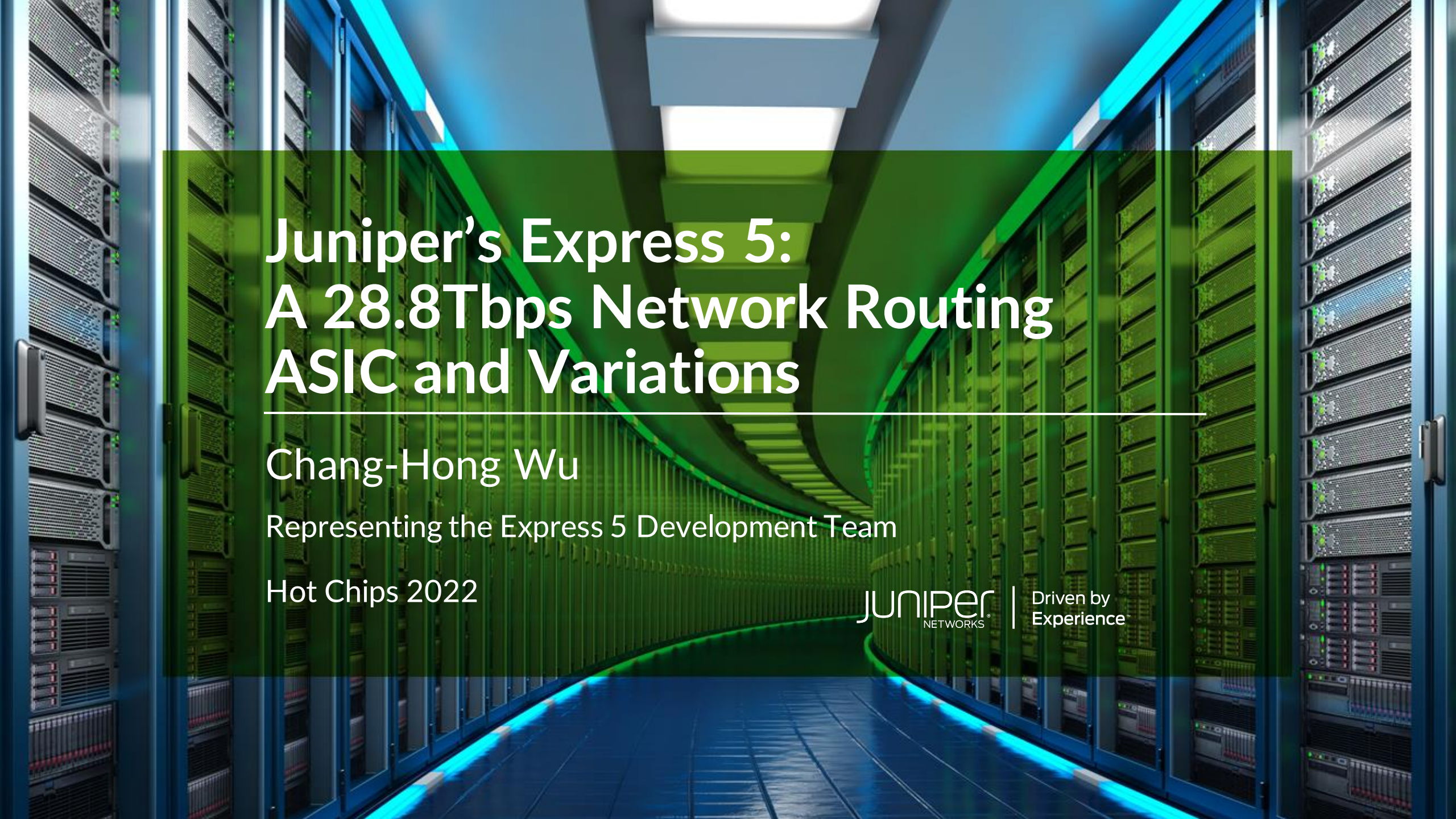


Juniper's Express 5: A 28.8Tbps Network Routing ASIC and Variations

Chang-Hong Wu

Representing the Express 5 Development Team

Hot Chips 2022

JUNIPER NETWORKS | Driven by Experience™

Disclaimer

This presentation describes the intended capabilities of Juniper's silicon. No commitment is made or implied on delivering these features or products, which are subject to change at any time without notice.

No purchasing decision should be made contingent upon Juniper Networks delivering any feature or functionality depicted in this presentation.

© 2022 Juniper Networks

What is Express 5?

