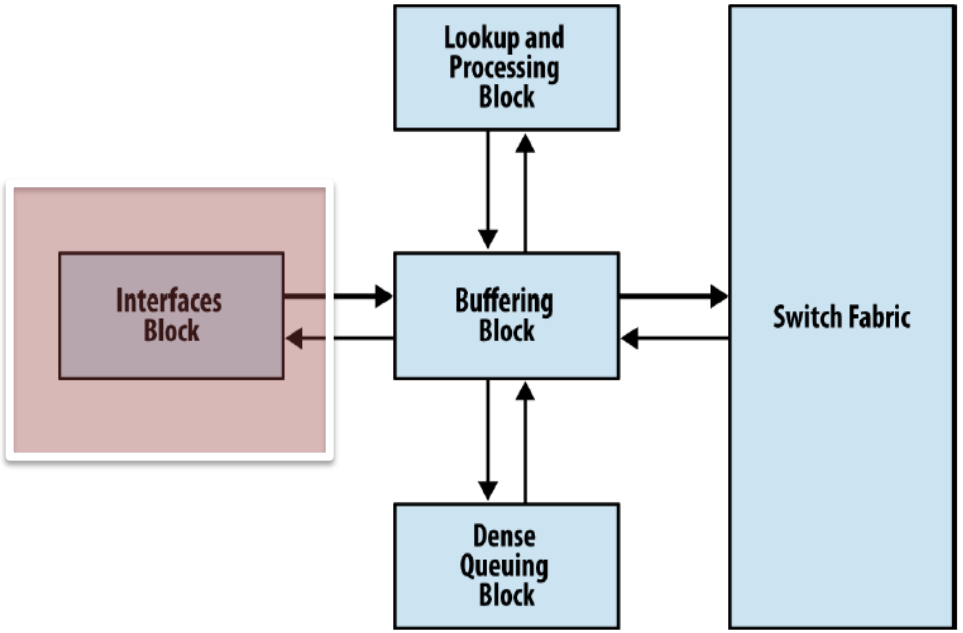
- Route lookups
- MAC lookups
- Class of Service (QoS) Classification
- Firewall filters
- Policers
- Accounting
- Encapsulation
- Statistics



•Deep Packet Inspection: 256 bytes into the packet

Courtesy of Juniper Networks Education Services

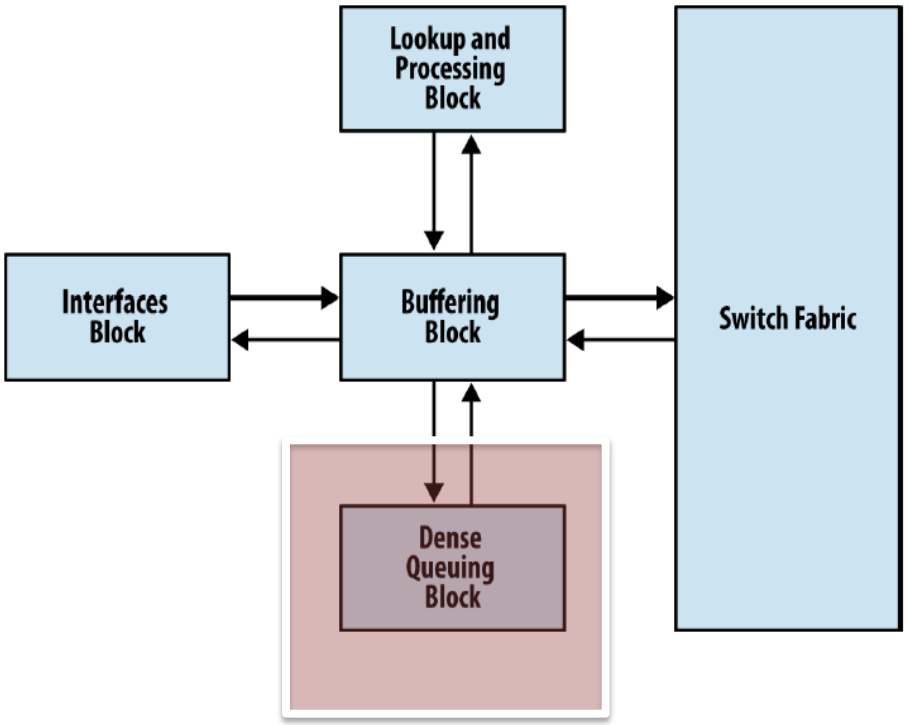
Interfaces Block



Each packet is inspected at line rate, and attributes such as Ethernet Type Codes, Protocol, and other Layer 4 information are used to evaluate which buffers to enqueue the packet towards the Buffering Block. Preclassification allows the ability to drop excess packets as close to the source as possible, while allowing critical control plane packets through to the Buffering Block.

Courtesy of Juniper Networks Education Services

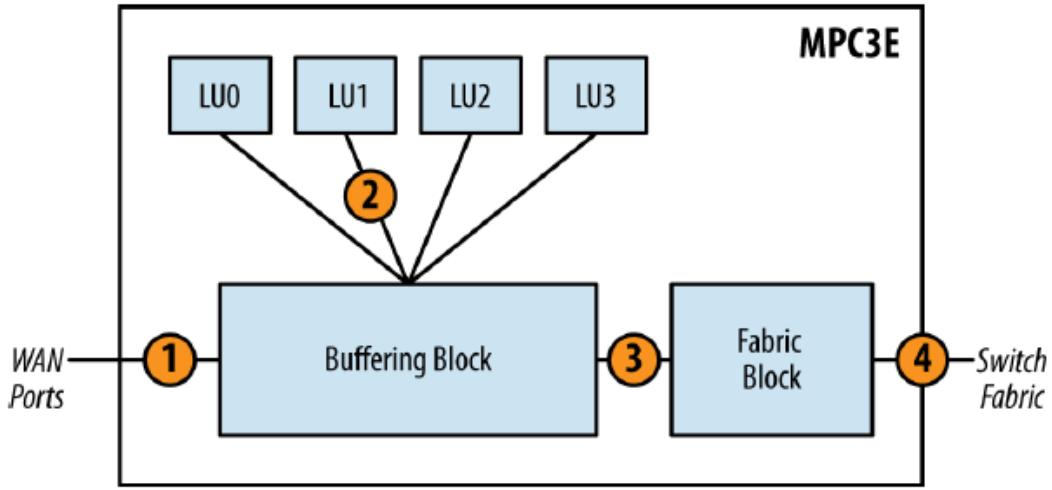
Dense Queuing Block



Depending on the line card, Trio offers an optional Dense Queuing Block that offers rich Hierarchical QoS that supports up to 512,000 queues with the current generation of hardware. This allows for the creation of schedulers that define drop characteristics, transmission rate, and buffering that can be controlled separately and applied at multiple levels of hierarchy.

Courtesy of Juniper Networks Education Services

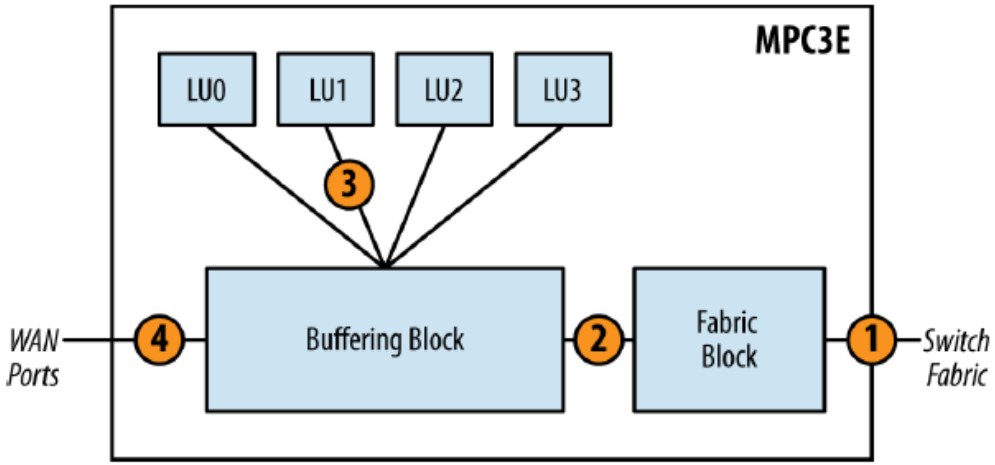
MPC3E Packet WalkThrough: Ingress



1. The packet enters the Buffering Block from the WAN ports and is subject to pre-classification. Depending on the type of packet, it will be marked as high or low priority. The Buffering Block will enqueue the packet as determined by the pre-classification at service the high-priority queue first. A Lookup Block is selected via round-robin and the packet is sent to that particular Lookup Block.
2. The packet enters the Lookup Block. A route lookup is performed and any services such as firewall filters, policing, statistics, and QoS classification are performed. The Lookup Block sends the packet back to the Buffering Block.
3. The packet is sent back to the Fabric Block and is enqueued into the switch fabric where it will be destined to another PFE. If the packet is destined to a WAN port within itself, it will simply be enqueued back to the Interfaces Block.
4. The packet is sent to the switch fabric.

