400G, 800G, and Terabit Pluggable Optics:

What You Need to Know

Mark Nowell - Cisco Fellow

BRKOPT-2699

Webex App

Questions?

Use the Webex app to chat with the speaker after the session

How

1. Find this session in the Cisco Events mobile app
2. Click “Join the Discussion”
3. Install the Webex app or go directly to the Webex space
4. Enter messages/questions in the Webex space

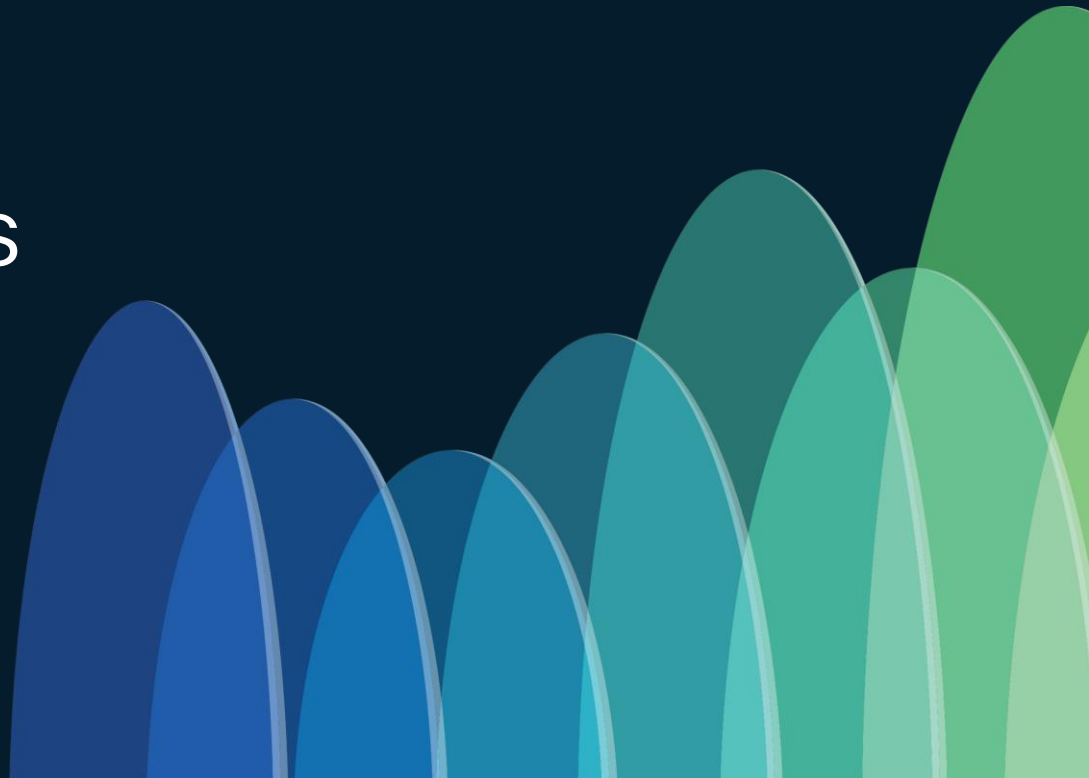
Webex spaces will be moderated by the speaker until February 28, 2025.

Agenda

- What is driving the optics interconnect market right now?
- Deploying 400G & 800G
  - Solutions & options
  - Deployment considerations
- What does AI/ML mean for optics?
  - New speeds & altering priorities
  - New implementations
- Wrap Up

Acknowledgements: This presentation would not exist without the inputs, expertise, and patience of many Cisco colleagues!

Market dynamics

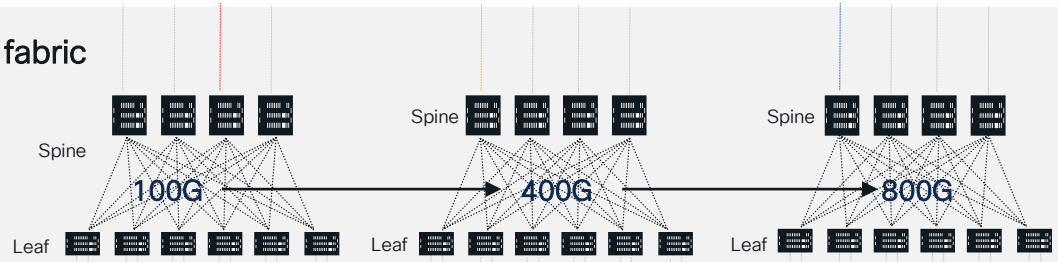


Speed evolution in the data center

Inter-Data Center

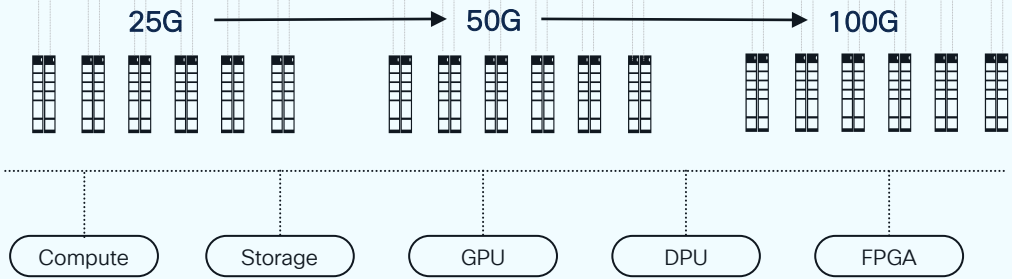
- Significant pluggable DCI (DWDM coherent)
- Open Line System

Switch fabric



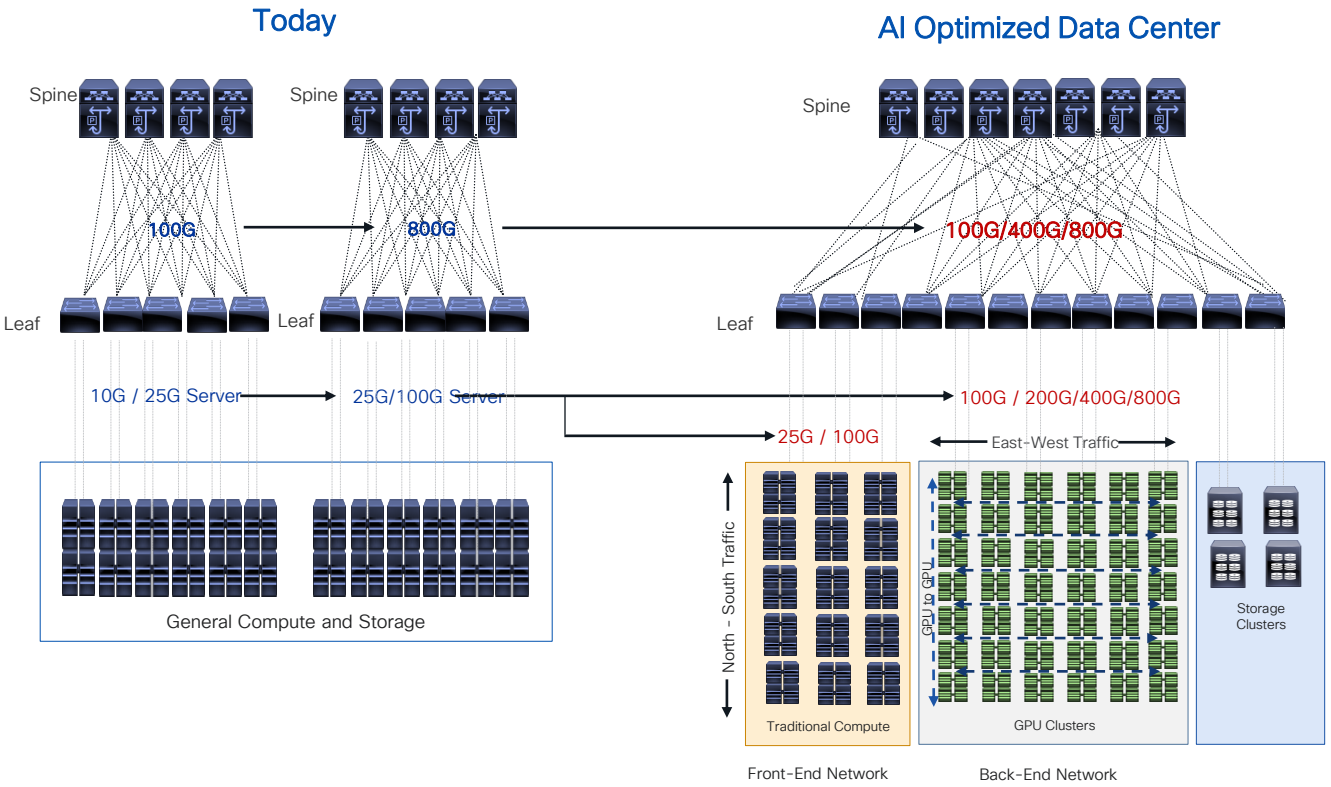
- Switch silicon bandwidth growing due to higher Radix and faster Serdes speeds
- Switch ASIC throughput growing: 6.4 Tbps to 12.8 Tbps to 25.6 Tbps to 51.2 Tbps
- Optics increasing from 40Gbps to 100G Gbps to 400Gbps to 800Gbps

