Pivotal Issues for 400 Gb/s Ethernet

Jeffery J. Maki
Distinguished Engineer, Optical
Juniper Networks

JUNIPER NETWORKS

Outline

400G Reach Objectives

Optical Lanes

Cabling

Potential PMDs

Form Factor Evolution

Duplex Fiber Approaches

Parallel Fiber Approaches

Relation to 100G

Relation to 40G

PMD Generations

Earlier Presentations

IEEE 802.3 Industry Connections Higher Speed Ethernet Ad Hoc

- September 2012: *100G Link Infrastructure Requirements to Support Future 400G PMDs*
  http://www.ieee802.org/3/ad_hoc/hse/public/12_09/maki_hse_01a_0912.pdf

IEEE 802.3 400 Gb/s Ethernet Study Group

- May 2013: *400GE at the Cost of 4 x 100GE*
  http://www.ieee802.org/3/400GSG/public/13_05/maki_400_01a_0513.pdf
- July 2013: *Long Shelf-Life Electrical Interfaces*
  http://www.ieee802.org/3/400GSG/public/13_07/maki_400_01_0713.pdf

400G Reach Objectives

Physical-Layer Specification	Status
Backplane	No Objective
Copper Cable Assembly	No Objective
MMF (At least 100 m)	Adopted November 2013
SMF (At least 500 m)	Adopted November 2013
SMF (At least 2 km)	Adopted November 2013
SMF (At least 10 km)	Adopted November 2013
SMF (At least 40 km)	No Objective

Optical Lanes

25G-λ

- Non-return to zero (NRZ)

50G-λ

- Non-return to zero (NRZ)
- Pulse amplitude modulation, four level (PAM-4), 25 Gbaud

100G-λ

- Pulse amplitude modulation, four level (PAM-4), 50 Gbaud
- Discrete multi-tone (DMT)

Cabling

Multi-Mode Fiber

- Parallel-fiber cabling assumed
- No proposals using duplex MMF
  (No 400G on a single fiber)

Single-Mode Fiber

- Parallel-fiber cabling
- Duplex-fiber cabling
  (400G on a single fiber proposed)

Distinct identity can take into account the choice of cabling

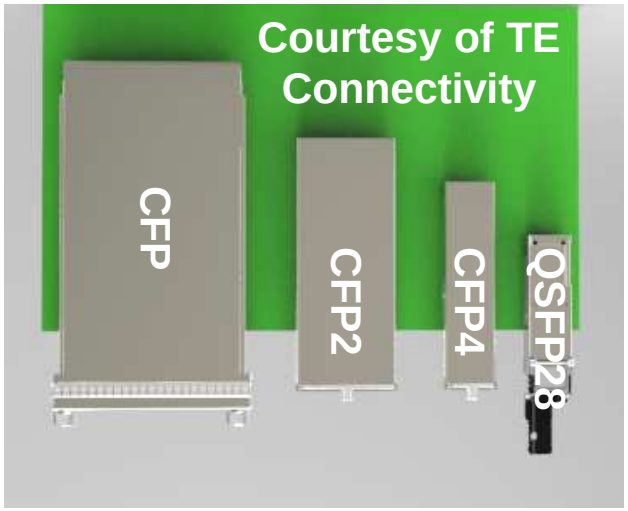
(Should consider to choose two solutions for the same reach differentiated by choice of cabling.)

PMD and Media (Other Solutions Possible)

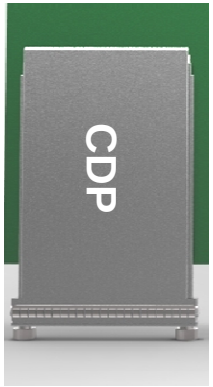
PMD	Media		
	Parallel MMF	Parallel SMF	Duplex SMF
MMF (At least 100 m)	4 x 100GBASE-SR4		
	400GBASE-SR16 (25G-λ: NRZ)		
	400GBASE-SR8 (50G-λ: NRZ or PAM-4)		
SMF (At least 500 m)		4 x 100GBASE-LR4	
		400GBASE-PSM16 (25G-λ: NRZ)	16-Ch 25G-λ WDM (NRZ)
		400GBASE-PSM8 (50G-λ: NRZ or PAM-4)	8-Ch 50G-λ WDM (NRZ or PAM-4)
		400GBASE-PSM4 (50G-λ: NRZ or PAM-4)	
		400GBASE-PSM4 (100G-λ: PAM-4 or DMT)	4-Ch 100G-λ WDM (PAM-4 or DMT)
SMF (At least 2 km)			8-Ch 50G-λ WDM (NRZ or PAM-4)
			4-Ch 100G-λ WDM (PAM-4 or DMT)
SMF (At least 10 km)			8-Ch 50G-λ WDM (NRZ or PAM-4)
			4-Ch 100G-λ WDM (PAM-4 or DMT)

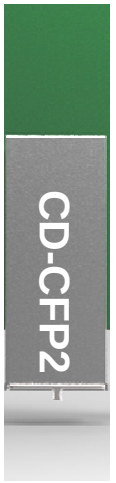
Form Factor Evolution

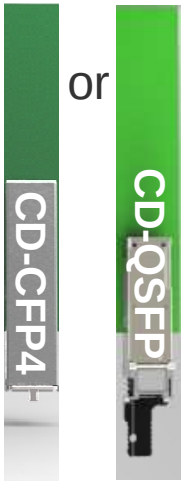
100G



400G



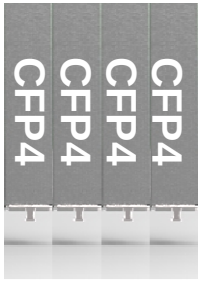




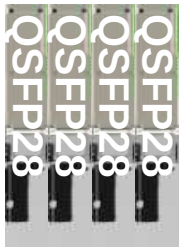
16x25G electrical lanes

8x50G electrical lanes

4x100G electrical lanes



Roman Numerals
XL = 40
C = 100
CD = 400



JUNIPER NETWORKS

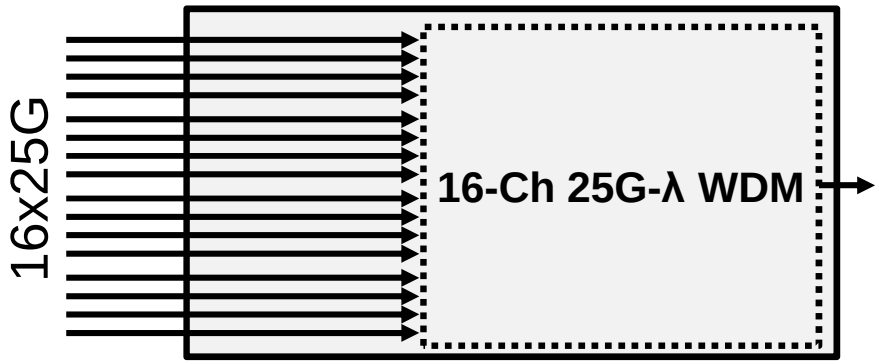
Wavelength Count

16 x 25G-λ

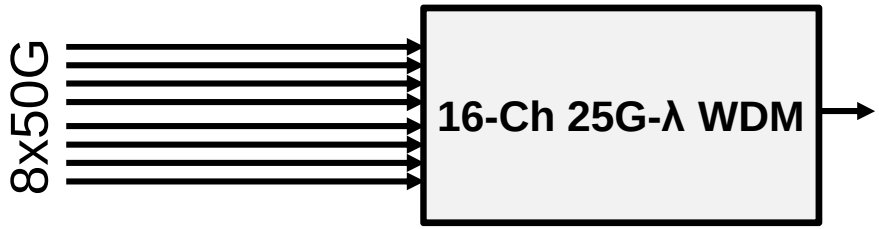
8 x 50G-λ

4 x 100G-λ

400G Duplex SMF, 25G-λ



CDP(LC)