Courtesy of Juniper Networks Education Services

MPC3E Packet WalkThrough: Egress

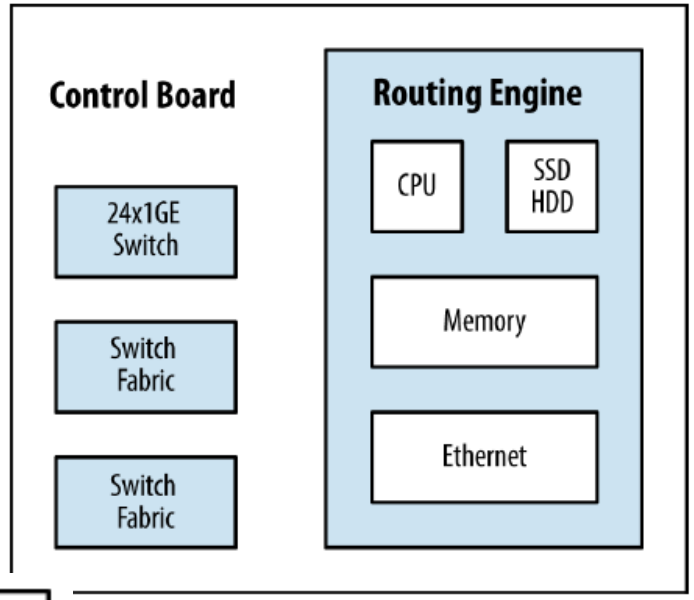


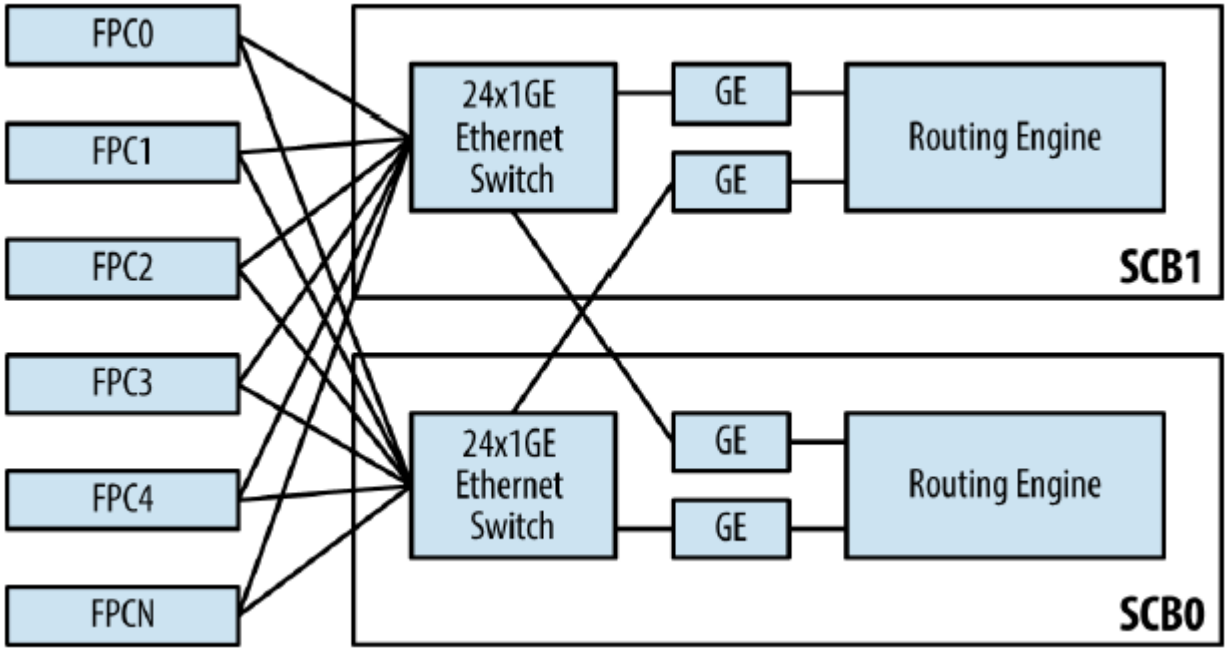
1. The packet is received from the switch fabric and sent to the Fabric Block. The Fabric Block sends the packet to the Buffering Block.
2. The packet enters the Buffering Block. The packet will then be subject to scheduling, shaping, and any other class os service as required. Packets will be enqueued as determined by the class of service configuration. The Buffering Block will then dequeue packets that are ready for transmission and send them to a Lookup Block selected via round-robin.
3. The packet enters the Lookup Block. A route lookup is performed as well as any services such as firewall filters, policing, statistics, and QoS classification. The Lookup Block sends the packet back to the Buffering Block.
4. The Buffering Block receives the packet and sends it to the WAN ports for transmission.

Courtesy of Juniper Networks Education Services

Switch and Control Board

```
{master}
dhanks@R1-RE0> show chassis ethernet-switch
Displaying summary for switch 0
Link is good on GE port 1 connected to device: FPC1
  Speed is 1000Mb
  Duplex is full
Link is good on GE port 2 connected to device: FPC2
  Speed is 1000Mb
  Duplex is full
```