Access



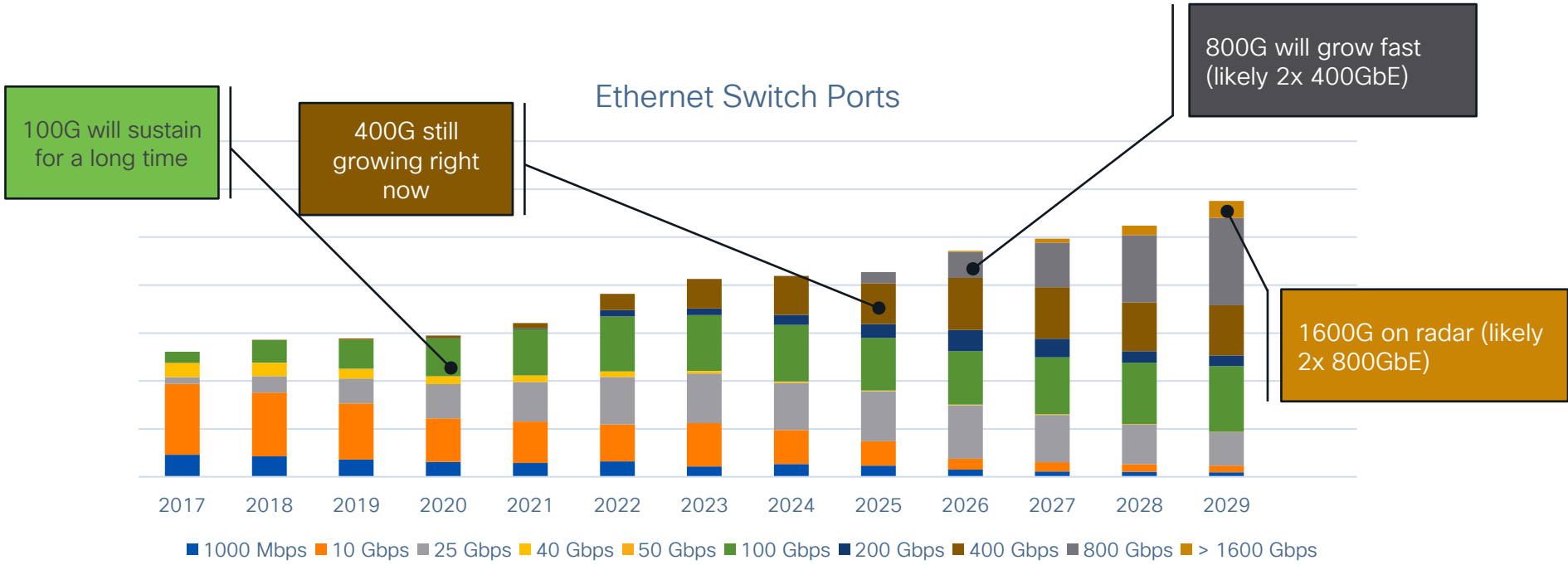
- Server network connectivity evolves with server processor upgrade cycles as data center traffic grows
- Server port speed is transitioning from 1/10 Gbps to 25 Gbps to 100 Gbps
- Storage, GPU, DPU, FPGA driving connectivity bandwidth, PCIe speed increase

Moving from a General-Purpose Compute Design to an AI/ML Optimized Infrastructure



- **AI/ML workloads:** pushing data centers to evolve their network architecture
- **AI-Specific Networking**: a dedicated Back-End network for AI workloads to isolate them from other data center traffic and ensure low-latency communication.
- **Back-End AI/ML clusters:** consists of hundreds to thousands of AI/ML accelerators, CPUs, storage devices, Switches, and Network Interface Cards (NICs) connected to GPUs
- **High-Speed Interconnects:** Backend network requires high speed 100G/200G or 800G optics to connect servers and network switches. These high bandwidth connections are essential for handling the data generated by AI workloads

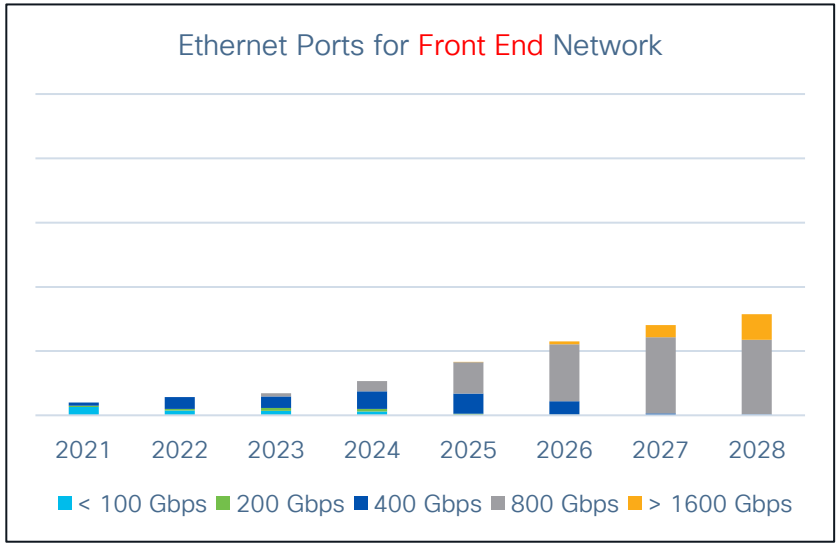
Ethernet Data Center Switch Port Speed Transitions

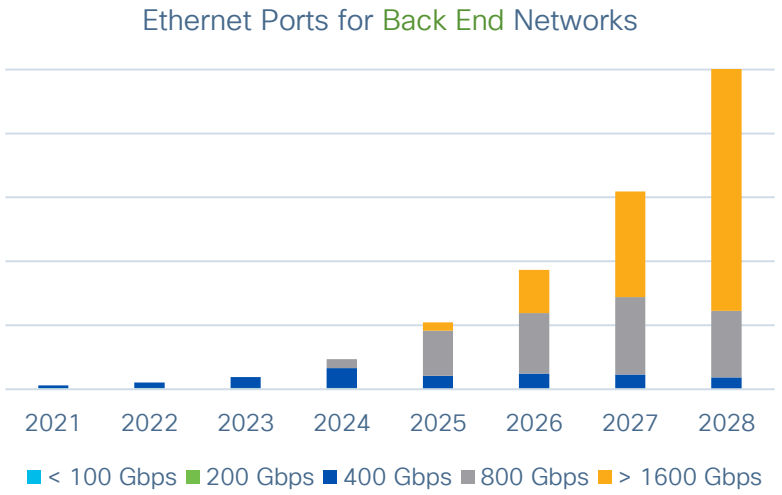


- Majority of the highest speed transitions are webscale (top 8) customers
- Webscale will drive the speed transitions quickly to scale
- Rest of the market will leverage that scale in their own time frame

Source: Dell'Oro's Ethernet Switch – Data Center Five Year Forecast Report 2025 – 2029

Ethernet Speed Transitions in AI Networks



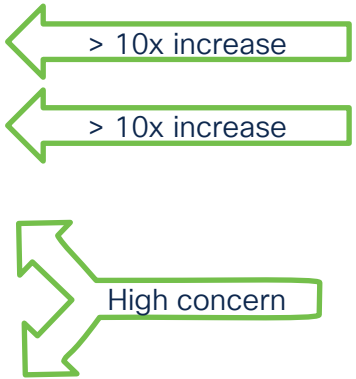


Majority of the switch ports in AI back-end Networks to be 800 Gbps in 2025 and 1600 Gbps in 2027, showing a very fast migration to the highest speeds available in the market.