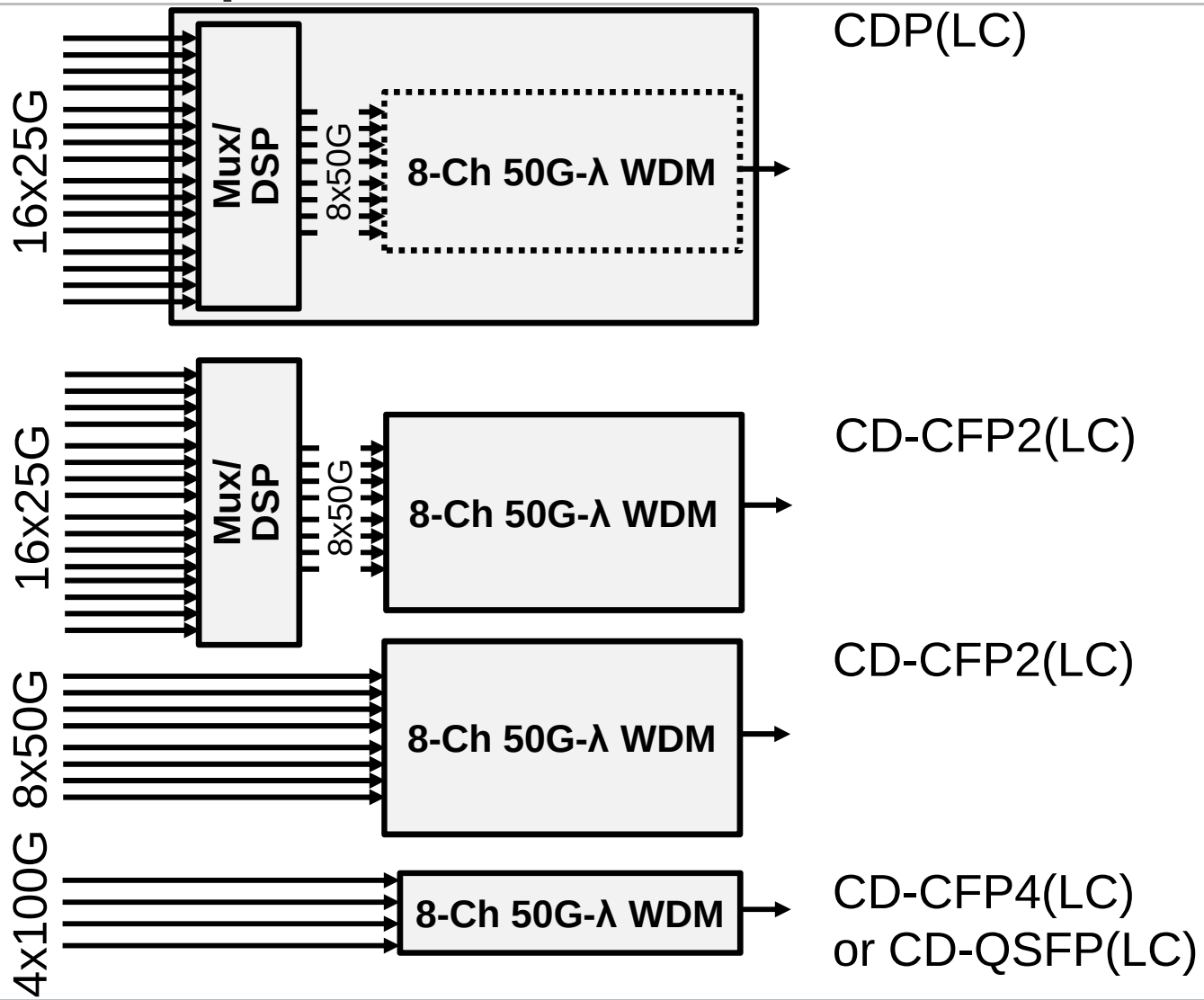


CD-CFP2(LC)

Problematic at this density, need reverse mux in module, not lowest cost

Transmit side only depicted.