JUNIPER NETWORKS

Routers

~~Layer 3 Devices~~

Deep Packet Buffers

High Logical Scale

Switches

~~Layer 2 Devices~~

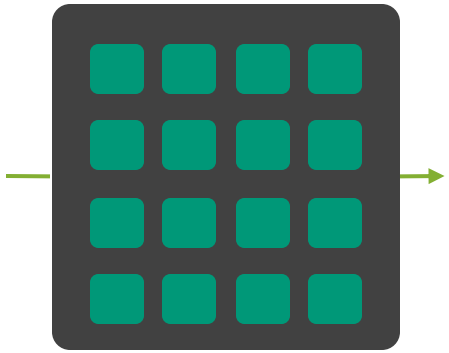
Shallow Packet Buffers

Low Logical Scale

© 2022 Juniper Networks

Juniper Routing ASICs

Trio ASICs for MX

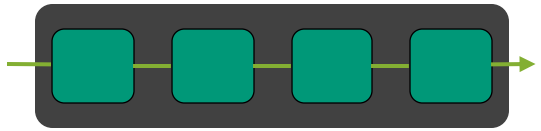


Multiple Packet Processing Engines Run-to-Completion

Flexibility / Logical Scale / ML Enabled

Latest generation: Trio 6

Express ASICs for PTX



Programmable Pipeline

Bandwidth and Energy Efficiency

Latest generation: Express 5

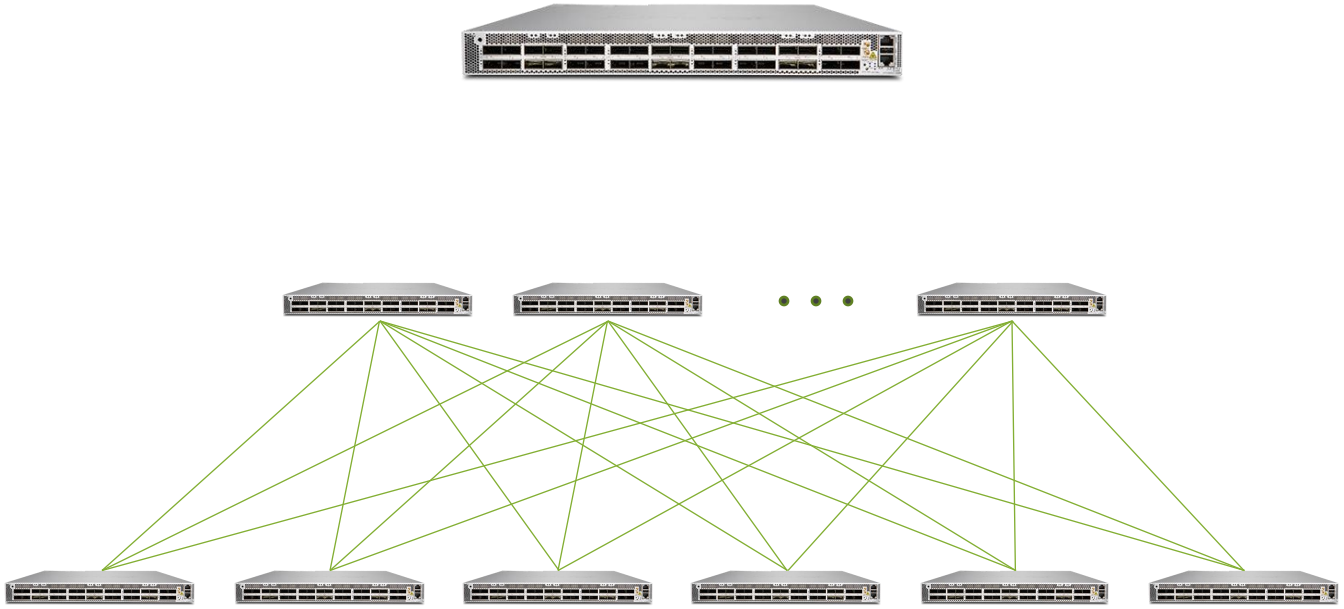
© 2022 Juniper Networks

Express 5 Goals

JUNIPER NETWORKS

Scale Out

Using Many Small-Form-Factor Systems



Scale Up

Using Fewer Large Modular Systems

© 2022 Juniper Networks

Scale Out: Why 28.8Tbps Single ASIC for Express 5?