Dell'Oro: 95N01_Advanced_Research_AI_Networks_For_AI_Workloads_Report_3Q24

AI/ML is a disruptive event for equipment design

		Traditional Front-end DC	AI/ML Back-end DC
Rack Bandwidth (ToR/MoR)		3.2T-12.8T	>> 100T
Rack power		~10 kW	100 kW+
Packet Loss impact (reliability)		Low importance	Critical importance
Latency importance	Absolute	Low	Low
	Tail	-	High



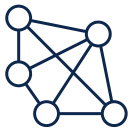
Optimal switch and interconnect design is affected by these requirements

400G/800G/1.6T use cases

Cloud service providers

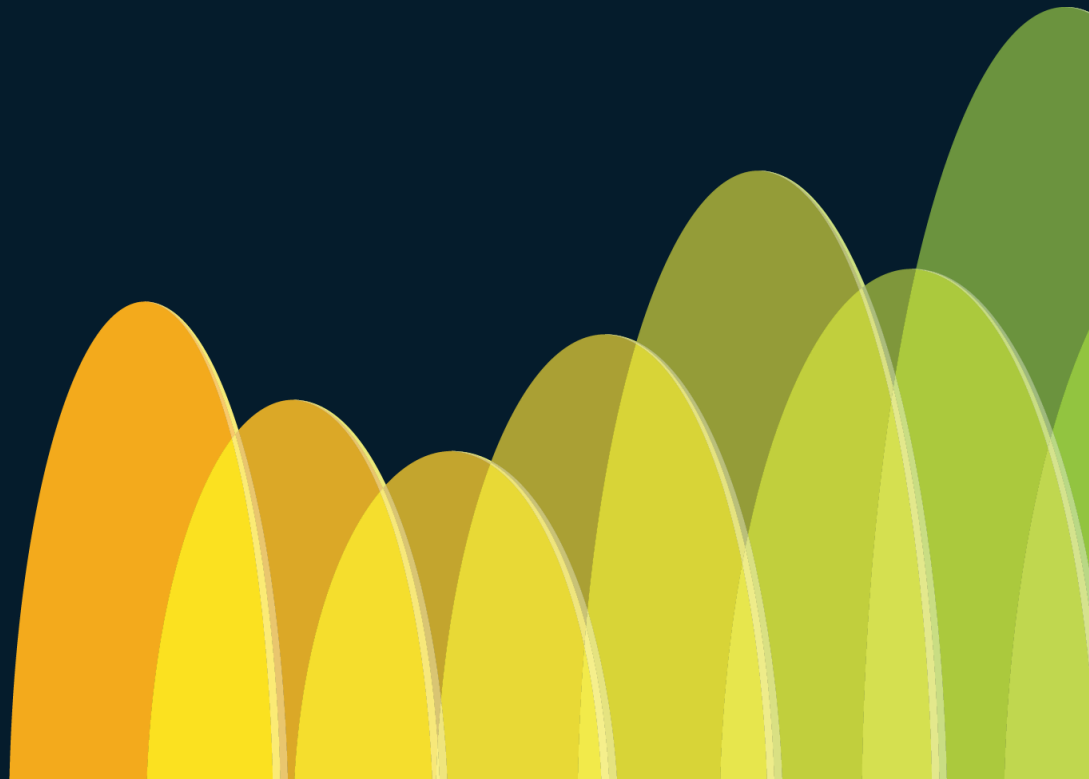
Telco service providers

Enterprise



Media networks

Deploying 400G & 800G



cisco Live!

Is 400G & 800G ready for broader adoption?

(AI generated pic)

The early progress of optics & interconnects is steered by the needs and use cases of the early adopters.

All adopters benefit from the economies of scale and technology maturity that results.

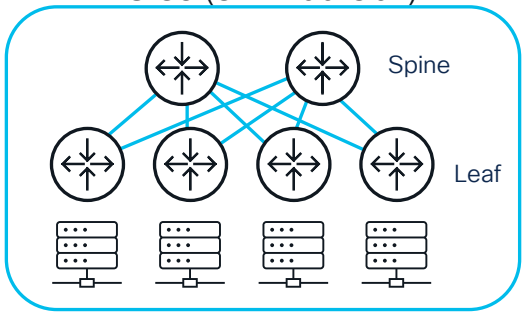
Looking at 400G/800G today we already have:

- 400 GbE & 800 GbE Ethernet defined
- Standardization of DCI interconnect (400ZR,400ZR+, 800ZR, 800ZR+)
- Denser switches and optics, breakout deployments

Why move to higher speeds?

400G → 800G example (same is true for 100G→ 400G)

25.6T user capacity using multiple switches with 12.8T ASICs (32x 400 GbE)



50 Gb/s ASIC IO (SerDes)

64 ports of 400GbE
(256 ports of 100 GbE)

~3000 Watts
26,280 kWh/year

=

25.6T user capacity using single switch with 25.6T ASIC (32x 800 GbE)

100 Gb/s ASIC IO (SerDes)

32 ports of 800G
(64 ports of 400 GbE
256 ports of 100 GbE)

~400 Watts
3,504 kWh/year

=

Up to 87%
Energy Savings

83% less space/fans

400G & 800G Today

Client (aka IMDD) (Short reach)		Coherent (Long Reach)
400 GbE, 800GbE	Mature standards	400G (OIF, Open ZR+)
100 Gb/s	Mature technology	400ZR: 16QAM @ 60 Gbd 800ZR: 16QAM @ 118 Gbd
800 GbE @ 200 Gb/s	Next Gen technology	1600ZR: 16QAM @ 240 Gbd
Incl. Breakout	Mature Deployment	400ZR, Bright 400ZR+

Most 400G/800G Standards are done

Cisco led many of these efforts

✓ Complete

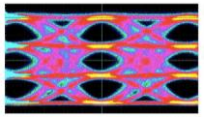
Cisco-led

Standards	IEEE 802.3bs ✓		400 GbE & 200 GbE MAC & Initial Interfaces
	IEEE 802.3cd ✓	cisco	50 GbE MAC & Interfaces (also 100 GbE & 200 GbE PMDs)
	IEEE 802.3cm ✓		400 GbE MMF (BiDi and SR8)
	IEEE 802.3cn ✓		Extended reach (40km) 50 GbE, 200 GbE, 400 GbE
	IEEE 802.3ct ✓		100GbE Coherent 80km
	IEEE 802.3cu ✓	cisco	100G-FR, 100G-LR, 400G-FR4, 400G-LR4-6
	IEEE 802.3ck ✓	cisco	100GE serdes
	IEEE 802.3db ✓		100/200/400GE MMF (100Gb/s short wavelength)
	OIF400ZR ✓	cisco	400 GbE Coherent 120km
	OIF800ZR	cisco	800 GbE Coherent 120km
	OIF1600ZR	cisco	800 GbE Coherent 120km
	Open ZR+ ✓	cisco	Long Distance 400G
	802.3df ✓	cisco	200G/400G/, 800G Ethernet Task Force @ 100Gb/s per lane
	802.3dj	cisco	200G/400G/800G/1.6T Ethernet Task Force @ 200Gb/s per lane
	802.3dk		Greater than 50 Gb/s Bidirectional Optical Access PHYs Task Force.
MSAs*	LPO MSA	cisco	100G Linear pluggable
	100G Lambda MSA ✓	cisco	100G-FR, 100G-LR, 400G-FR4, 400G-LR4
	QSFP-DD MSA ✓	cisco	400G/800G/1.6T Form factor
	OSFP MSA ✓	cisco	400G/800G/1.6T Form factor
	400G-BiDi MSA ✓	cisco	400 GbE MMF BiDi