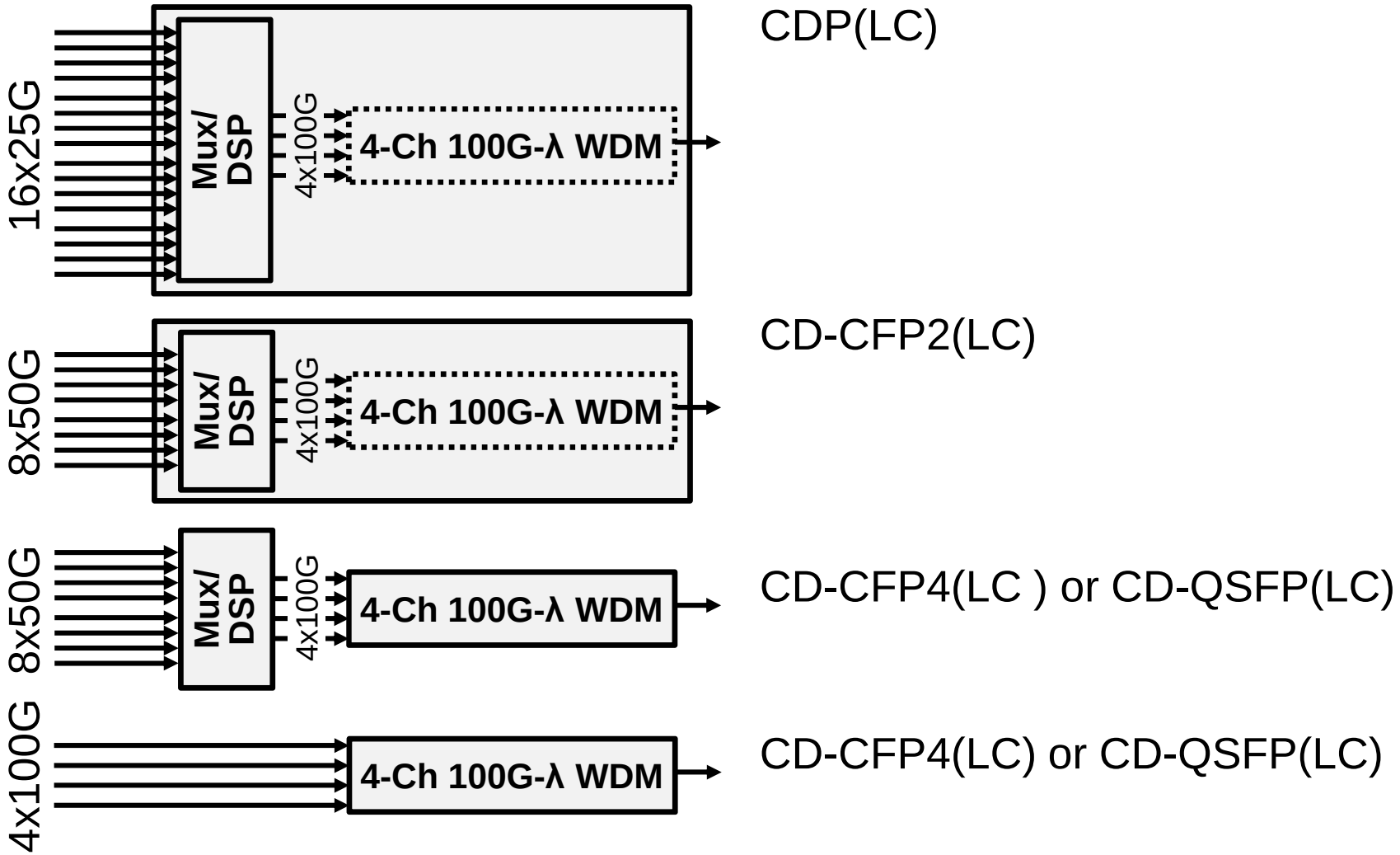
400G Duplex SMF, 50G-λ



Problematic at this density, need reverse mux in module, not lowest cost

Transmit side only depicted.

400G Duplex SMF, 100G-λ



Transmit side only depicted.

Influence of Wavelength Count

16 x 25G-λ

- One form factor generation supported before reverse mux is required
- CDP does not provide front-panel bandwidth density improvement over 4 x QSFP28
- Suitable for applications not requiring interop over form factor generations such as mega data-center applications

8 x 50G-λ

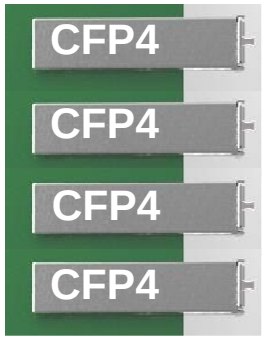
- Two form factor generations supported before reverse mux is required
- Second generation doubles front-panel bandwidth density
- Not desirable for router applications since front panel bandwidth density is only doubled while preserving interop over form factor generations

4 x 100G-λ

- Three form factor generations supported with progressive cost reduction
- Quadruples front panel bandwidth density while preserving interop over form factor generations

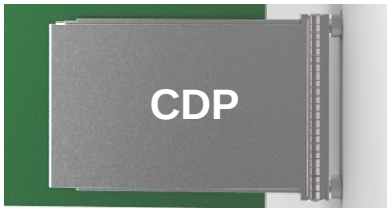
Parallel Fiber

4 x CAUI-4
(16 x 25G)



or

CDAUI-16
(16 x 25G)



or

CDAUI-8
(8 x 50G)

Current Presumptions of Lowest Cost Implementations of Parallel Optics

Electrical lane rate matched to the optical lane rate

- More accurate to say the baud is matched
- Electrical and optical lane counts are the same
  (WDM allows the optical lane count to be different than the fiber count)

Electrical and optical modulation is matched

FEC encode/decode avoided in the module

- FEC layer located in the host ASIC

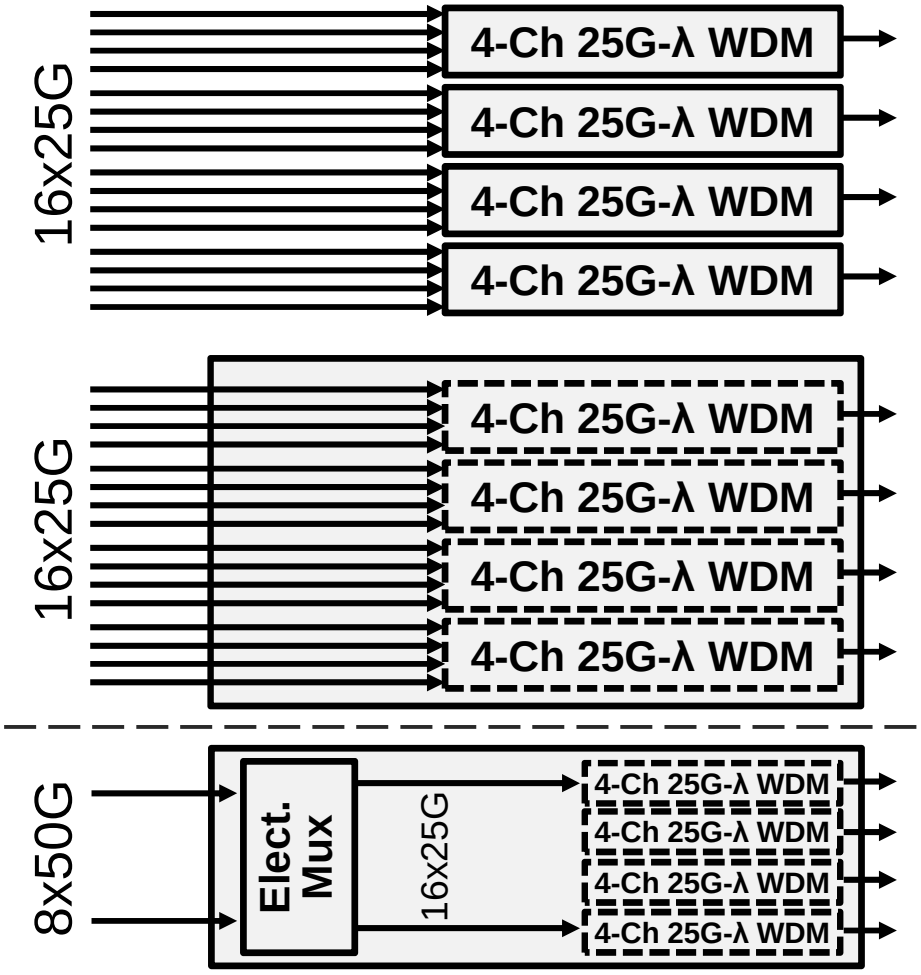
Lane adapting reverse mux employed in the module only when cost is justified to enable higher density of a legacy Ethernet rate