And not 19.2Tbps

36 Pluggable Optical Cages Fit in Standard 19" Racks

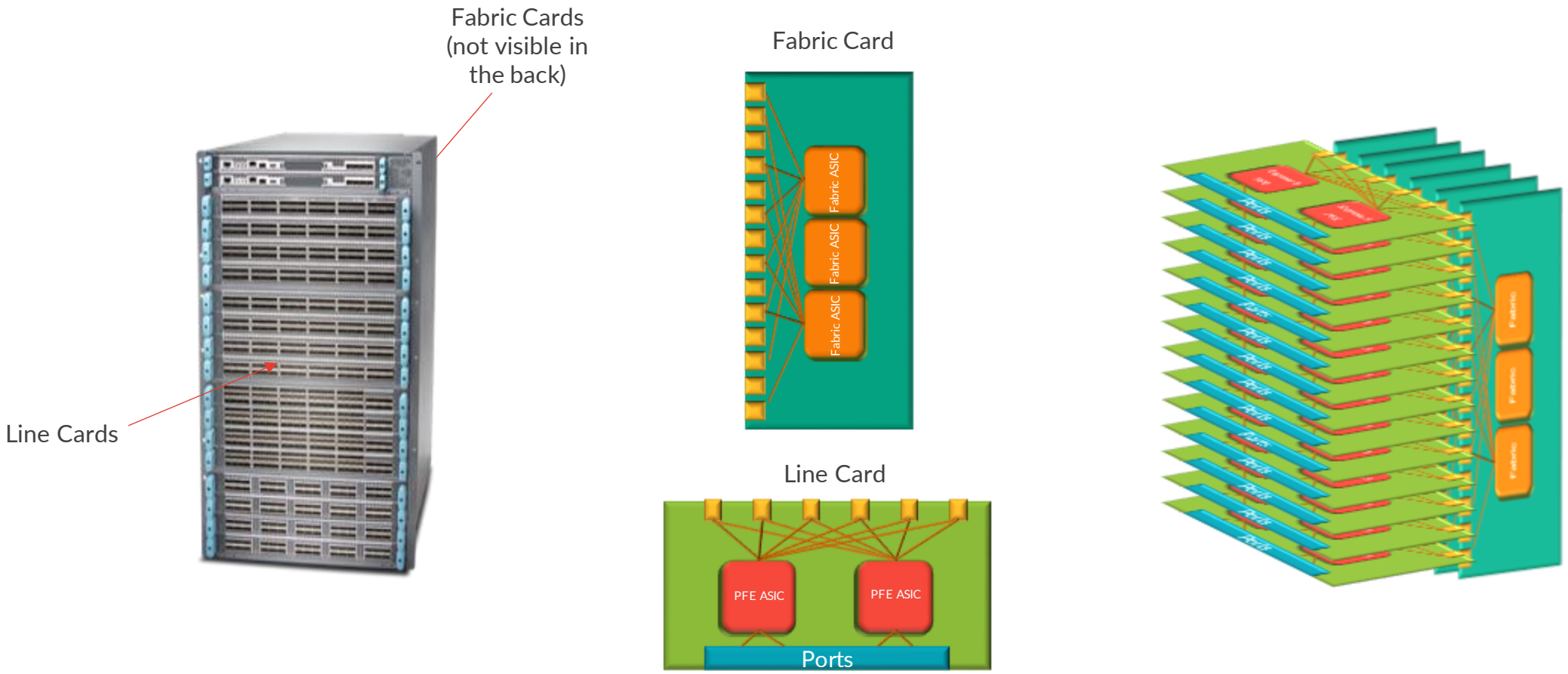
Common Radix: 32 or 36

36 x 800Gbps = 28.8Tbps

© 2022 Juniper Networks

Scale Up: Modular Chassis

Built using internal fabric connected components



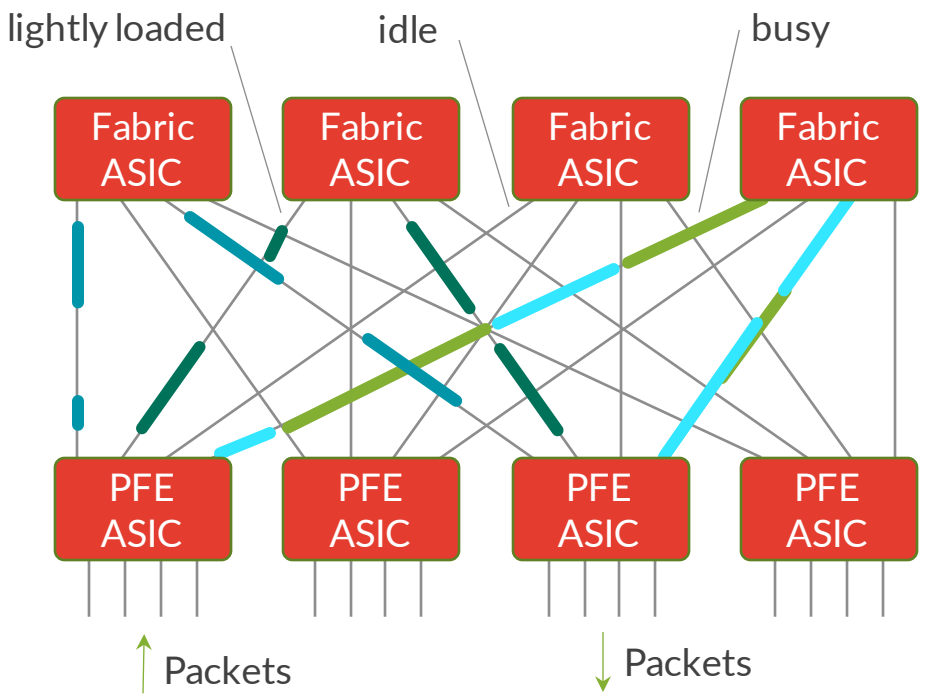
Example: 16 Line Cards x 28.8Tbps per Line Card = 460.8Tbps