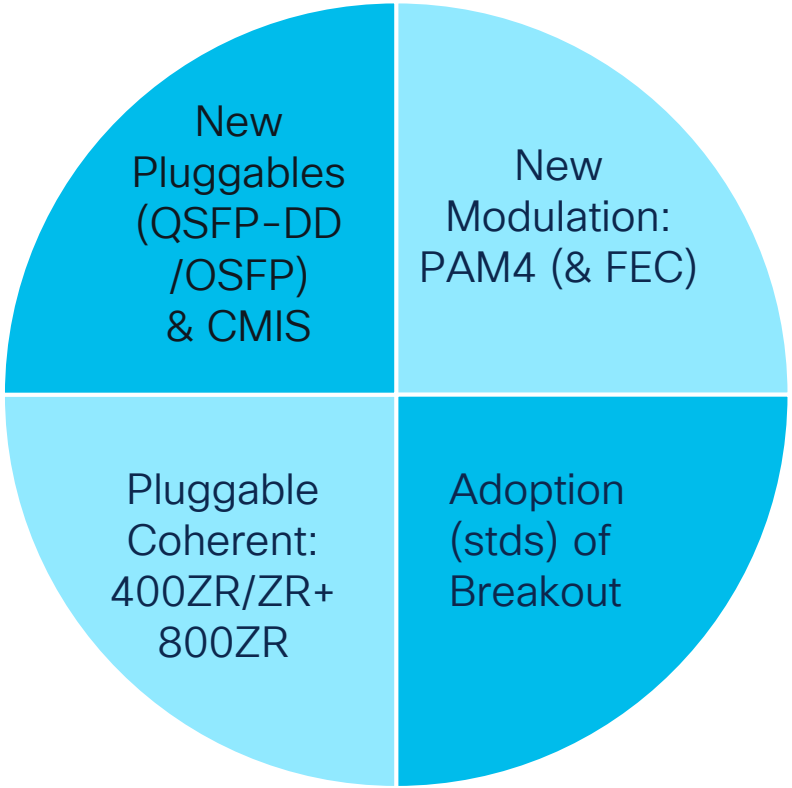
Flexibility of 400G / 800G pluggable modules

- Copper cables
- Multimode Fiber – 100m
- Single Mode Fiber inside DC – 500m & 2km
- Single Mode Fiber Campus – 10 km
- Outside plant, DCI – 100-1300 km

Key elements of 400G & 800G optics



New Pluggables with 8-wide connector to support 400G and 800G ports

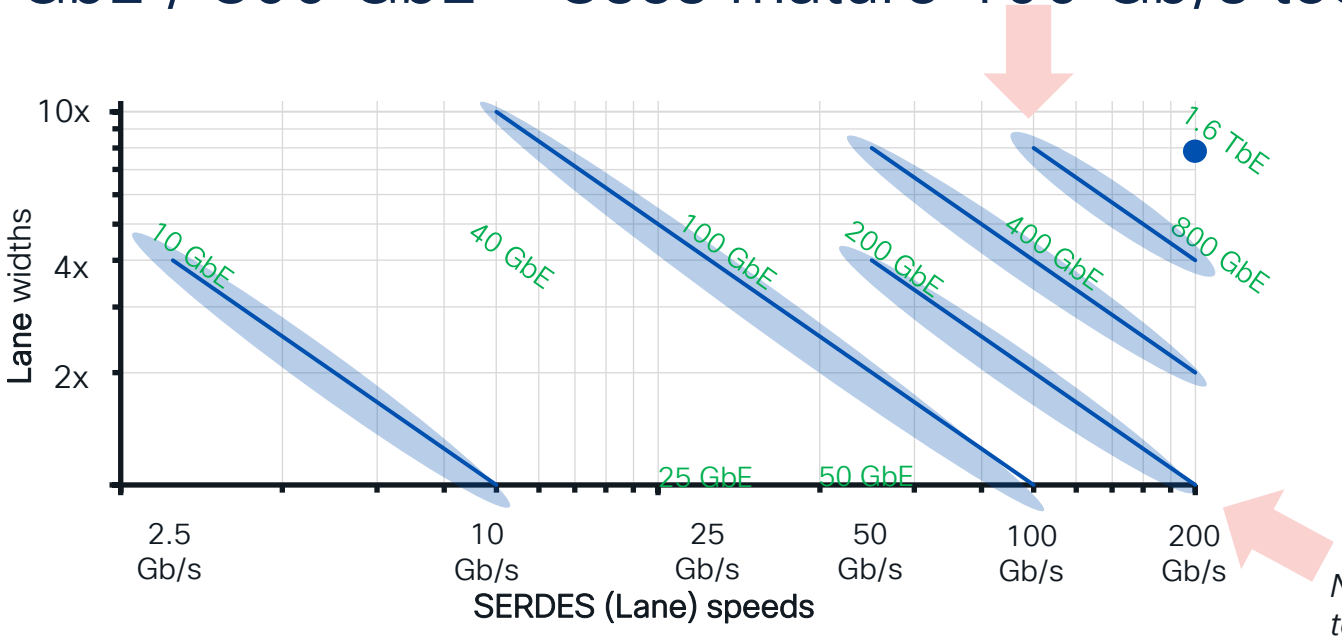


Higher speed interfaces adopted PAM4 modulation.
Ubiquitous use of FEC.

Long reach coherent without any system port density reduction
→ Routed Optical Networking

Pluggable modules supporting multiple lower speed interfaces

400 GbE / 800 GbE – Uses mature 100 Gb/s technology



Newer 200Gb/s technology in development now. See later.

Technology maturity of 100 Gb/s technology enables:

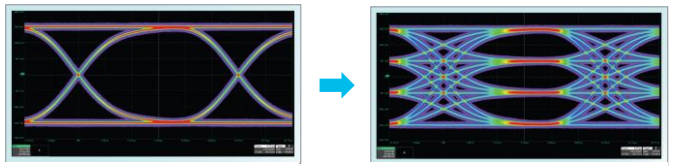
- Excellent performance
- Excellent interoperability
- Broad adoption

Technologies used for 400G, 800G optics

Enabling higher performance but lowering cost

Advanced Modulation

Client optics (IMDD)



NRZ
1 bit/sec/symbol

PAM4
2 bit/sec/symbol

Coherent optics

Coherent QPSK
2 bit/sec/symbol

Coherent 16-QAM
4 bit/sec/symbol

Using more complex modulation allows us to increase the data rate (Gb/s) without increasing the signaling speed (Gbaud)

Forward Error Correction (FEC)

Allows correction of errors at receiver

Enables use of relaxed specs (saves $) to get same performance or enables much higher performance.

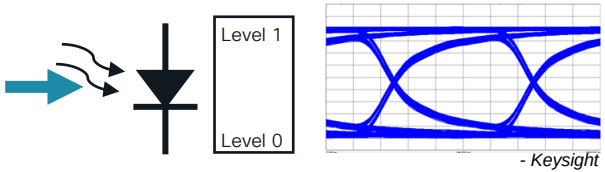
Usually embedded in Ethernet switch or module DSP ASIC

Some overhead (faster data rate needed)

Detection: Direct Detect (IMDD) vs Coherent Detect

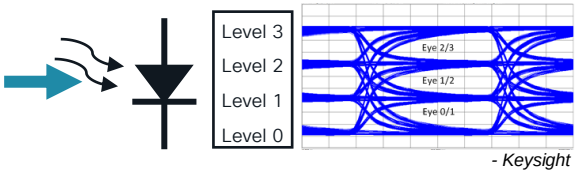
Direct Detection of Intensity Modulation

NRZ



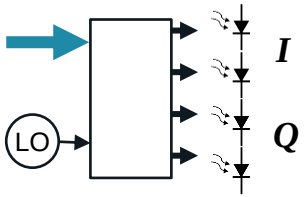
- Detect intensity envelope
- Simple implementation
- Little receive DSP

PAM4



- Detect multi-level intensity
- Medium design complexity
- High-speed ADC/DACs and DSP typically utilized

Coherent Detection



- Detect amplitude & phase of light
- More advanced DSP implementation
- DSP also provides digital compensation of optical impairments enabling much longer reaches

Common Management Interface Spec (CMIS)

Standardized now in OIF – ubiquitous adoption for 400G modules and above

Host (switch/router)

High speed electrical interface
(OIF CEI, IEEE, etc)

Pluggable Module

Optical/Copper
(IEEE, MSAs, OIF)

Management
- configure
- initialize
- monitor

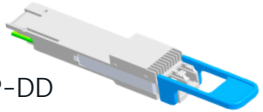
Often overlooked, CMIS standardizes management between modules and hosts

Consistent definition for configuration, initialization, monitoring, telemetry, firmware update etc.