- Needed on routers
- Cost prohibitive on mega data-center platforms

25G-λ

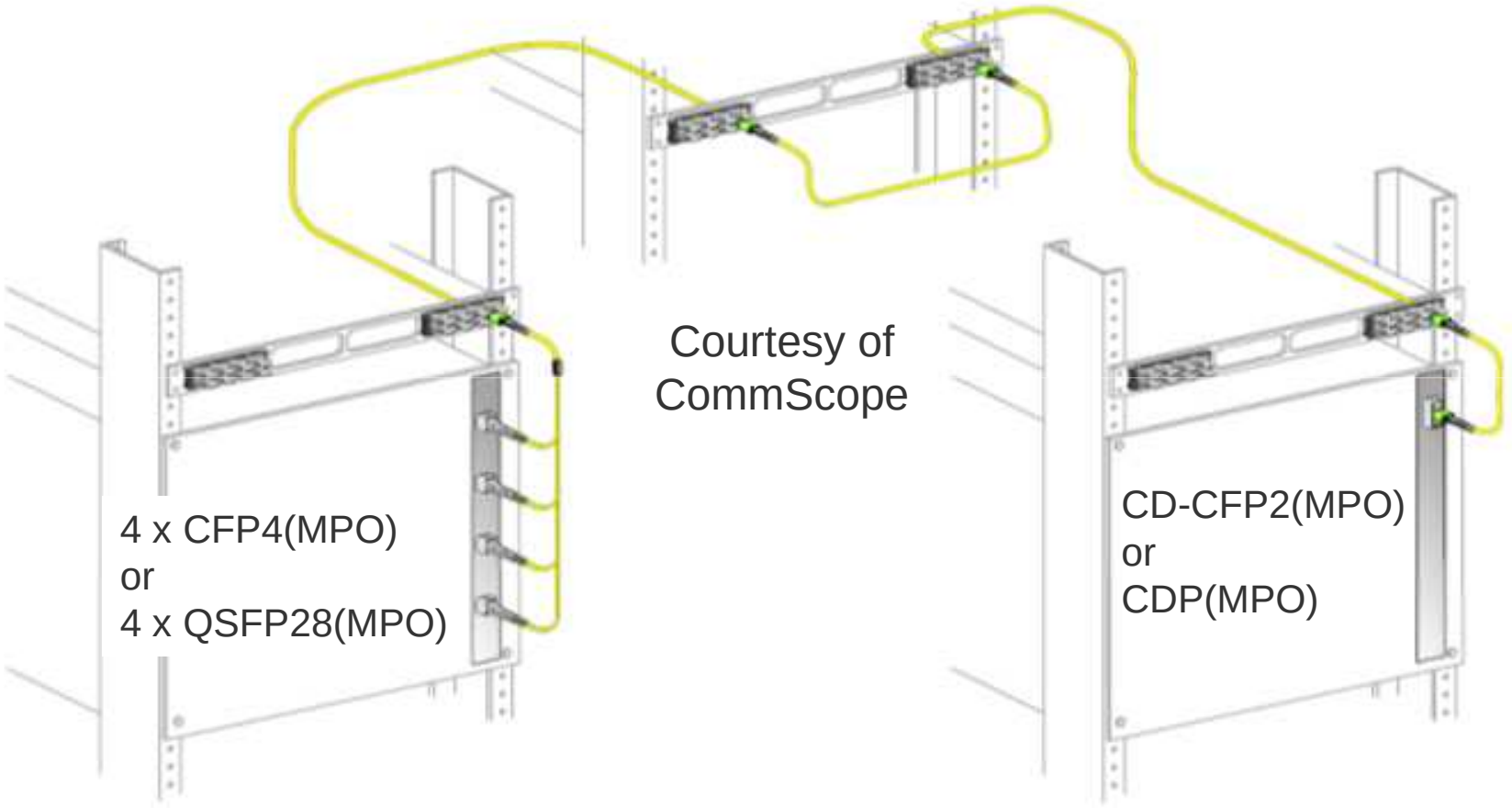
4-Lane Parallel Single Mode via Technology Reuse: 4 x 100GBASE-LR4 or 4 x 100G-CWDM4

Transmit side only depicted.



Also enables high-density 100GE for routers

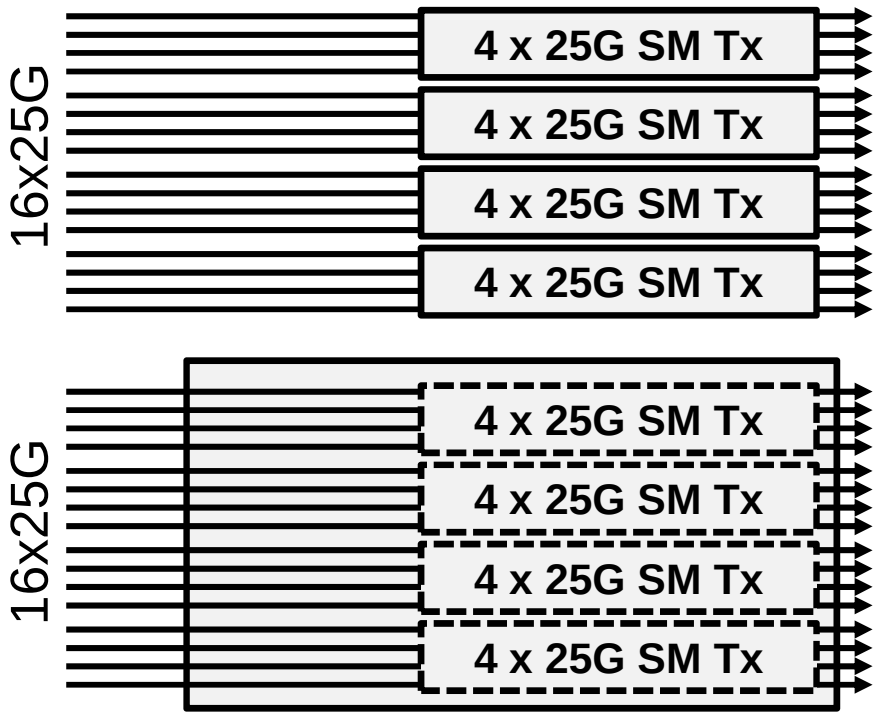
4-Lane Parallel SMF based on Technology Reuse



4 x 100GBASE-LR4 or 4 x 100G-CWDM4 for up to 500 m of SMF

16-Lane Parallel Single Mode: 4 x 100G-PSM4 or 400GBASE-PSM16

Transmit side only depicted.