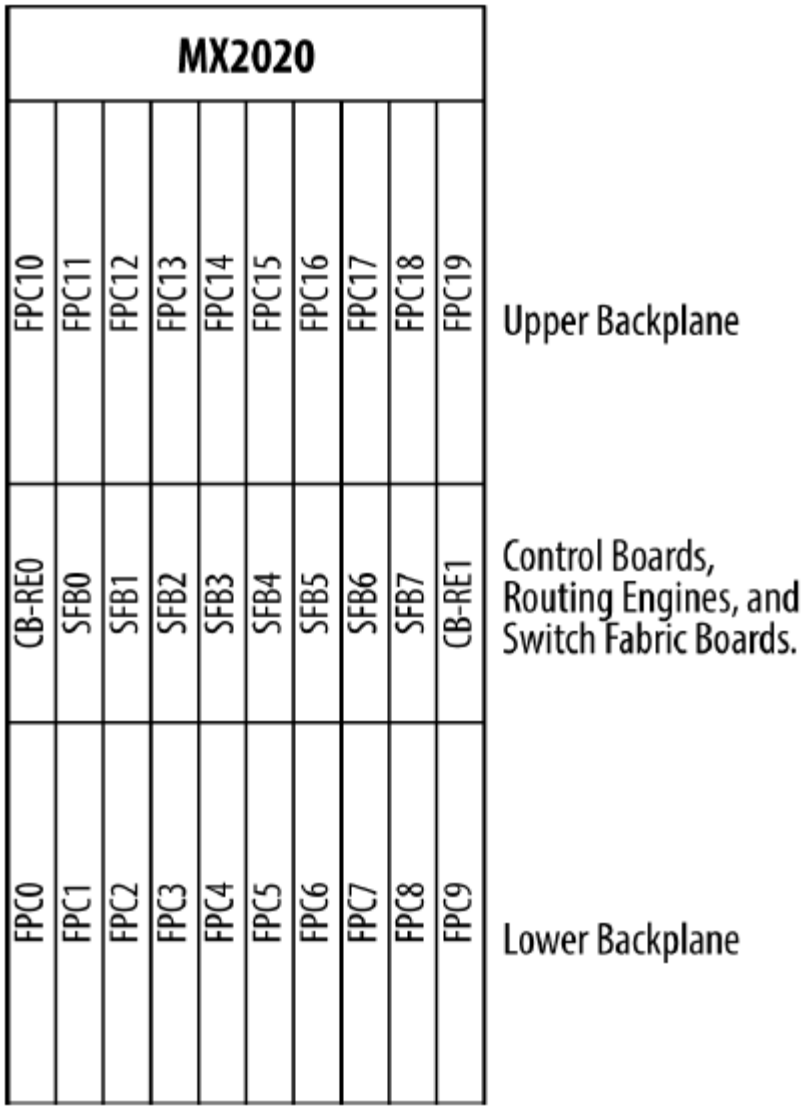


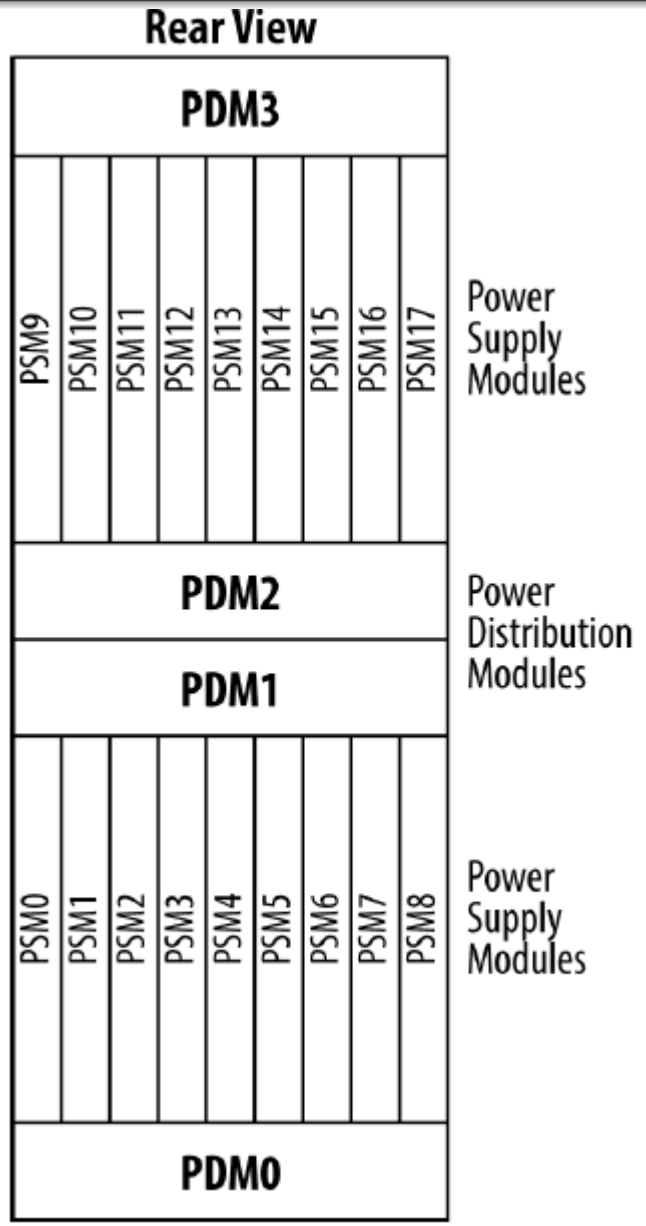

```
Link is good on GE port 12
connected to device: Other RE
  Speed is 1000Mb
  Duplex is full
```

```
Link is good on GE port 13
connected to device: RE-GigE
  Speed is 1000Mb
  Duplex is full
```

Courtesy of Juniper Networks Education Services

MX2020





Courtesy of Juniper Networks Education Services

State of the art Routers Architecture (Juniper, Cisco, Huawei)

Cisco ASR 1000 Building Blocks

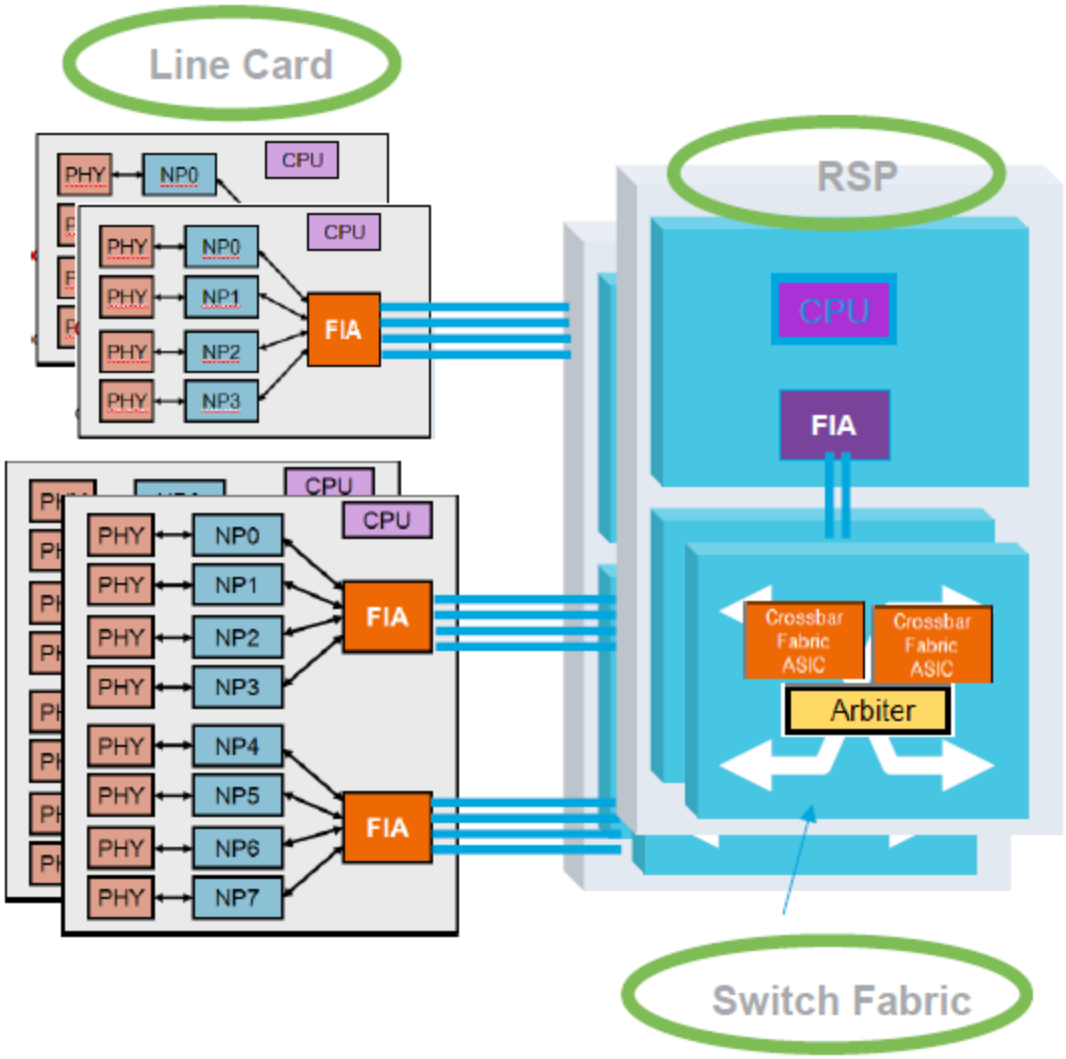
Fully Distributed Architecture for High Performance and High Multi-dimensional Control Plane Scale

- Data forwarding is fully distributed on the line cards
- Control plane split among RSP and LC CPU (same type of CPU as RSP)
- L2 protocols, BFD, CFM, Netflow runs on the LC CPU for high scale

True Modular OS for HA and Operational Simplicity

- Micro-kernel based, true modular OS
- High availability and System stability
- SW patch granularity for operational simplicity