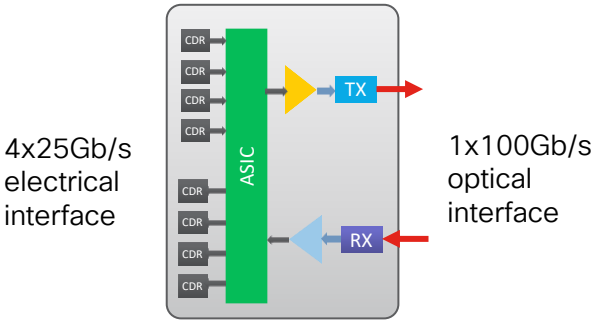
Migration to 100G Single Lambda Optics

100G Lambda MULTI-SOURCE AGREEMENT

100G QSFP28

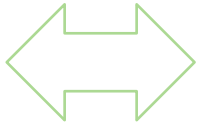


400G or 800G QSFP-DD

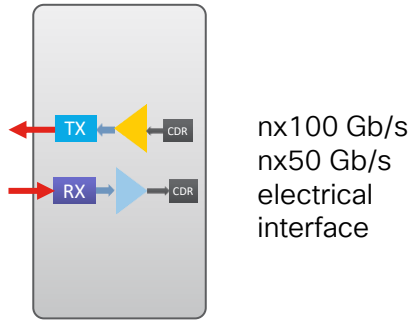


4x25Gb/s electrical interface

1x100Gb/s optical interface



nx100Gb/s optical interface



nx100 Gb/s nx50 Gb/s electrical interface

As your network evolves, the optical interface is THE point of interoperation.

Equipment and electrical serdes can evolve through 3 generations (25 Gb/s, 50 Gb/s or 100 Gb/s) without changing the optical interface that interconnects your equipment.

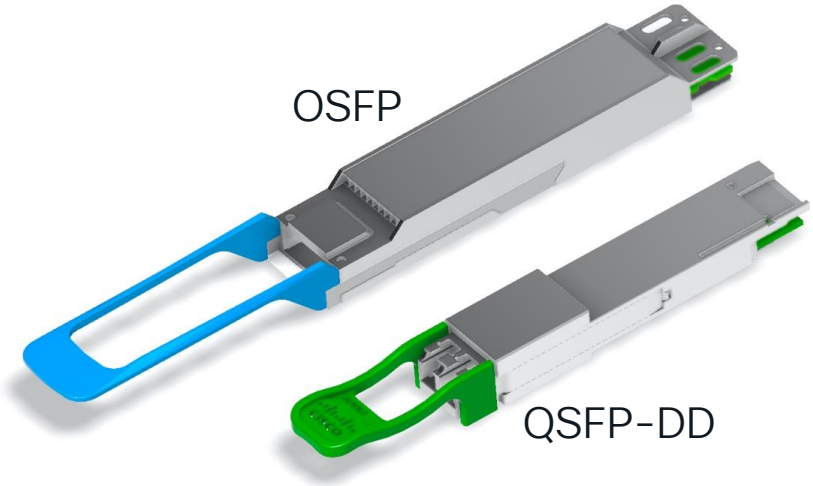
Module Roadmap for 200/400/800G

Roadmap
Orderable
Q2CY25

Distance	1-3+ m*	50/100 m	500 m-2 km	10 km	100+ km
200G QSFP56	QSFP-200-CUxM QDD-2Q200-CUxM	QSFP-200G-SL4 QSFP-200G-SR4-S	QSFP-200G-FR4-S		
400G QSFP-DD	QDD-400-CUxM QDD-2Q200-CUxM QDD-4ZQ100-CUxM QDD-400-AOCxM	QDD-400G-SR8-S QDD-400G-SR4.2-BD QDD-400G-VR4	QDD-400G-DR4-S QDD-4X100G-FR-S QDD-400G-FR4-S	QDD-400G-LR4-S QDD-2X100-LR4-S	QDD-400G-ZR QDD-400G-ZRP
400G QSFP112		QSFP-400G-VR4	QSFP-400G-DR4 QSFP-400G-DR4-2 QSFP-400G-FR4		
800G QSFP-DD	QDD-800-CUxM QDD-2Q400-CUxM	QDD-800G-VR8	QDD-800G-DR8 QDD-8X100G-FR QDD-2X400G-FR4	QDD-800G-DR8-10 QDD-2X400G-LR4	
800G OSFP/R	OSFP-800-CUxM OSFP-2Q400-CUxM OSFP-4Q200-CUxM	OSFP-800G-VR8 OSFPR-800G-VR8	OSFP-800G-DR8 OSFP-800G-DR8-2 OSFP-2X400G-FR4 OSFPR-800G-DR8	OSFP-800G-DR8-10 OSFP-2X400G-LR4	

New FF

Pluggable Optical Modules: QSFP-DD or OSFP



400G/800G/1.6T variants

Both variants support all the technical requirements:

- 32 ports in 1 RU is feasible
- 64 ports in 2 RU is feasible

QSFP-DD is compatible with QSFP

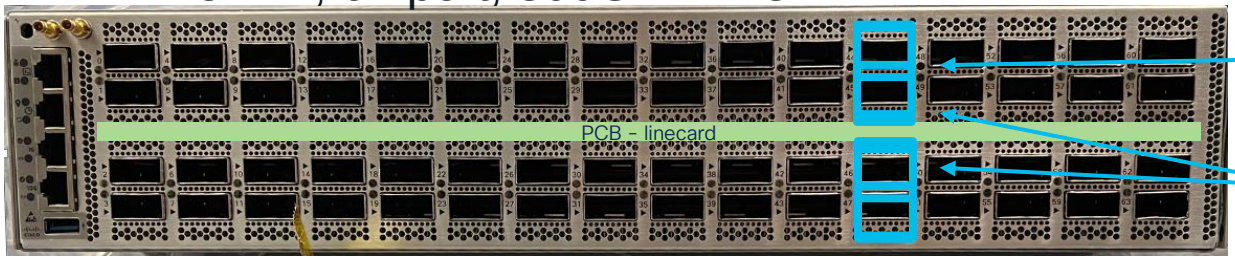
- QSFP-DD enables better system co-design which has thermal advantages

Breakout optical connector options[1]

[1] only QSFP-DD shown but similar on OSFP

800G Optical Modules: QSFP-DD or OSFP

51.2T, 64 port, 800G in 2RU



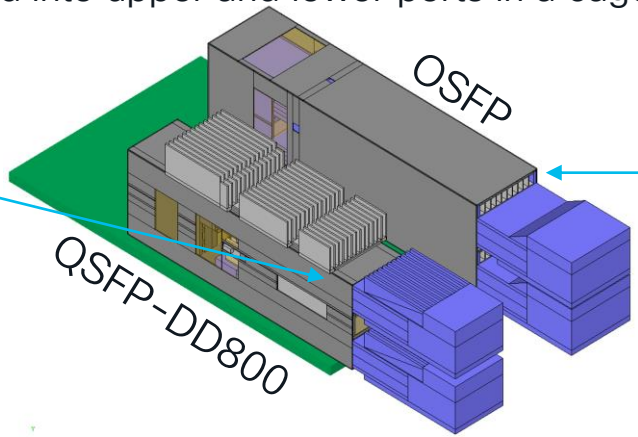
Stacked cages (two modules)

Both above and below the linecard

Showing two modules inserted into upper and lower ports in a cage. Heatsink differences are the key.

Module slides under heatsinks.

- Heatsink design can change and optimize per equipment design



Heatsink is integrated with module

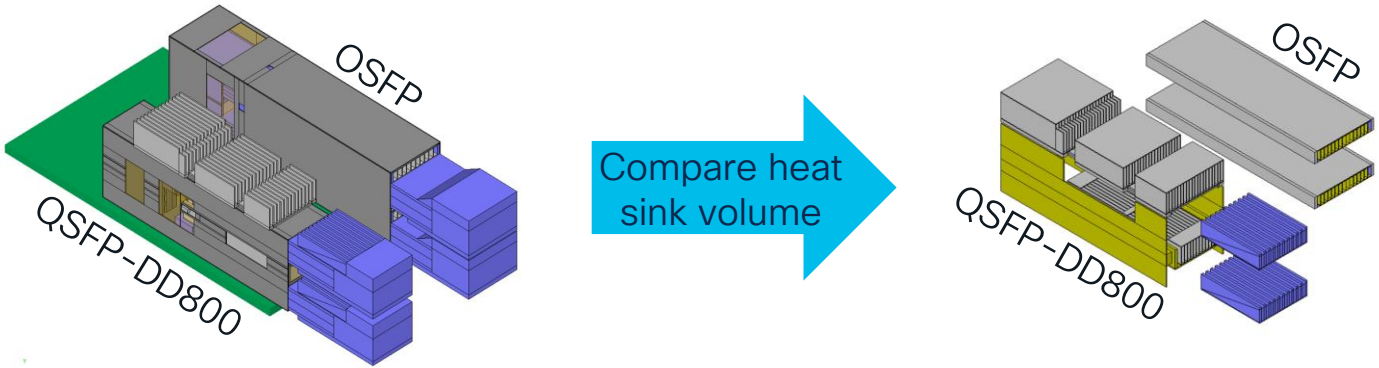
Upper port

Lower port

QSFP-DD – avoids the “risk” of the integrated heatsink

QSFP-DD’s flat top (and bottom) case design: allows systems to optimize riding heat sinks.