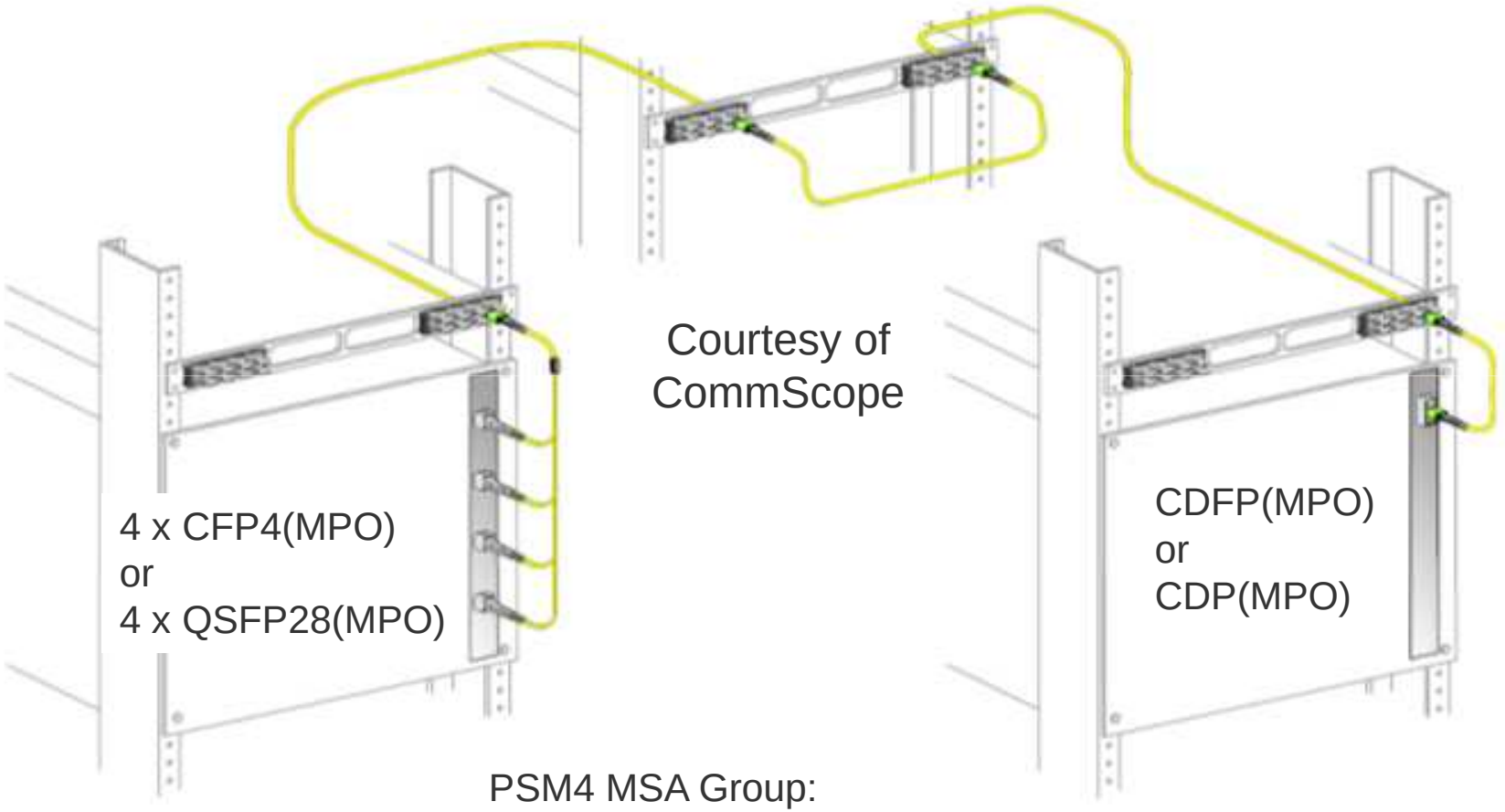
4 x CFP4(LC) or 4 x QSFP28(LC)

CDP(MPO) or CDFP(MPO)

Enables high-density 100GE in CDFP

PSM16 (or even PSM8) is thought only to be useful for limited adoption, where PSM4 is preferred

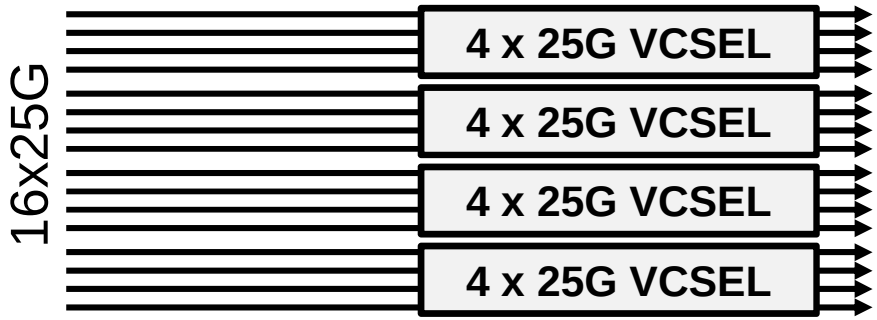
16-Lane Parallel SMF based on Technology Reuse



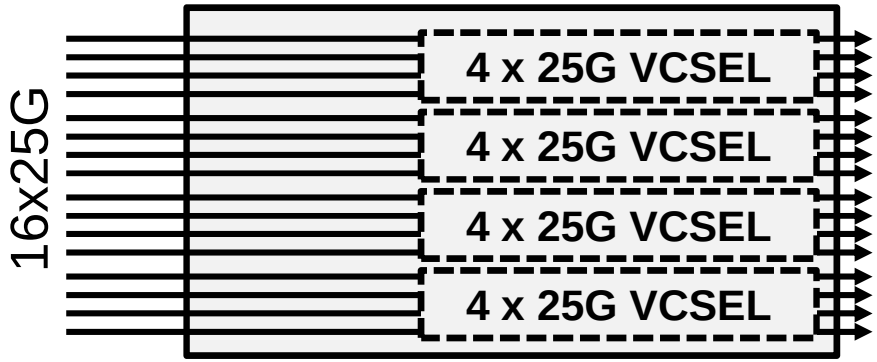
PSM4 MSA Group:
4 x 100G-PSM4 for up to 500 m of SMF

16-Lane, Parallel Multi-Mode via Technology Reuse: 4 x 100GBASE-SR4 or 400GBASE-SR16

Transmit side only depicted.

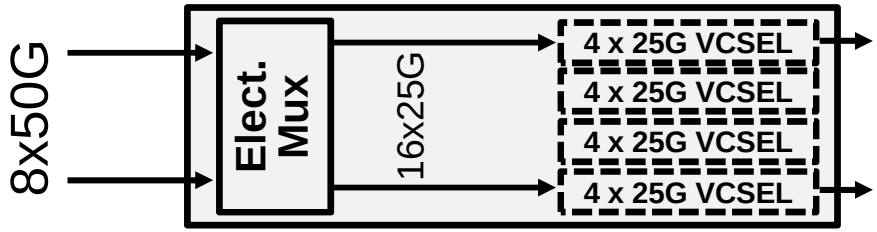


4 x CFP4(LC)



CDP(MPO) or CDFP(MPO)