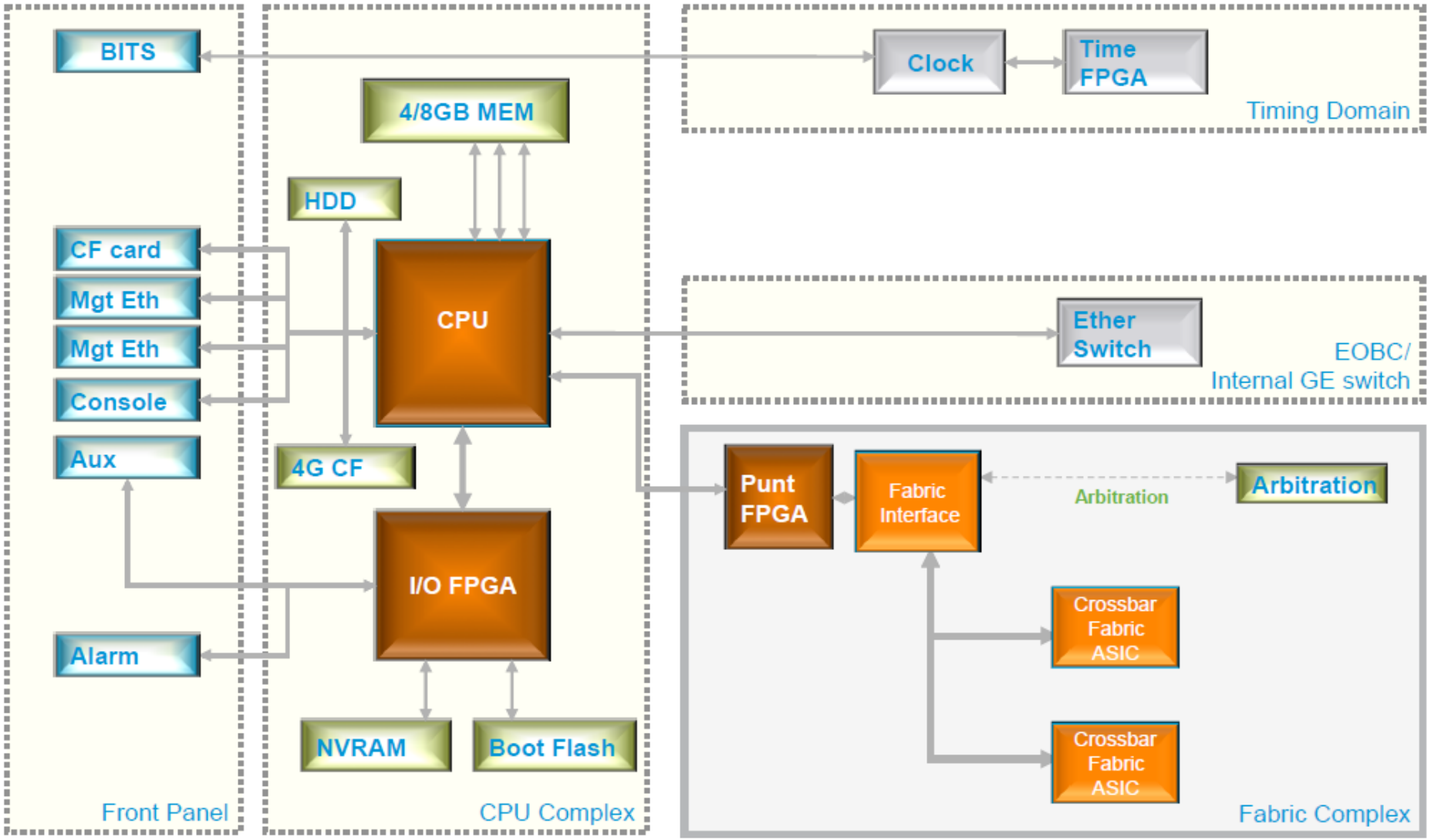
FIA = Fabric Interface ASIC



Active-Active Switch Fabric
Guarantee "0" packet loss during RSP failover

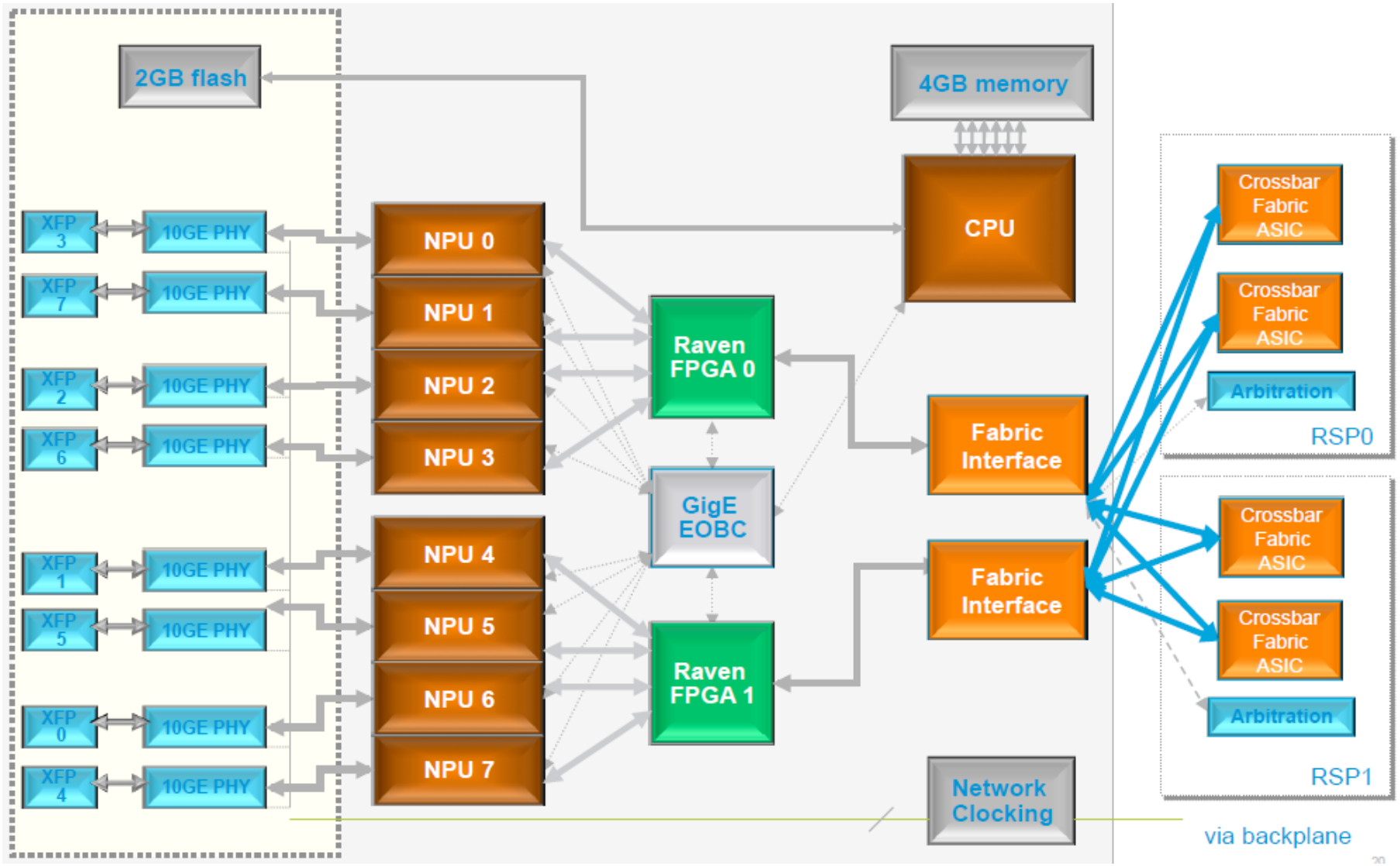
Courtesy of Cisco Education Services

Generic RSP Enging Architecture



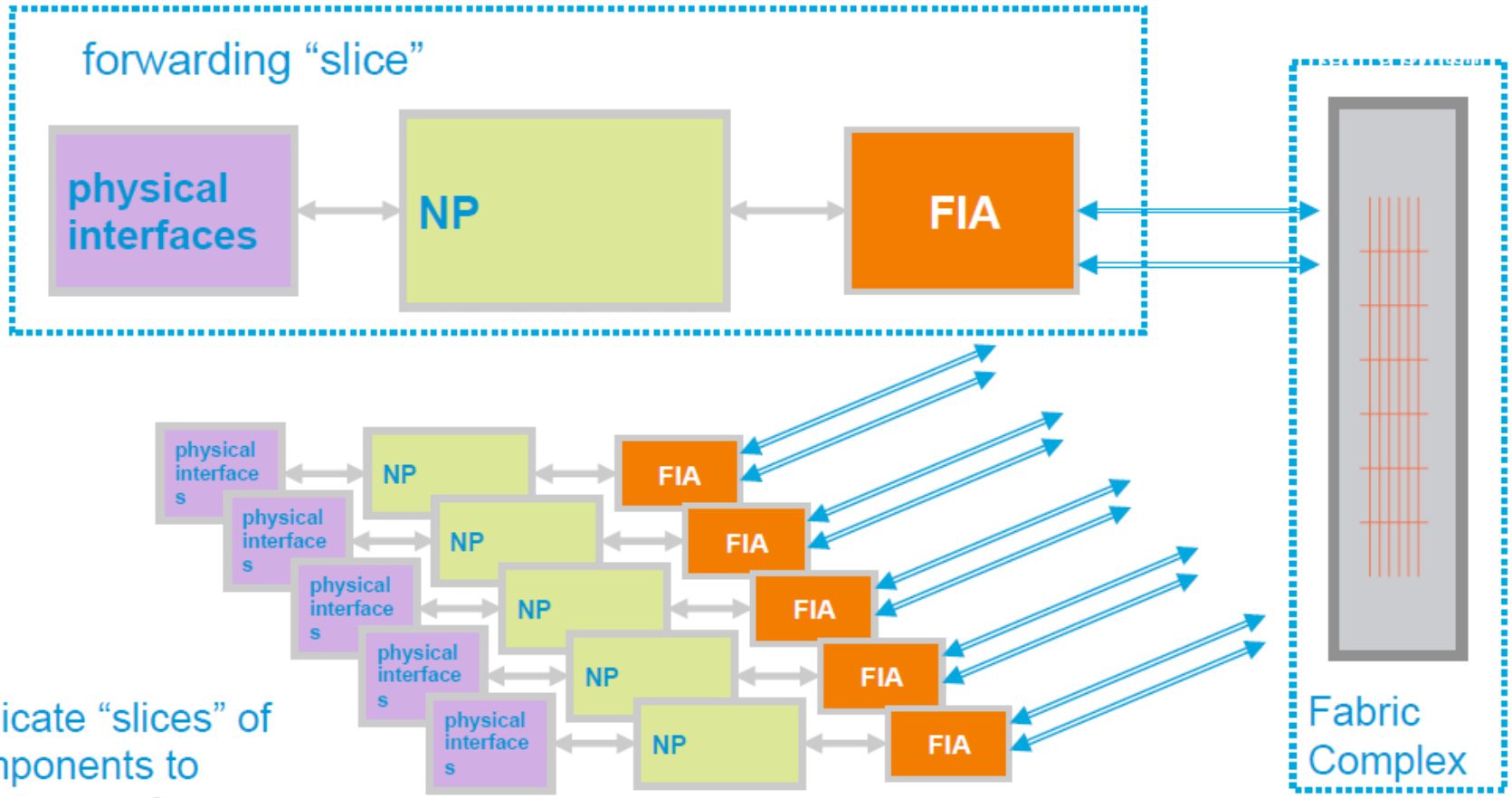
Courtesy of Cisco Education Services

80G Line Rate Line Card



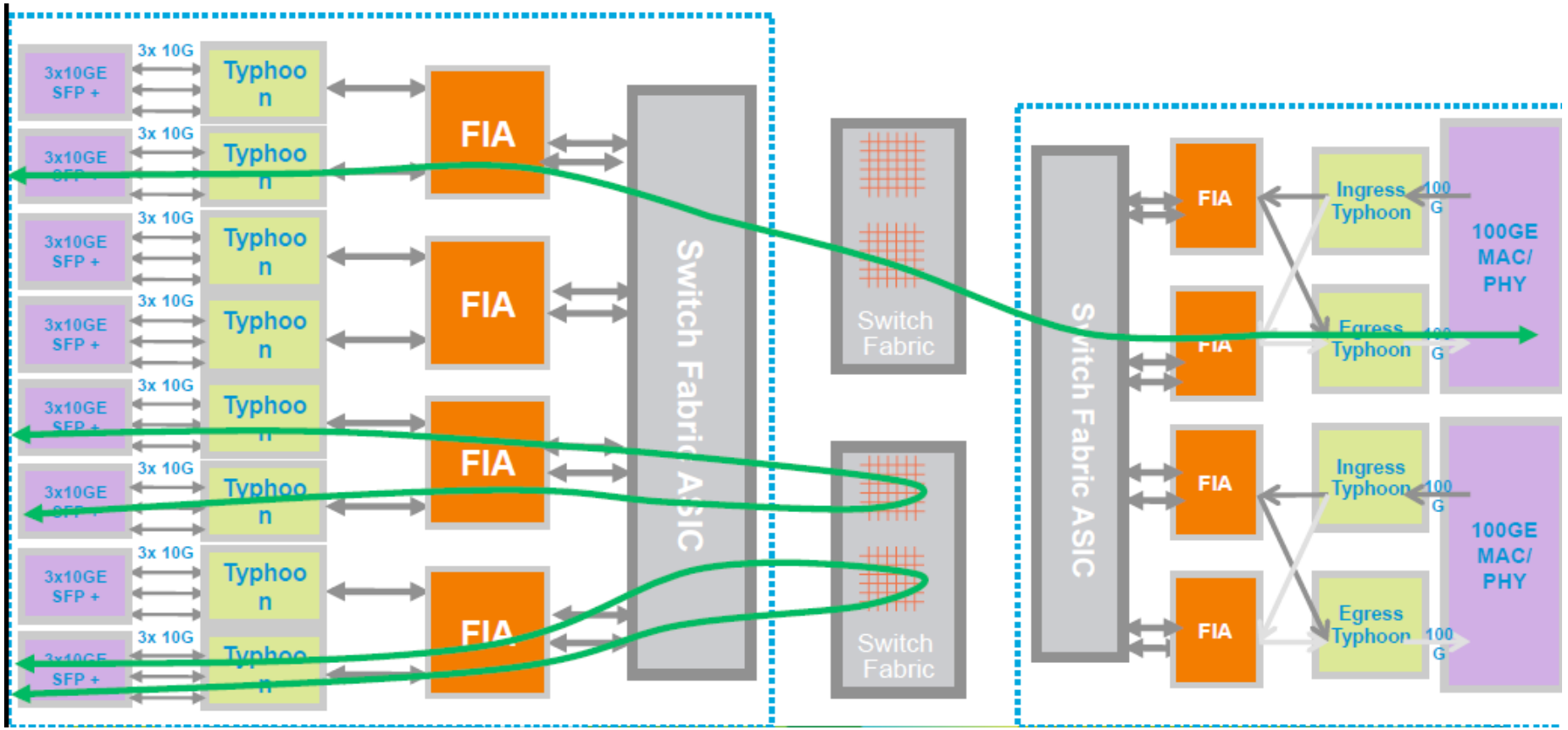