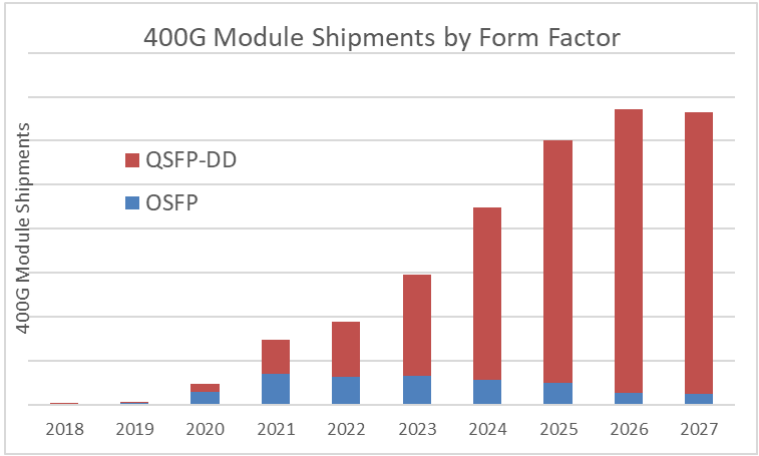
- *This fundamentally is the BREAK of that intuitive perspective*
- In an air-cooled system, the best way to cool anything is to increase the heat sink fin area.
- QSFP-DD is innovation friendly: yields constant improvements to cage design, faceplate design, heatsink design, air flow balance etc. Friendly to ASIC cooling
- While an integrated heat sink seems better, a QSFP-DD design can have nearly 3x the heatsink volume. *OSFP heatsink was defined in ~2015 and can’t be changed*



Current deployment trends

400G:

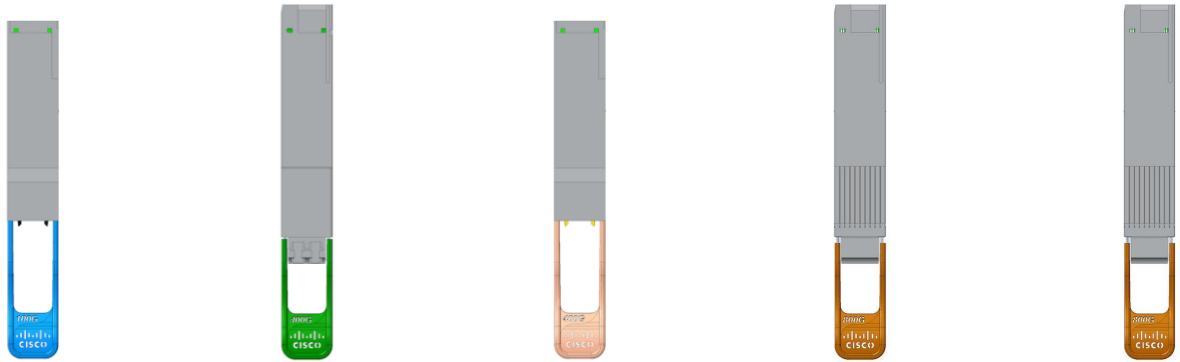
- QSFP-DD saw broad adoption across all markets (Hyperscaler → Enterprise)



800G:

- OSFP adoption growing. The perceived advantage of integrated heatsink swayed decisions at early adopter hyperscalers
- Nvidia NICs based on OSPF-RHS
- Survey of Cisco's Enterprise DC customers shows strong interest in QSFP-DD due to backwards compatibility

QSFP-DD Extends to 1.6T



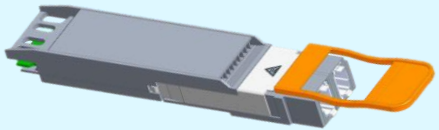
100 Gb/s	400 Gb/s	400 Gb/s	800 Gb/s	1600 Gb/s
25 Gb/s SerDes	50 Gb/s SerDes	100 Gb/s SerDes	100 Gb/s SerDes	200 Gb/s SerDes
QSFP28	**QSFP56-DD**	**QSFP112**	**QSFP-DD800**	**QSFP-DD1600**

Complete backwards compatibility

Powerful value proposing to enable seamless network growth and investment protection. Technically superior solution.

Pluggable form factors – a wealth of options

OSFP

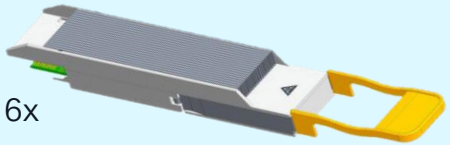


400G/800G/1.6T

OSFP-RHS

400G/800G/1.6T

OSFP-XD



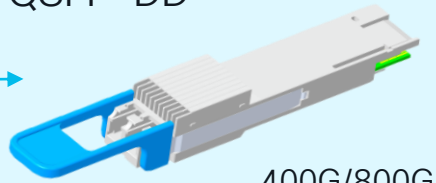
16x

1.6T

Switches

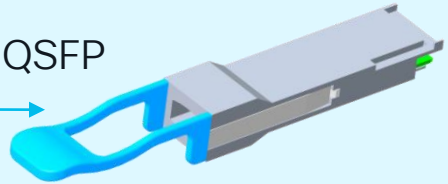
Compute NICs

QSFP-DD



400G/800G/1.6T

QSFP



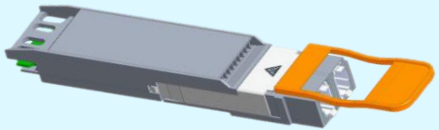
100G/200G/400G
QSFP28, QSFP56, QSFP112

Port compatibility

Everything within the square is compatible with each other

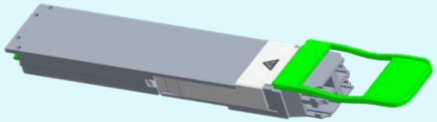
Pluggable form factors – interface interoperability

OSFP



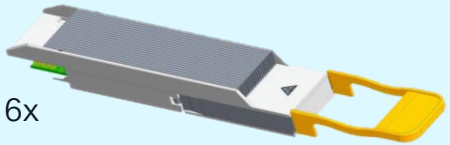
400G/800G/1.6T

OSFP-RHS



400G/800G/1.6T

OSFP-XD



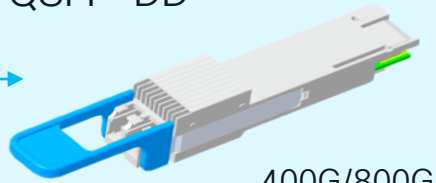
16x

1.6T

Ethernet plug and play as normal.

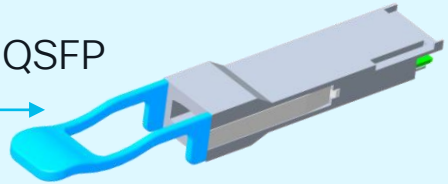
Breakout as well!

QSFP-DD



400G/800G/1.6T

QSFP

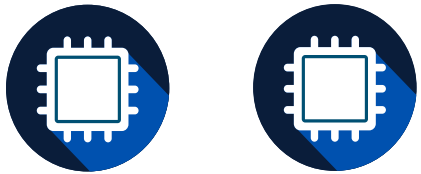


100G/200G/400G
QSFP28, QSFP56, QSFP112

Ethernet Interface compatibility

Port compatibility
Everything within the square is compatible with each other

800G does not necessarily mean 800 GbE



25.6T and 51.2T ASIC roadmap and system density requirements drive to 800G ports

- Dominant interest in 800G module today is to support 2x 400 GbE breakout
- No immediate initial network demand for 800 GbE
- 800G modules have the same issues:
  - Thermals and Signal integrity
  - Backwards compatibility?