Or 4608 x 100GbE ports

© 2022 Juniper Networks

Juniper Public

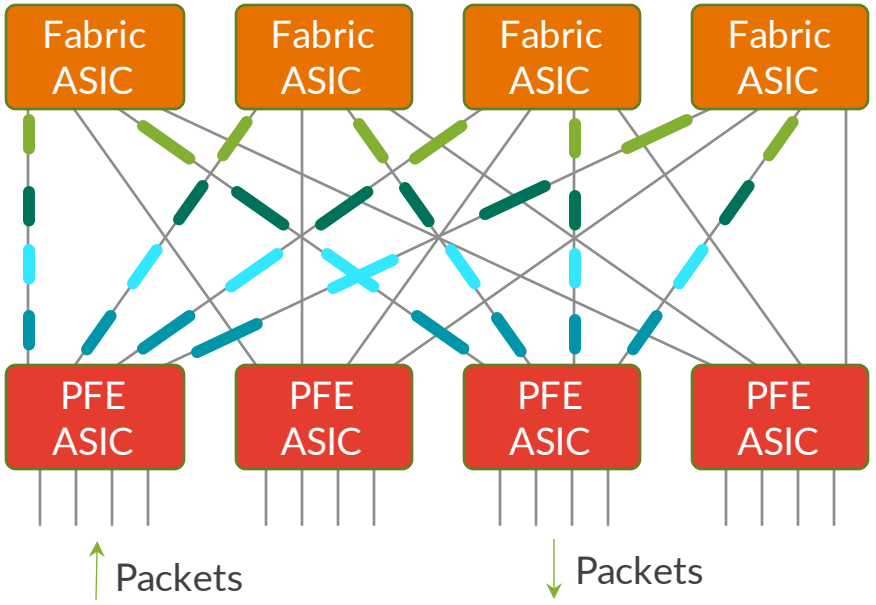
Pitfalls of Naïve Packet-Based Fabric Design



- Ethernet devices for PFE and fabric
- Lack of speedup for scheduling
- Unbalanced distribution of flows
- Dramatically different sized packets
- Low utilization of links
- Low energy efficiency

© 2022 Juniper Networks

Express 5: Cell-Based Fabric Design



- Even spray of small cells
- High link utilization
- Resiliency: retransmit-on-error
- VoQ based fabric protocol
- Low-latency cell switching
- High energy efficiency
- Interoperable with previous generation line cards

© 2022 Juniper Networks

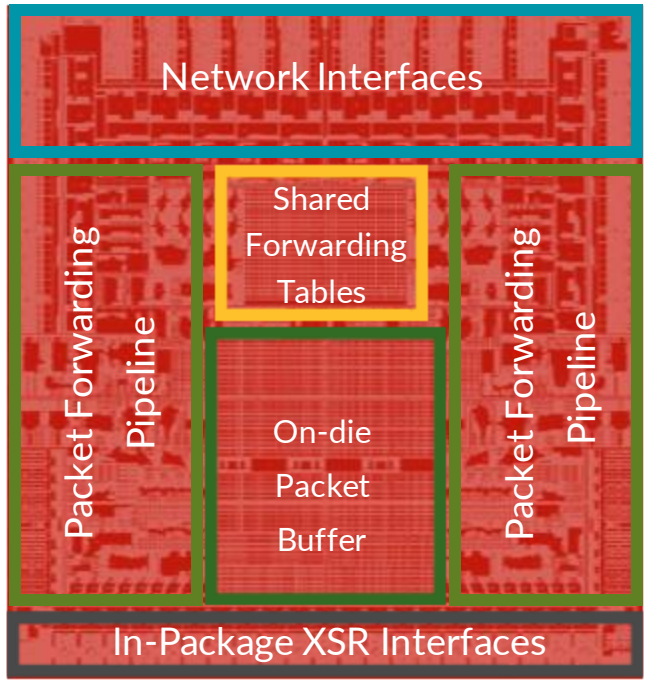
Express 5

Building Blocks

JUNIPER NETWORKS

X-Chiplet: eXpress forwarding chiplet

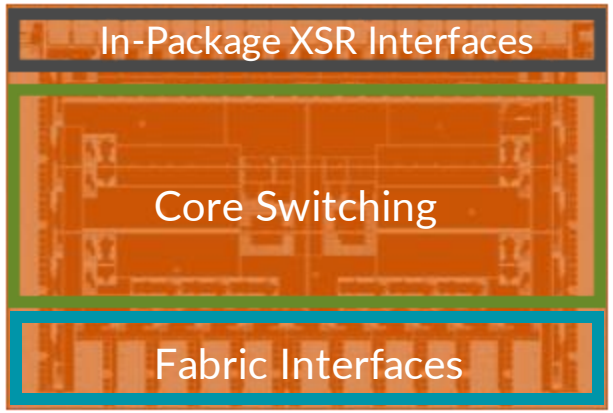
- TSMC 7nm
- 59 billion transistors
- 3 billion bits of on-die SRAM
- Multiple HBM2e Interfaces
- Multiple 112G Long Reach (LR) PAM4 SerDes
- Multiple 112G eXtreme Short Reach (XSR) PAM4 SerDes
- PCIe, Host Ethernet and other misc interfaces