Courtesy of Cisco Education Services

Generic Line Card Architecture



Courtesy of Cisco Education Services

Packet Flow Overview

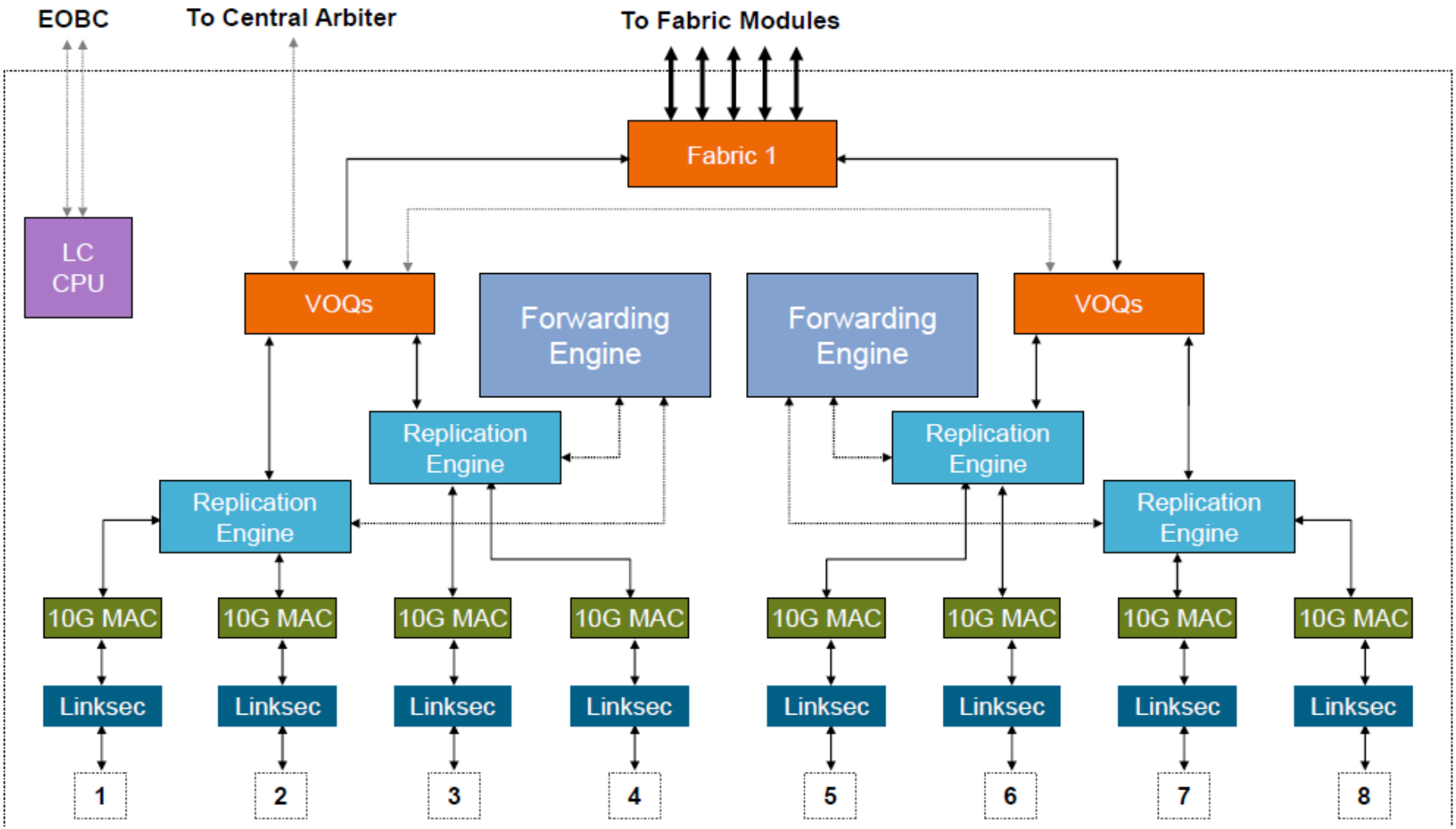


References:

1. Cisco Nexus 7000 Hardware Architecture, Cisco Live! 2012
2. Xander Thuijs, Understanding Cisco ASR 9000 Series Aggregation Services Routers Platform Architecture and Packet Forwarding Troubleshooting, Cisco Support Community Expert Series Webcast 2013