Fixed Configs

25.6T	1 RU	32p @ 800G
		QSFP-DD800, $OSFP_{800G}$
51.2T	2 RU	64p @ 800G
		QSFP-DD800, $OSFP_{800G}$

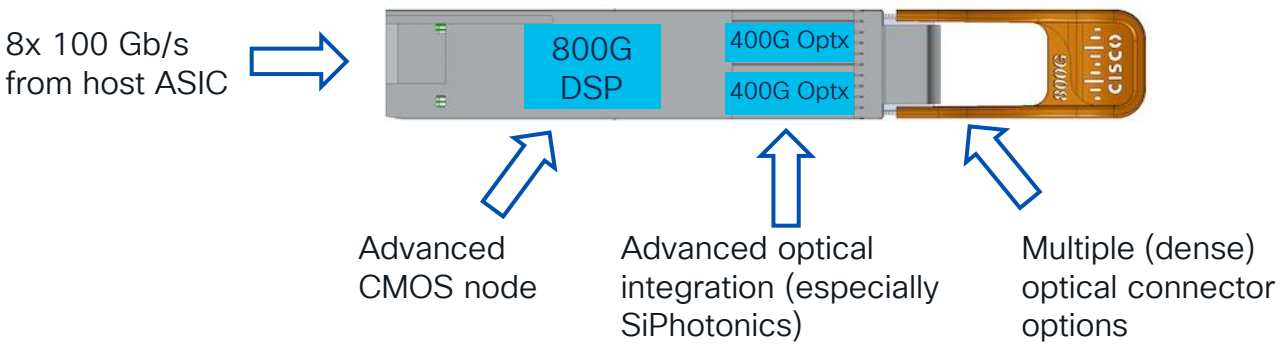
800G Pluggables supporting dense 400 GbE (aka breakout)

QSFP-DD

Both 400G & 800G form factor enables an economical way to implement breakout to lower speed Ethernet interfaces. This maximizes

- Cost and density on one end of link
- Compatibility and return on investment with existing equipment

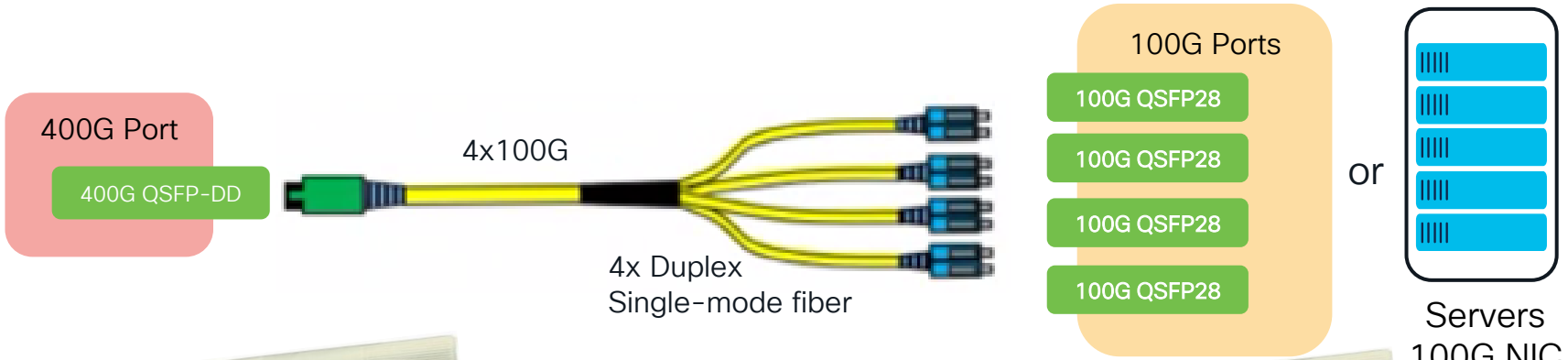
Example: 800G module



What is a Breakout?

- Breakouts take advantage of ports with multiple optical lanes for both the Tx and Rx
  - Optical lanes in this context means pairs of fibers
  - e.g. 400G DR4 optical connector has 4 pairs of fibers, each pair can be configured as a 100G-DR
- A breakout is when a port (module) is configured as multiple lower speed interfaces
- Breakout transceivers generally use MPO connectors which have multiple fibers for both the Tx and Rx
  - The port controls how the module will be configured either for breakout or non breakout operation
  - For 2x breakout, module's support regular LC connectors too
- Breakouts can also be done for copper cables and AOCs
  - Cables and AOCs are fixed for either breakout or non-breakout applications

Breakout enables operational efficiency



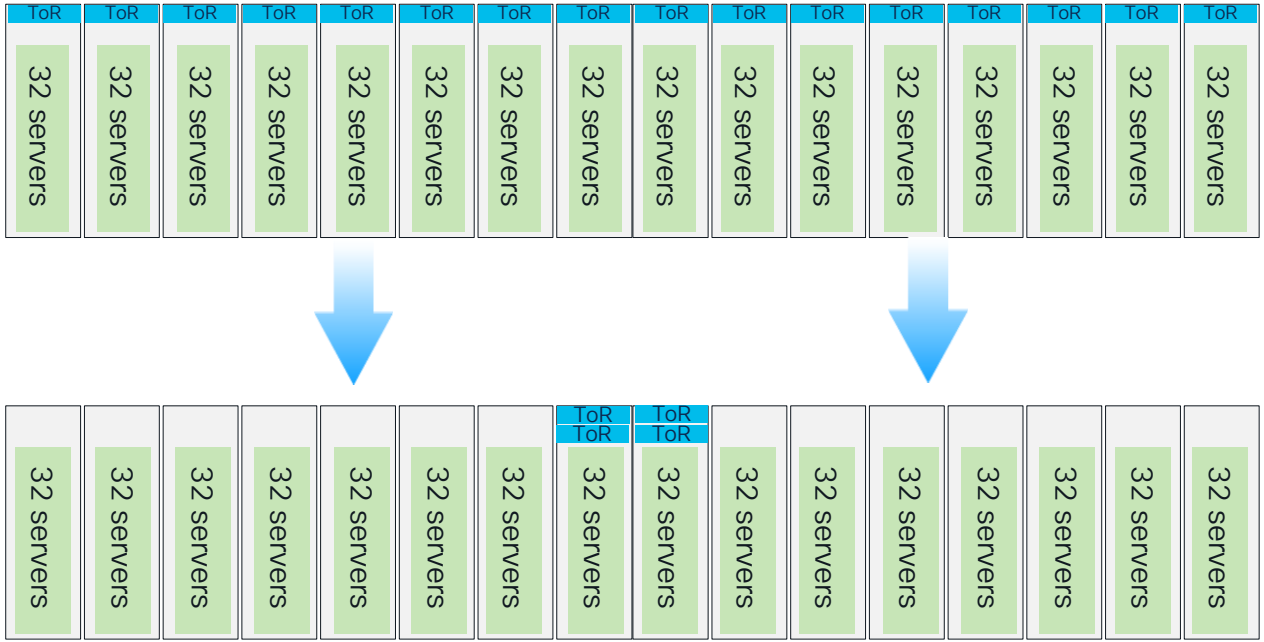
400G DR4 optics
Power: 12W
Space: 1 port

100G DR optics
Power: 4.3W x 4 = 17.3W
Space: 1 port x4 = 4 ports

Savings: 8 logical 100G ports with only 5 Physical ports → 5.3W power savings

Breakouts Promote Sustainability

32 1RU servers per cabinet w/ 16 cabinet row



Non Breakout Architecture
- 16 TOR = 811W x 16 = 12976W
- 512 DAC 0.5W per end = 512W
- Total power 13,744W

Breakout Architecture using 64 port switches
- 4x switch: 1324W x 4 = 5296W
- 128 DR4 modules 9W x 128 = 1152W
- 100G-DR modules 3W x 512=1536W
- Total Power 7984W

40% savings in power

Energy optimized architectures using breakout are explored in this whitepaper.