© 2022 Juniper Networks

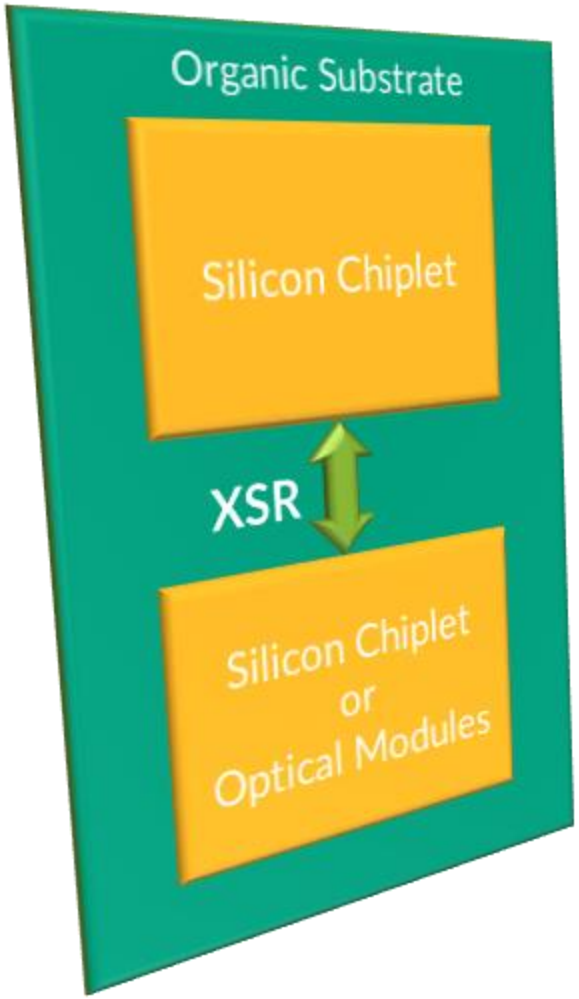
F-Chiplet: Fabric interface/switching chiplet

- TSMC 7nm
- 35 billion transistors
- 0.29 billion bits of on-die SRAM
- Multiple 112G Long Reach (LR) PAM4 SerDes
- Multiple 112G eXtreme Short Reach (XSR) PAM4 SerDes
- PCIe and other misc interfaces



© 2022 Juniper Networks

Why CEI-112G-XSR-PAM4 In-Package Interconnect?



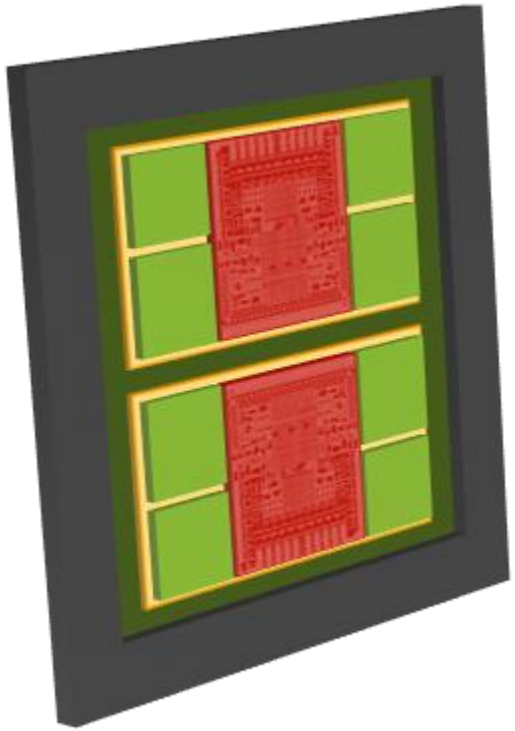
- Low power and area efficient
- High bandwidth on organic substrate channels
- Standardized by OIF (proposed by Juniper & partners)
- Allows for future Co-Packaged Optics

© 2022 Juniper Networks

Express 5

ASIC Variations

ASIC 1: 28.8Tbps Network Routing Device



- 2 X-chiplets, connected through 112G-XSR interfaces, with >32Tbps of aggregate bandwidth
- 118 billion transistors, 6 billion SRAM bits
- Multiple HBM2e stacks
- 2 silicon interposers, each ~1.5x reticle size
- 85mm x 85mm organic substrate
- 6756 BGA balls, 1mm pitch, square patterns
- Bare dies with stiffener ring