Cost optimal

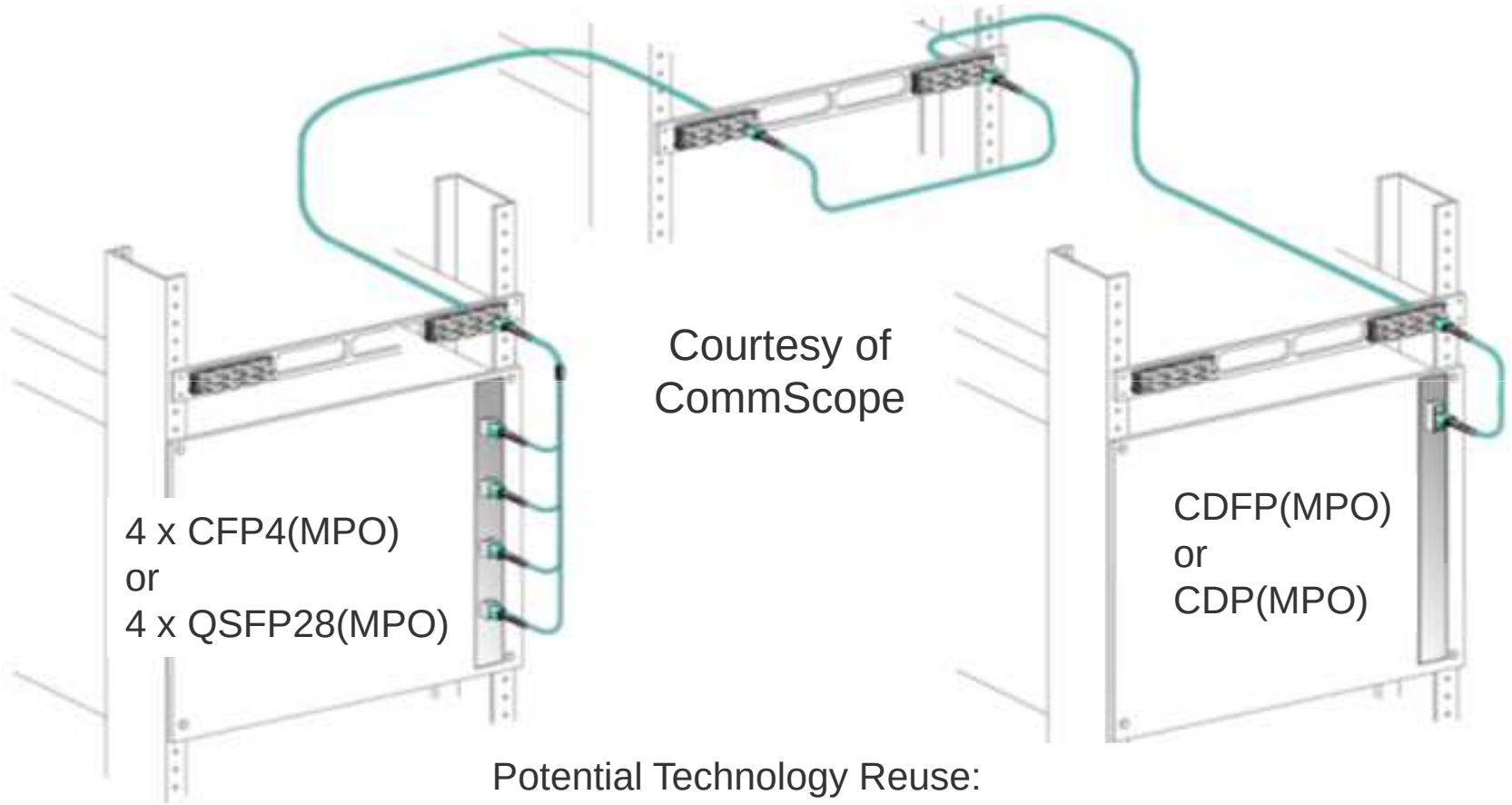


CD-CFP2(MPO)

Also enables high-density 100GE for routers

JUNIPER NETWORKS

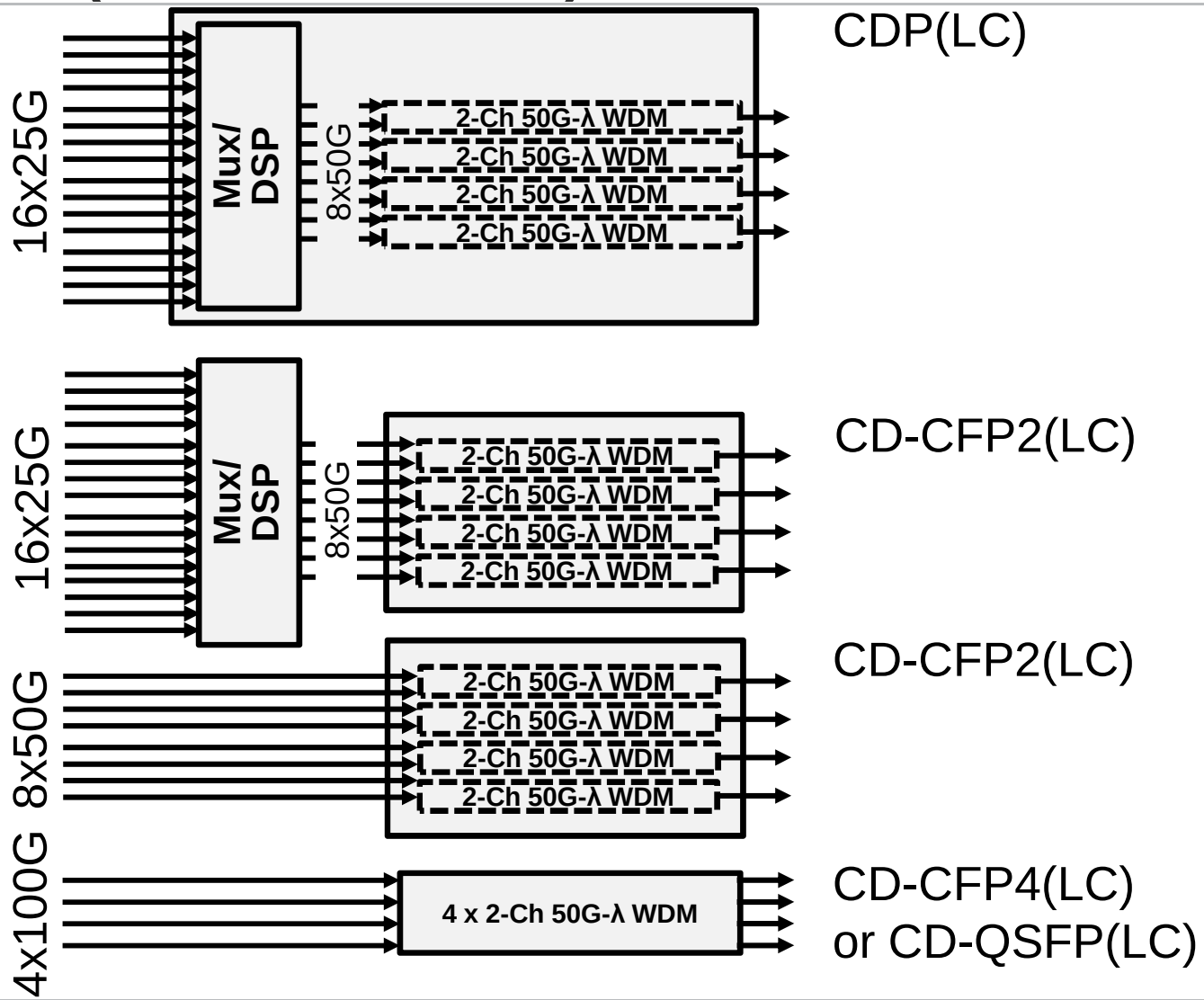
16-Lane Parallel MMF based on Technology Reuse