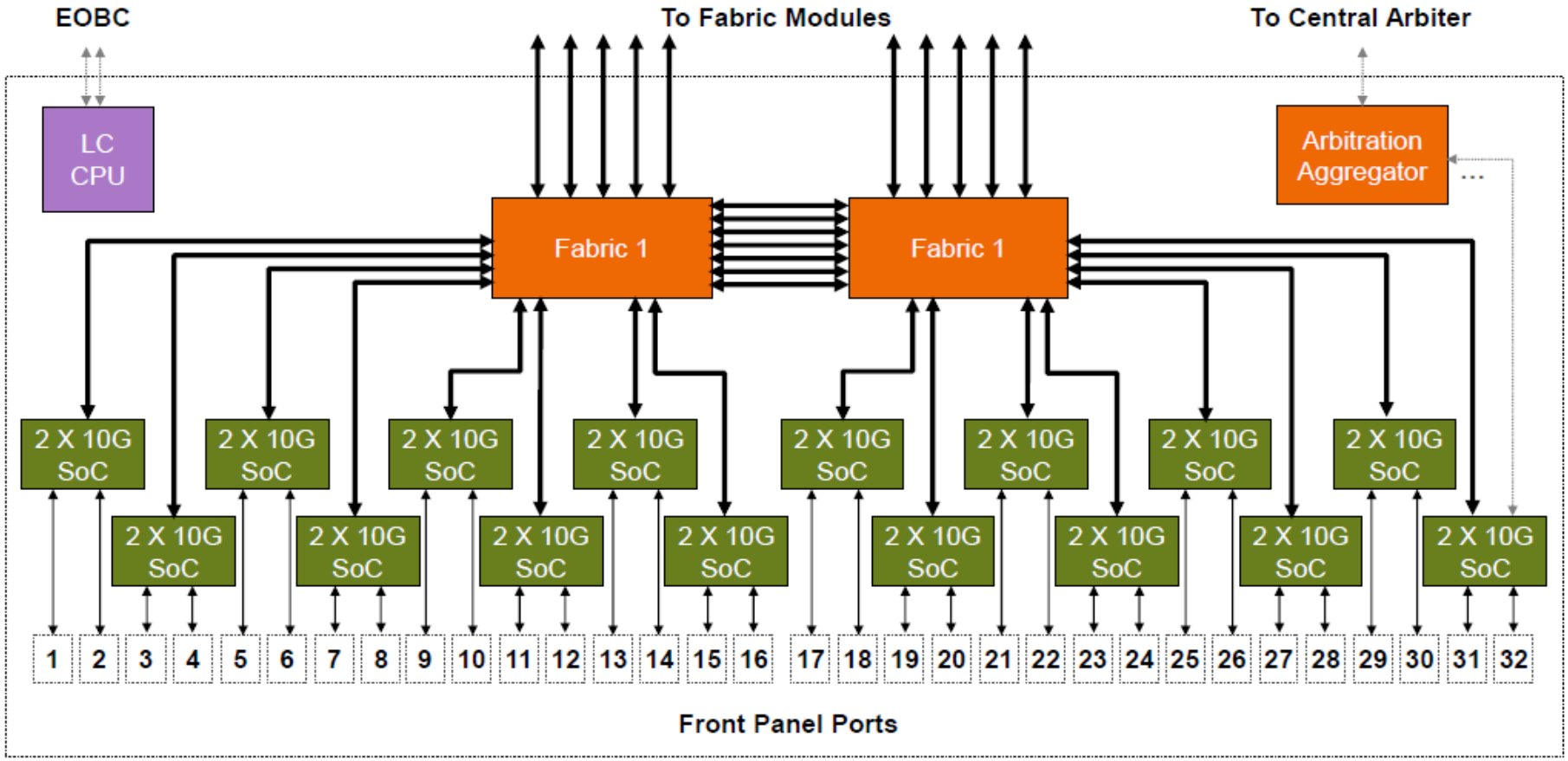
Courtesy of Cisco Education Services

Nexus 7000 M1 Architecture



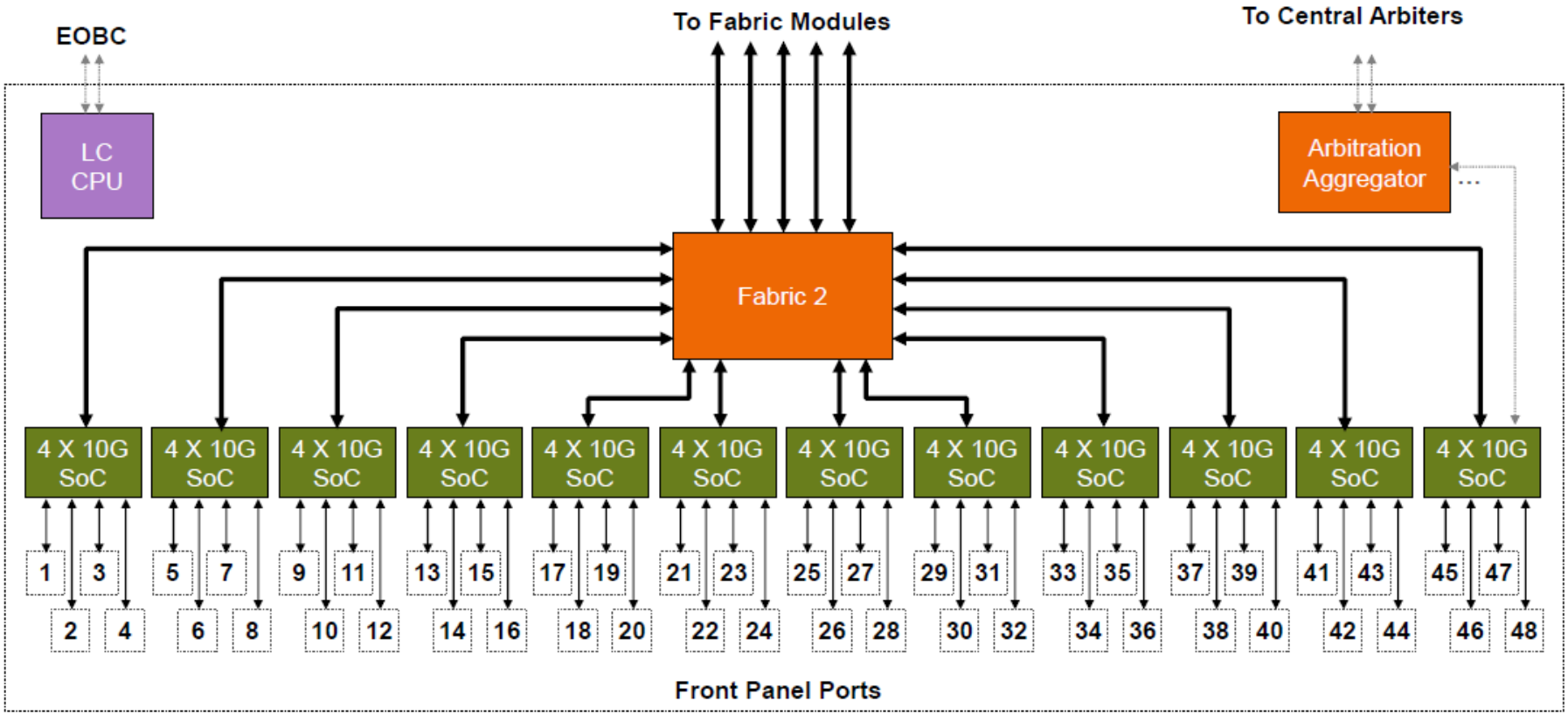
Courtesy of Cisco Education Services

Nexus 7000 F1 Architecture



Courtesy of Cisco Education Services

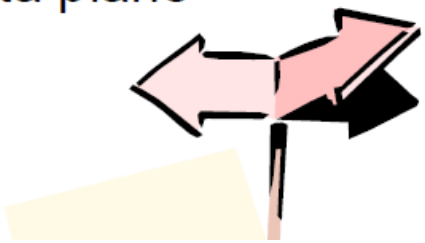
Nexus 7000 F2 Architecture



Courtesy of Cisco Education Services

Nexus 7000 IP Forwarding

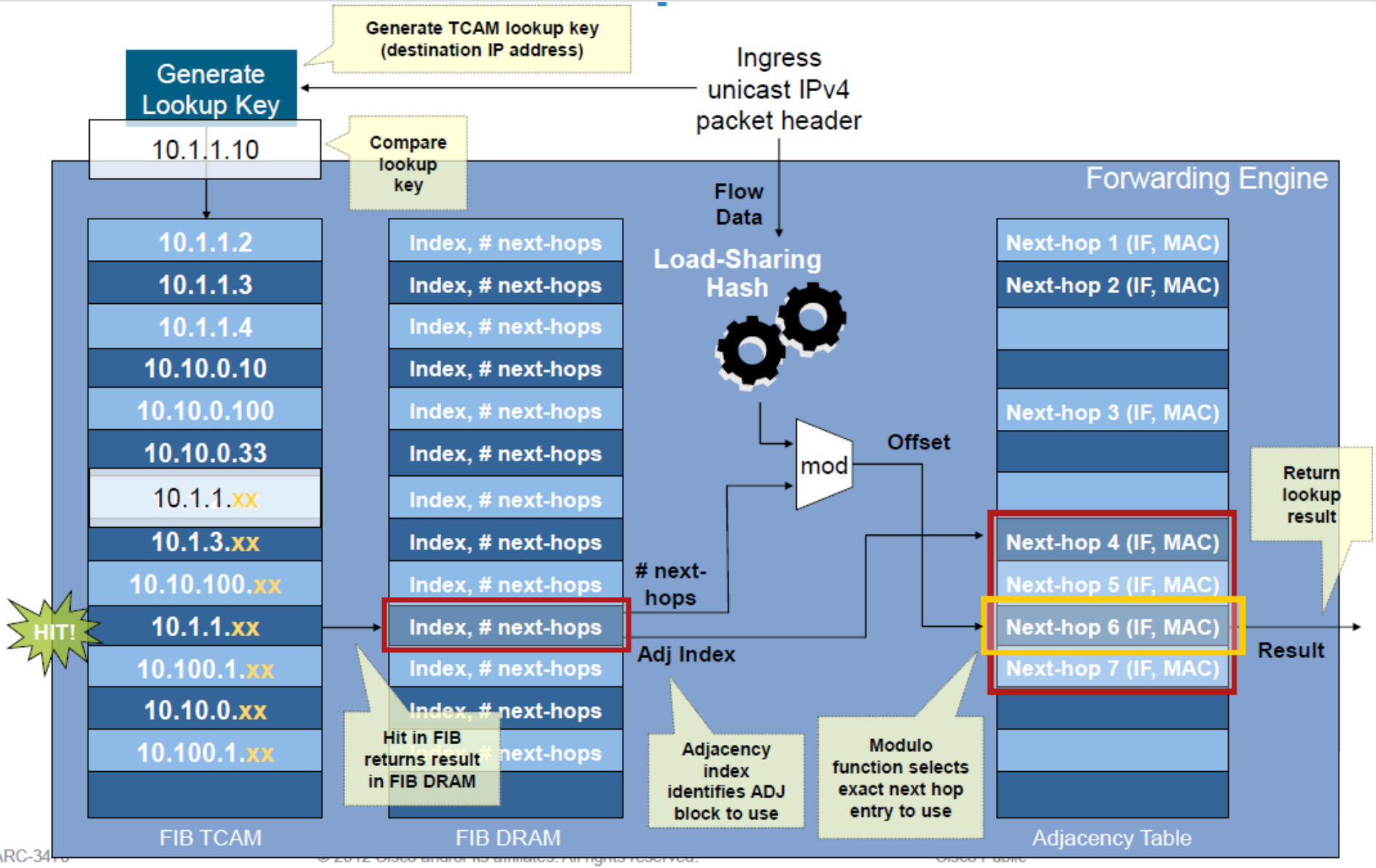
- Nexus 7000 decouples control plane and data plane
- Forwarding tables built on control plane using routing protocols or static configuration
  - –OSPF, EIGRP, IS-IS, RIP, BGP for dynamic routing
- Tables downloaded to forwarding engine hardware for data plane forwarding
  - –FIB TCAM contains IP prefixes
  - –Adjacency table contains next-hop information



Hardware IP Forwarding Process

- FIB TCAM lookup based on destination prefix (longest-match)
- FIB “hit” returns adjacency, adjacency contains rewrite information (next-hop)
- Pipelined forwarding engine architecture also performs ACL, QoS, and NetFlow lookups, affecting final forwarding result

IPv4 FIB TCAM Lookup



Courtesy of Cisco Education Services

State of the art
Routers Architecture
(Juniper, Cisco, Huawei)

MPU B4

System Control and Management Unit

•Route calculation: All routing protocol packets are sent by the forwarding engine to the MPU for processing. In addition, the MPU broadcasts and filters packets, and downloads routing policies from the policy server.

•Outband communication between boards: The LAN switch modules integrated on the MPU provide outband communications between boards. In this manner, messages can be controlled, maintained, and exchanged between SFUs and LPUs.