© 2022 Juniper Networks

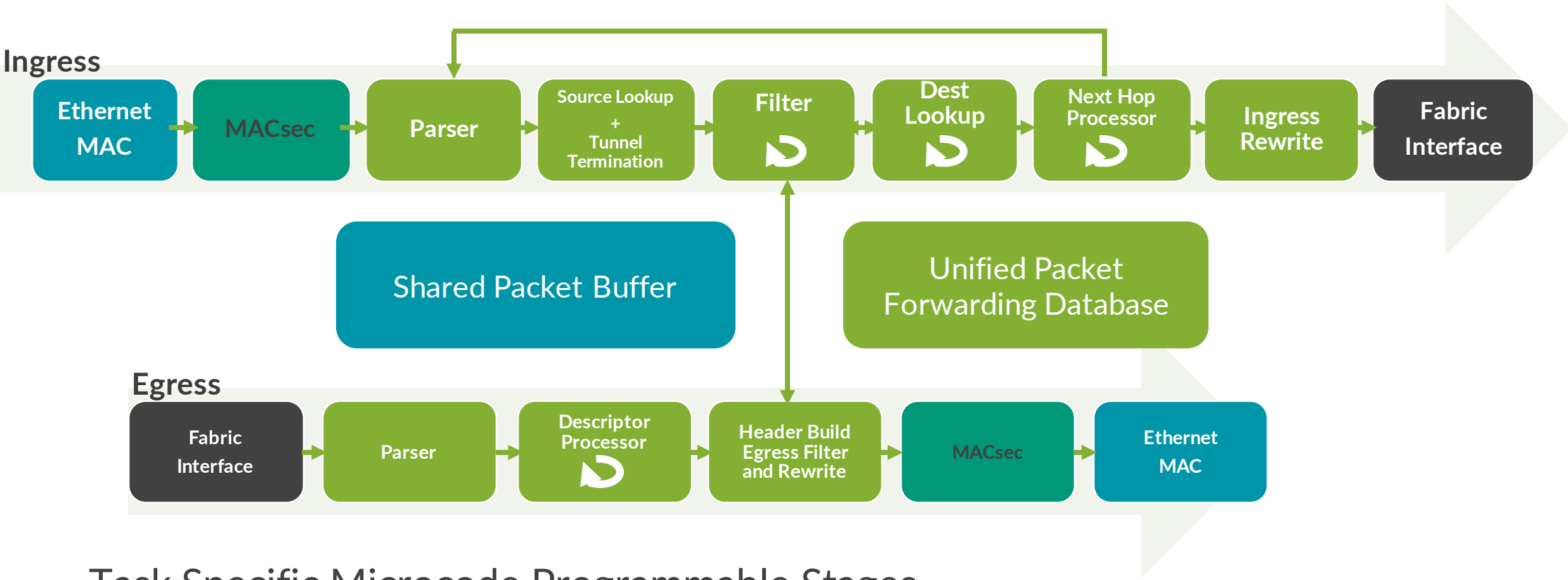
Juniper Public

Features

Speed	Scale	Flexibility	Security	Visibility	QoS
288 x 112G LR SerDes for 36 x 800GbE or 72 x 400GbE or 144 x 200GbE or 288 x 100GbE or mixture of ports Up to ~10 billion packets per second	10+M entries Internet+ sized routing table Unified packet forwarding database Multi-dimensional scaling Extendable to HBM	Programmable pipeline stages New protocol support including SRv6, BIER	High performance/ high capacity packet filters Integrated MACSec on all ports at line rate	8M counters Native IPFIX export Inband Network Telemtry	Hybrid on-die and HBM packet buffer Congested queues to HBM HQoS support

© 2022 Juniper Networks

Packet Forwarding Pipelines



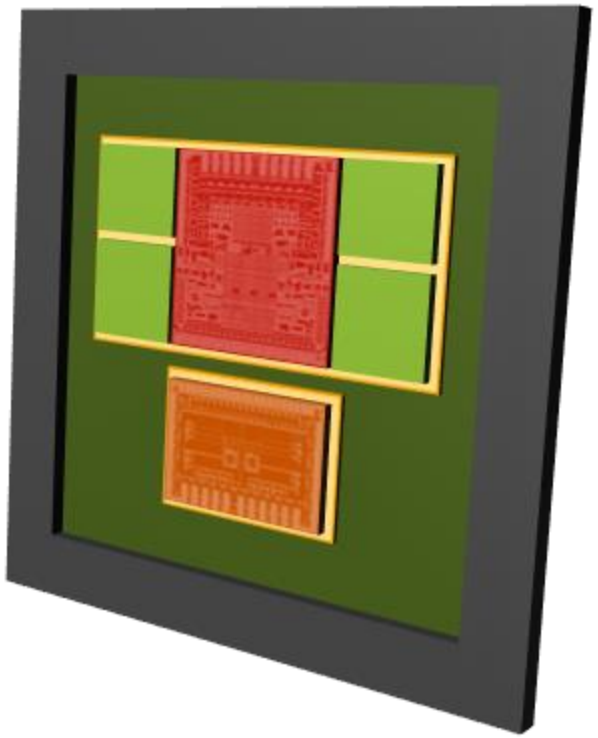
Task Specific Microcode Programmable Stages

P4 Runtime Support

© 2022 Juniper Networks

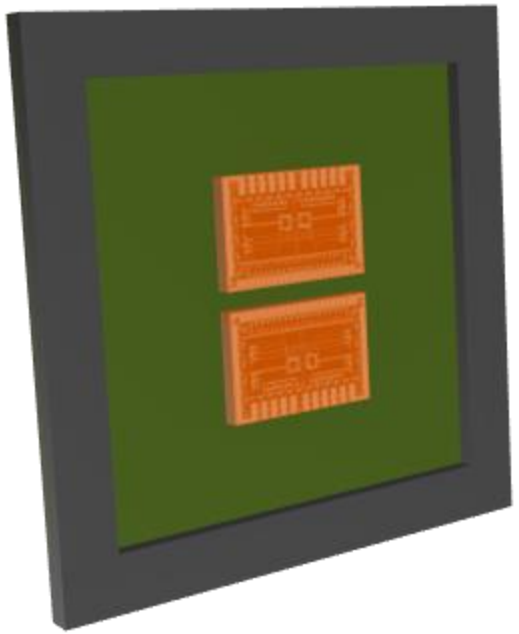
ASIC 2: Line Card Packet Forwarding Device with Fabric Interconnect