Potential Technology Reuse:
4 x 100GBASE-SR4 for 100 m on OM4

50G-λ

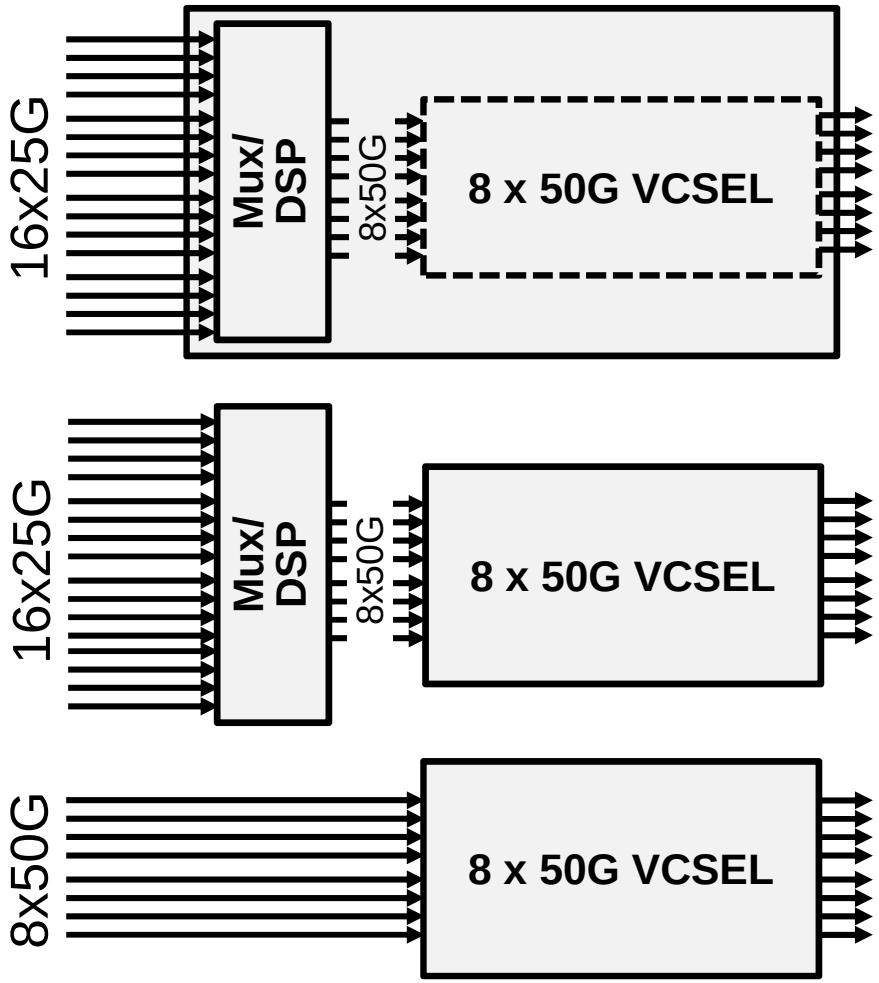
4-Lane, Parallel Single Mode: 4 x (2-Ch 50G-λ WDM)



Problematic at this density, need reverse mux in module, not lowest cost

Transmit side only depicted.

8-Lane, Parallel Multi-Mode
400G Parallel MMF, 50G-λ



CDP(LC)

CD-CFP2(LC)

CD-CFP2(LC)

Cost optimal

Transmit side only depicted.

PMD Standard Longevity

"Long Shelf Life" PMD Standards, Invariant to Implementation

- Router-to-router and router-to-transport applications need good longevity of any standard to enable interoperation among competitors with various time to market goals
- Router designs can be pursued with increased font-panel bandwidth density without having to harmonize on new standards, otherwise competition is stifled
  - Choice of electrical interface should not hamper this goal
  - MAC and PCS standard would best be invariant to choice of electrical interface and PMD

"Limited Shelf Life" PMD Standards

- Mega data center applications need cost optimized standards
- Interop over form-factor generation is not required
- PMD should be optimized to use the prevalent electrical lane technology for lowest cost

Electrical Interface Evolution—The Sweet Spots

The ultimate electrical interface for a given Ethernet rate to achieve lowest cost is typically serial

The subsequent Ethernet rate can leverage the serial electrical interface of the previous Ethernet rate as lanes for an electrical interface with broad adoption and longevity

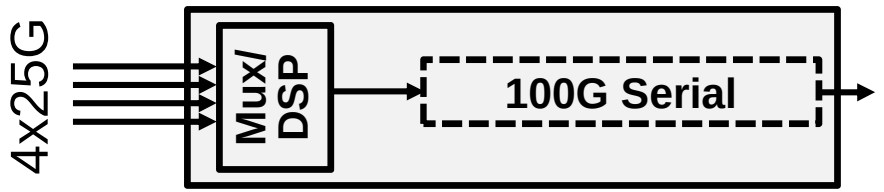
- Today: 10G electrical lane for 10GE leads to 4 x 10G electrical lanes for 40GE
- Future: 100G electrical lane (e.g., CDAUI-1) for 100GE leads to 4 x 100G electrical lanes (e.g., CDAUI-4) for 400GE

Connection to 100 Gb/s Ethernet

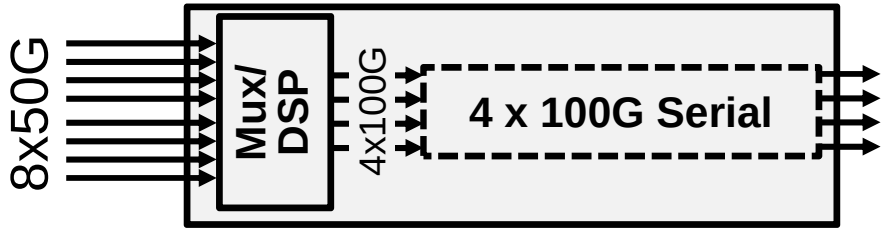
400 Gb/s Ethernet project will prompt technology investment

Market adoption of this new technology may first be for 100 Gb/s Ethernet applications

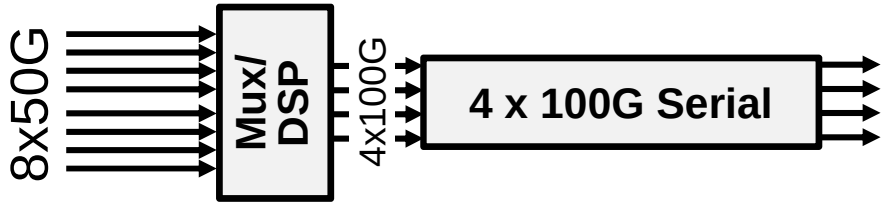
Adoption for 100 Gb/s Ethernet



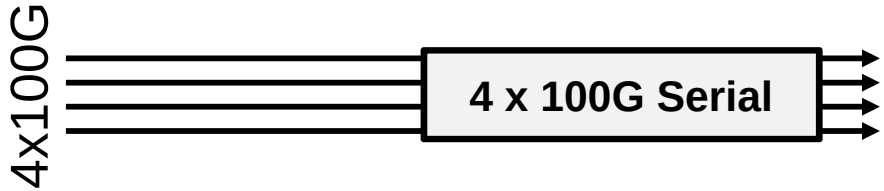
QSFP28(LC)
Single Port of 100GbE



CD-CFP2(MPO)
4 Independent Ports of 100GbE



CD-CFP4(MPO) or
CD-QSFP(MPO)
4 Independent Ports of 100GbE

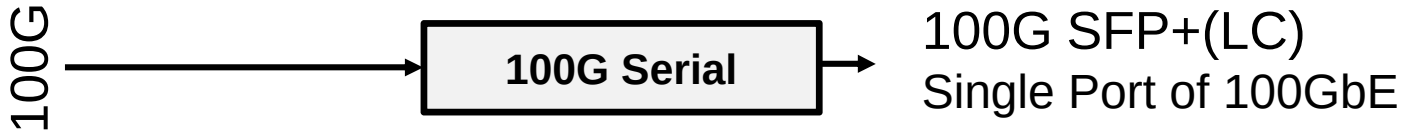


CD-CFP4(MPO) or
CD-QSFP(MPO)
4 Independent Ports of 100GbE

Transmit side only depicted.