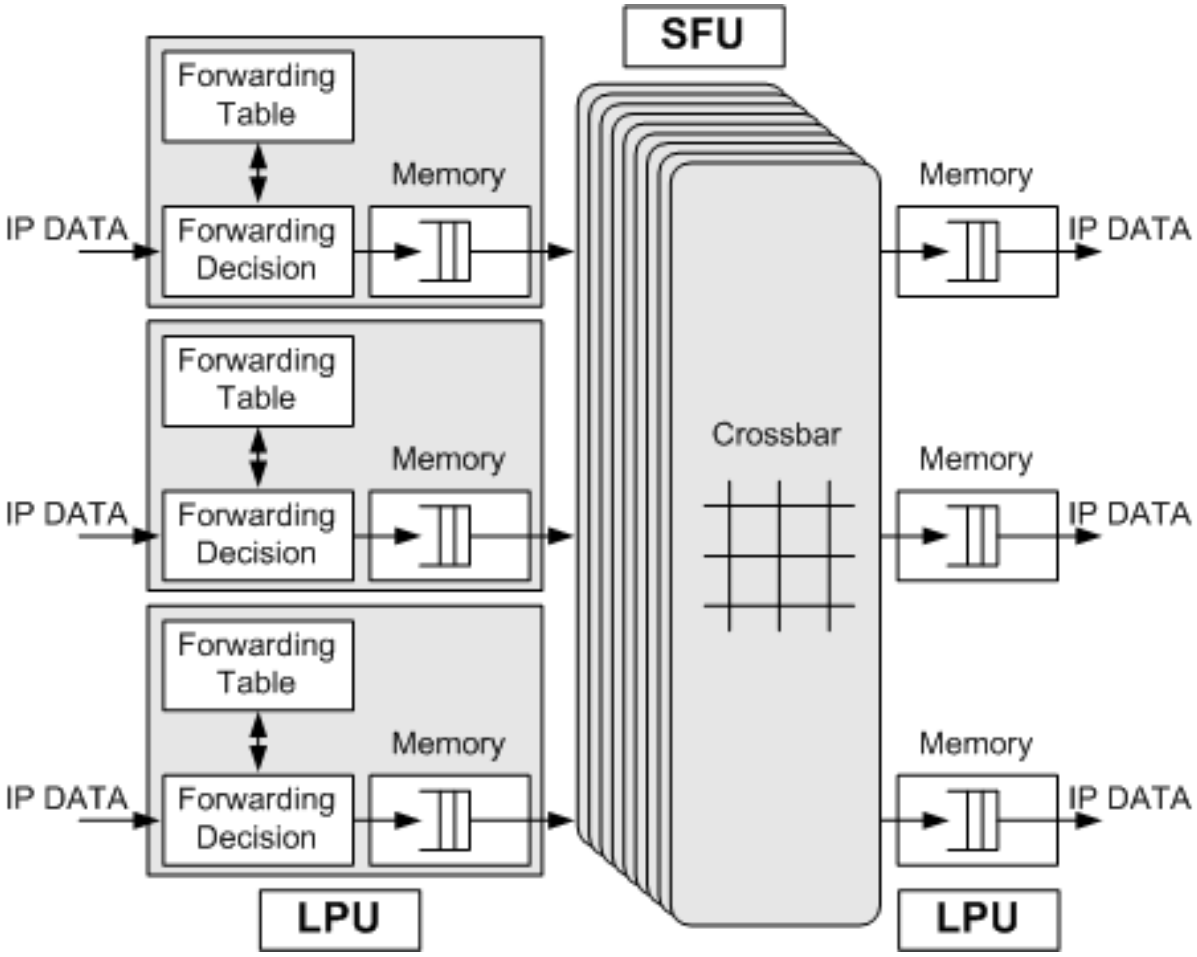
•Device management and maintenance: Devices can be managed and maintained through the management interfaces (serial interfaces) provided by the MPU.

•Data configuration: The MPU stores configuration data, startup files, charging information, upgrade software, and system logs.

•Data storage: The MPU provides two interfaces for CF cards, which serve as mass storage devices to store data files.

Courtesy of Huawei Education Services

Switching Network



References:

1. HUAWEI NetEngine5000E Core Router, NE5000E Product Description, 2011

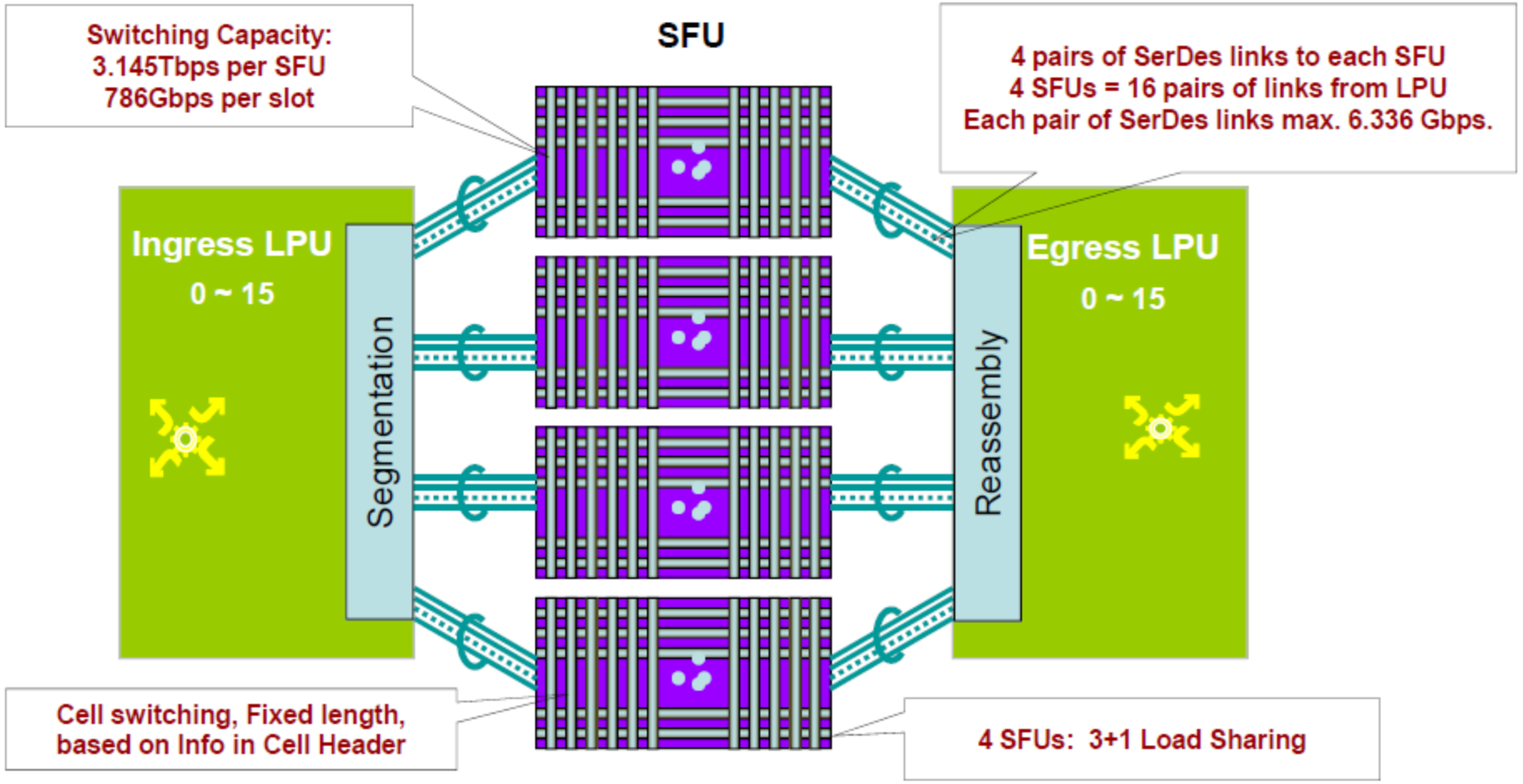
SFU-B

The Switch Fabric Unit B(SFUI-200-B) is the new generation switch board. It supports 200G full duplex capacity per slot. The Switch Fabric Unit B(SFUI-200-B) switches data for the entire system at line speed of 3.15 Tbit/s. This ensures a non-blocked switching network.

The NE40E-X16 has four SFUs working in 3+1 load balancing mode. The entire system provides a switching capacity at wire speed of 12.58 Tbit/s.

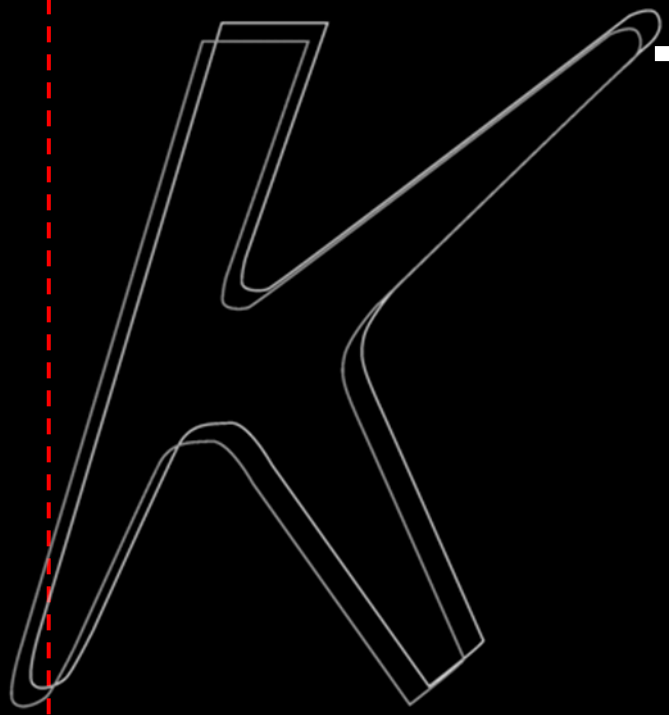
The four SFUs load balance services at the same time. When one SFU is faulty or being replaced, the other three SFUs automatically take over its tasks to ensure normal delivery of services.

Switching Fabric of NE5000E-X16



FIA = Fabric Interface ASIC

Courtesy of Huawei Education Services



Thank you

October, 15th 2013

Maurizio Ipsale
JNCI, CCSI, HCSI