- 1 X-chiplet and 1 F-chiplet, connected through 112G-XSR interfaces, with >32Tbps of aggregate BW
- Multiple HBM2e stacks
- 2 silicon interposers on an organic substrate
- 14.4Tbps of Ethernet ports
- Same packet-forwarding functionality
- >16Tbps cell-based VoQ fabric interconnect
- Forward Error Correction and retransmit-on-error for the fabric links



© 2022 Juniper Networks

ASICs 3 and 4: Cell-Based Fabric Switch Devices



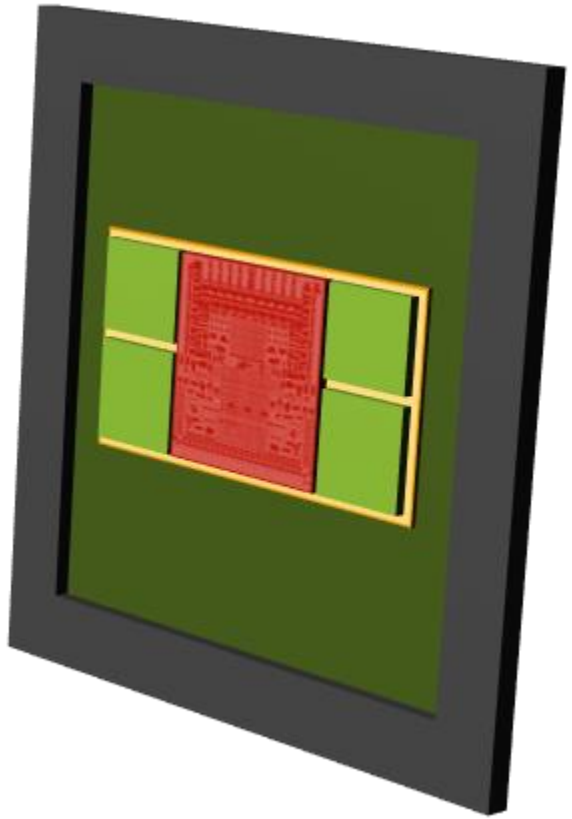
- > 32Tbps cell-based fabric switch ASIC
- Dual F-chiplets connected by XSR

- > 16Tbps cell-based fabric switch ASIC
- Single F-chiplet

© 2022 Juniper Networks

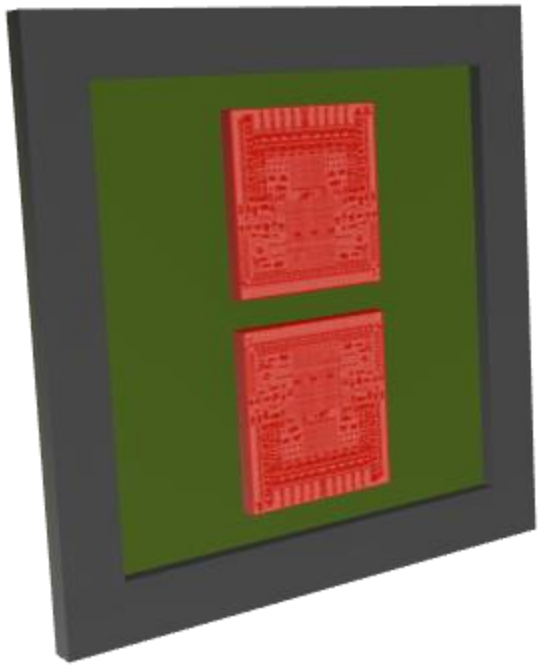
ASIC 5: 14.4Tbps Network Routing Device

- 1 X-chiplet, multiple HBM2e stacks, 1 silicon interposer
- 14.4Tbps of Ethernet ports