JUNIPER NETWORKS

Adoption for 100 Gb/s Ethernet Cont.



Transmit side only depicted.

Connection to 40 Gb/s Ethernet

Development of 40G electrical lanes for lowest cost 40G Ethernet will prompt common technology also to support 50G electrical lanes

8 x 50G electrical interfaces for 400G Ethernet will also support 8 x 40G electrical interfaces for high-density 40G Ethernet

50G electrical lanes are likely to be prevalent for low-cost data-center interconnects, which will prompt the need for PMD standards using 50G optical lanes

PMD Generation—Early Adopter

PMD	Media		
	Parallel MMF	Parallel SMF	Duplex SMF
MMF (At least 100 m)	4 x 100GBASE-SR4		
	400GBASE-SR16 (25G-λ: NRZ)		
	400GBASE-SR8 (50G-λ: NRZ or PAM-4)		
SMF (At least 500 m)		4 x 100GBASE-LR4	
		400GBASE-PSM16 (25G-λ: NRZ)	16-Ch 25G-λ WDM (NRZ)
		400GBASE-PSM8 (50G-λ: NRZ or PAM-4)	8-Ch 50G-λ WDM (NRZ or PAM-4)
		400GBASE-PSM4 (50G-λ: NRZ or PAM-4)	
		400GBASE-PSM4 (100G-λ: PAM-4 or DMT)	4-Ch 100G-λ WDM (PAM-4 or DMT)
SMF (At least 2 km)			8-Ch 50G-λ WDM (NRZ or PAM-4)
			4-Ch 100G-λ WDM (PAM-4 or DMT)
SMF (At least 10 km)			8-Ch 50G-λ WDM (NRZ or PAM-4)
			4-Ch 100G-λ WDM (PAM-4 or DMT)

PMD Generation—Ideal for Router to Router & Router to Transport

PMD	Media		
	Parallel MMF	Parallel SMF	Duplex SMF
MMF (At least 100 m)	4 x 100GBASE-SR4		
	400GBASE-SR16 (25G-λ: NRZ)		
	400GBASE-SR8 (50G-λ: NRZ or PAM-4)		
SMF (At least 500 m)		4 x 100GBASE-LR4	
		400GBASE-PSM16 (25G-λ: NRZ)	16-Ch 25G-λ WDM (NRZ)
		400GBASE-PSM8 (50G-λ: NRZ or PAM-4)	8-Ch 50G-λ WDM (NRZ or PAM-4)
		400GBASE-PSM4 (50G-λ: NRZ or PAM-4)	
		400GBASE-PSM4 (100G-λ: PAM-4 or DMT)	4-Ch 100G-λ WDM (PAM-4 or DMT)
SMF (At least 2 km)			8-Ch 50G-λ WDM (NRZ or PAM-4)
			4-Ch 100G-λ WDM (PAM-4 or DMT)
SMF (At least 10 km)			8-Ch 50G-λ WDM (NRZ or PAM-4)
			4-Ch 100G-λ WDM (PAM-4 or DMT)

PMD Generation—Data-Center 25G Elect. Ecosystem

PMD	Media		
	Parallel MMF	**Parallel SMF**	**Duplex SMF**
MMF (At least 100 m)	4 x 100GBASE-SR4		
	400GBASE-SR16 (25G-λ: NRZ)		
	400GBASE-SR8 (50G-λ: NRZ or PAM-4)		
SMF (At least 500 m)		4 x 100GBASE-LR4	
		400GBASE-PSM16 (25G-λ: NRZ)	16-Ch 25G-λ WDM (NRZ)
		400GBASE-PSM8 (50G-λ: NRZ or PAM-4)	8-Ch 50G-λ WDM (NRZ or PAM-4)
		400GBASE-PSM4 (50G-λ: NRZ or PAM-4)	
		400GBASE-PSM4 (100G-λ: PAM-4 or DMT)	4-Ch 100G-λ WDM (PAM-4 or DMT)
SMF (At least 2 km)			8-Ch 50G-λ WDM (NRZ or PAM-4)
			4-Ch 100G-λ WDM (PAM-4 or DMT)
SMF (At least 10 km)			8-Ch 50G-λ WDM (NRZ or PAM-4)
			4-Ch 100G-λ WDM (PAM-4 or DMT)

PMD Generation—Data-Center 50G Elect. Ecosystem

PMD	Media		
	Parallel MMF	**Parallel SMF**	**Duplex SMF**
MMF (At least 100 m)	4 x 100GBASE-SR4		
	400GBASE-SR16 (25G-λ: NRZ)		
	400GBASE-SR8 (50G-λ: NRZ or PAM-4)		
SMF (At least 500 m)		4 x 100GBASE-LR4	
		400GBASE-PSM16 (25G-λ: NRZ)	16-Ch 25G-λ WDM (NRZ)
		400GBASE-PSM8 (50G-λ: NRZ or PAM-4)	8-Ch 50G-λ WDM (NRZ or PAM-4)
		400GBASE-PSM4 (50G-λ: NRZ or PAM-4)	
		400GBASE-PSM4 (100G-λ: PAM-4 or DMT)	4-Ch 100G-λ WDM (PAM-4 or DMT)
SMF (At least 2 km)			8-Ch 50G-λ WDM (NRZ or PAM-4)
			4-Ch 100G-λ WDM (PAM-4 or DMT)
SMF (At least 10 km)			8-Ch 50G-λ WDM (NRZ or PAM-4)
			4-Ch 100G-λ WDM (PAM-4 or DMT)

PMD Generation—Data-Center 100G Elect. Ecosystem

PMD	Media		
	Parallel MMF	Parallel SMF	Duplex SMF
MMF (At least 100 m)	4 x 100GBASE-SR4		
	400GBASE-SR16 (25G-λ: NRZ)		
	400GBASE-SR8 (50G-λ: NRZ or PAM-4)		
SMF (At least 500 m)		4 x 100GBASE-LR4	
		400GBASE-PSM16 (25G-λ: NRZ)	16-Ch 25G-λ WDM (NRZ)
		400GBASE-PSM8 (50G-λ: NRZ or PAM-4)	8-Ch 50G-λ WDM (NRZ or PAM-4)
		400GBASE-PSM4 (50G-λ: NRZ or PAM-4)	
		400GBASE-PSM4 (100G-λ: PAM-4 or DMT)	4-Ch 100G-λ WDM (PAM-4 or DMT)
SMF (At least 2 km)			8-Ch 50G-λ WDM (NRZ or PAM-4)
			4-Ch 100G-λ WDM (PAM-4 or DMT)
SMF (At least 10 km)			8-Ch 50G-λ WDM (NRZ or PAM-4)
			4-Ch 100G-λ WDM (PAM-4 or DMT)

Conclusions

400 Gb/s Ethernet project has reach objectives with different market adoption time frames and longevity requirements

- First need 2-km reach on duplex SMF

Might see to choose a solution with less than desired longevity in favor of meeting a time to market goal while planning a more durable (better longevity) standard to supersede it that is perhaps for a different reach objective

Electrical interfaces will be required that are likely not within technical feasibility at this time but should not be precluded by how the 400 Gb/s MAC and PCS is defined