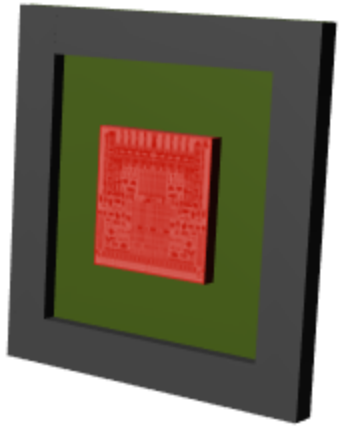


© 2022 Juniper Networks

ASICs 6 and 7: Networking Devices without HBM



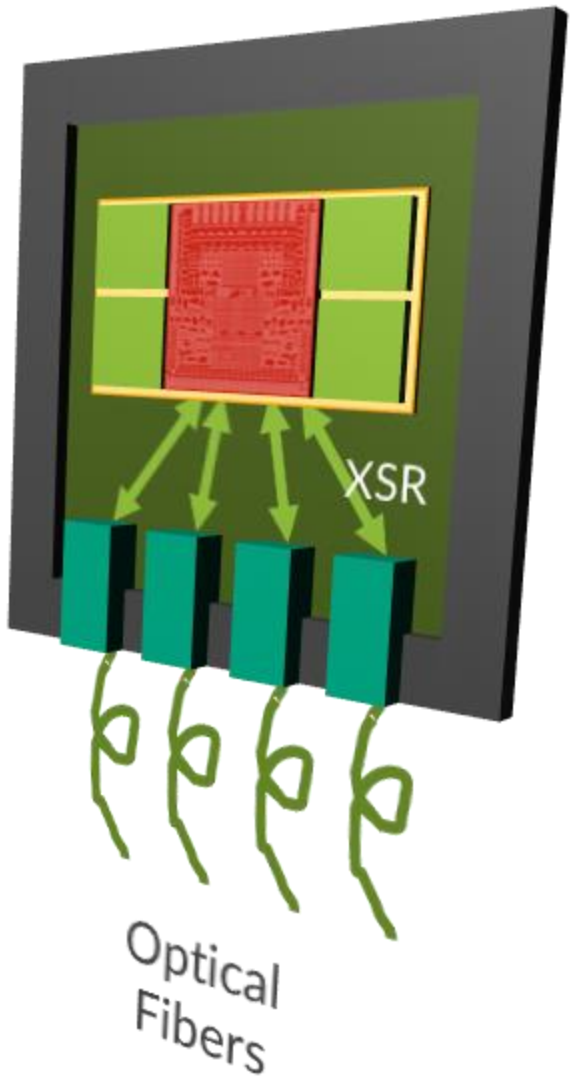
- 28.8Tbps of Ethernet ports
- 2 X-chiplets, connected by XSR



- 14.4Tbps of Ethernet ports
- 1 X-chiplet

© 2022 Juniper Networks

ASIC 8: Future Integration of Co-Packaged Optics

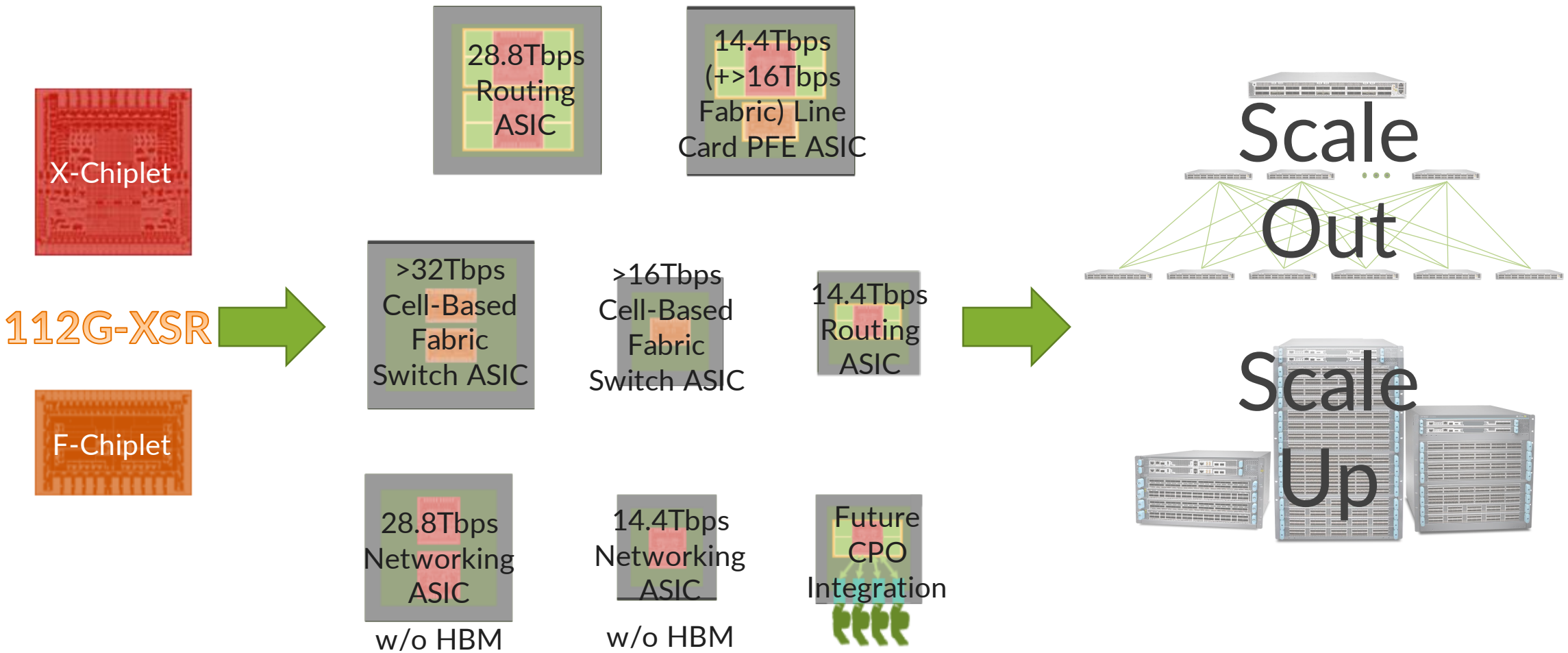


© 2022 Juniper Networks

Summary

JUNIPER NETWORKS

Express 5 – High Bandwidth, High Scale, Programmable ASIC Family



© 2022 Juniper Networks


Thank you
JUNIPER NETWORKS
Driven by Experience™