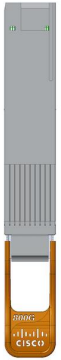
Implementing Dense 400 / 800 GbE

QSFP and QSFP-DD allow wide range of deployment options between switch and compute equipment with compatible modules

100 Gb/s
25 Gb/s SerDes
QSFP28

400 Gb/s
50 Gb/s SerDes
QSFP56-DD

400 Gb/s
100 Gb/s SerDes
QSFP112

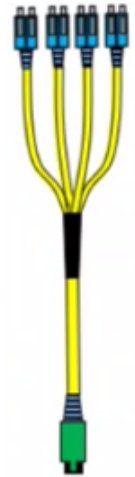


800 Gb/s
100 Gb/s SerDes
QSFP-DD800

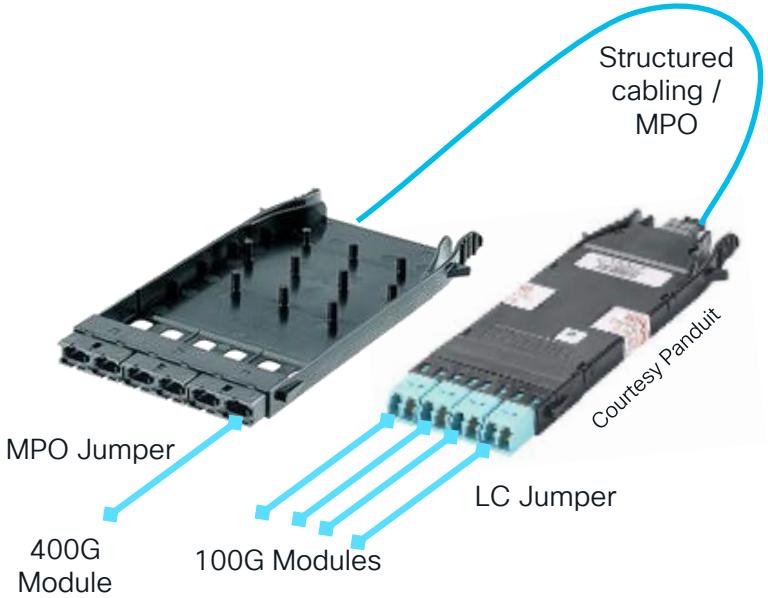
100 GbE	1x	4x	4x	8x
200 GbE	-	2x	2x	4x
400 GbE	-	1x	1x	2x
800 GbE	-	-	-	1x

Deploying Breakout

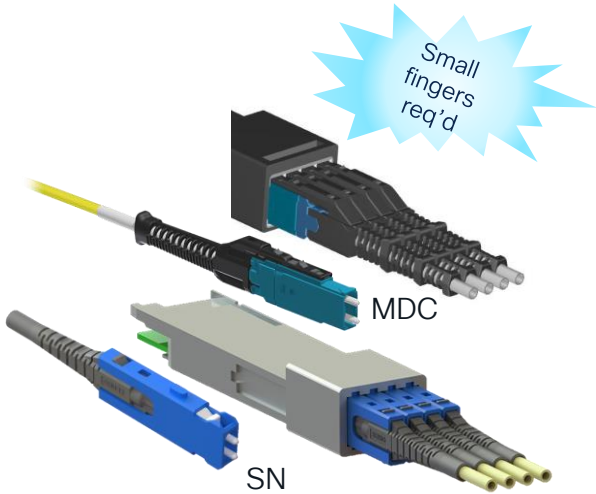
- Multiple options exist



Breakout cables



Using structured cabling and breakout cassettes



Dense VSFF connectors mostly used in patch panels

Network Operator Hurdles for Breakouts

- Reliability – What if one lane fails, do I have to interrupt non-affected traffic?
  - Nearly all systems design include redundancy or fail-over paths
  - Cisco optical modules have world class reliability
    - Typical field failure rate <<100ppm
  - Breakout modules do not have higher field failure rates than single lane modules

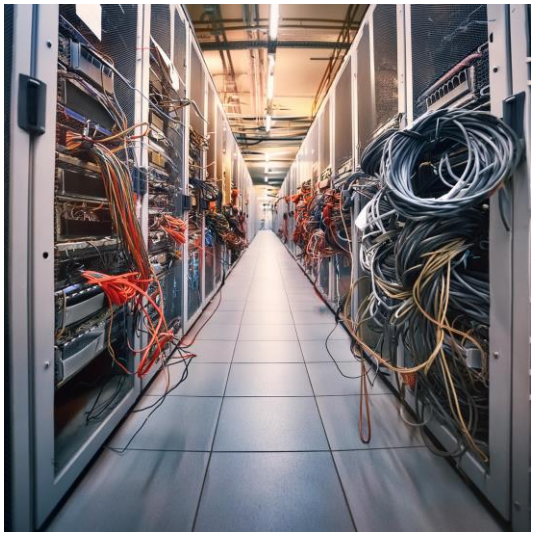
- Fiber management – Managing more connections in front of the switch will be a nightmare
  - Cisco sells a high-density patch panel system that eases fiber management
    - Supported by Cisco!

- Why transition to a higher BW switch now?
  - Manage migration to higher speed with the latest features on a port-by-port basis.
  - Ports are backwards compatible and can accommodate existing optics
  - Use breakout modules to connect to lower speed ports while network evolves
  - Use higher speed modules when needed as the rest of the network is upgraded
  - Reduce power consumption for equivalent bandwidth

Managing Breakout Cabling: Patch Panels

Recommended to use high density breakout patch panels

Can address any concerns about fiber management with breakouts



When you say “Breakout”, is this what you are thinking? (AI generated pic ☺)

Example using Cisco patch panels

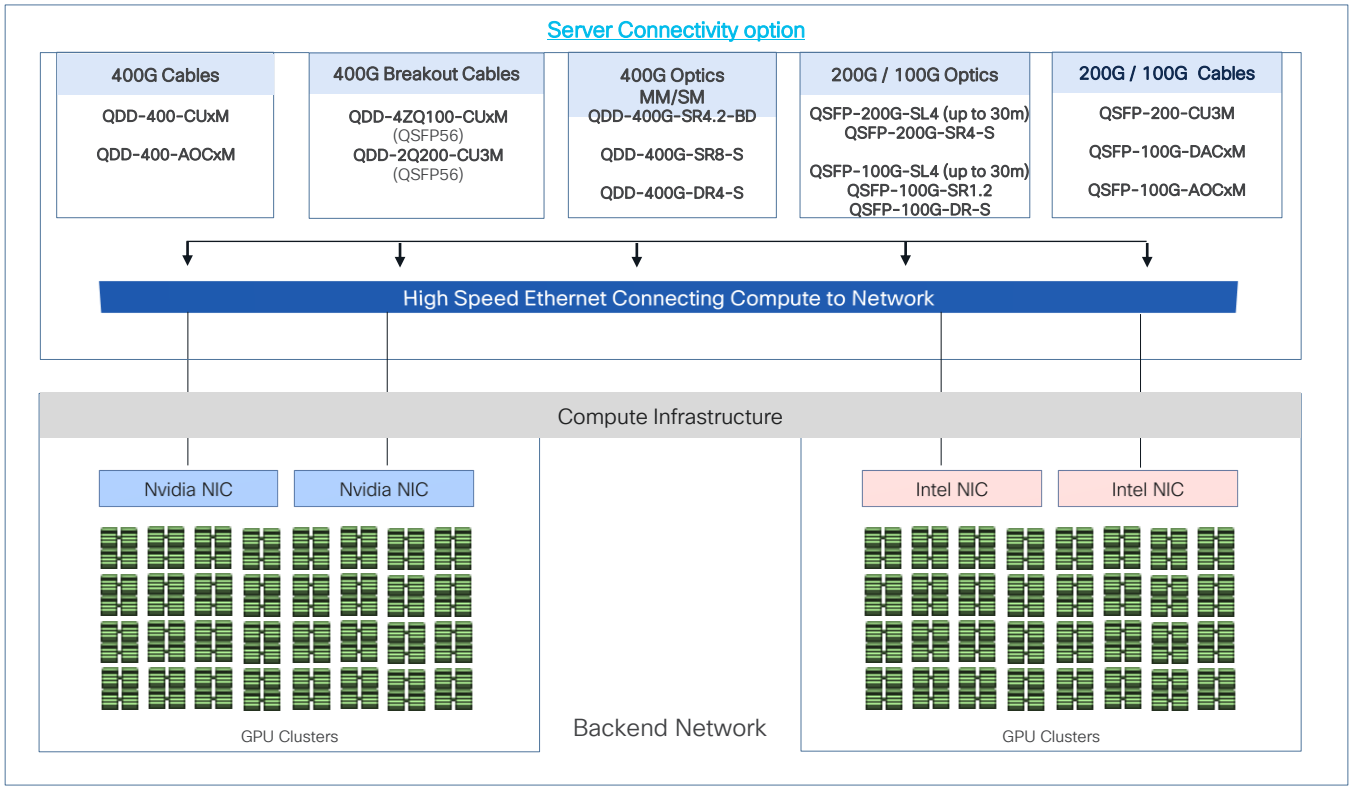
- 24 RU’s of patch panels
- 432 jumpers

Supporting :

- Cisco 8812 Chassis (172 Tbps)
- 12 Line Cards
- 36 Ports of 400G
- 4x100G Modules broken out
- 1728 x 100G connections

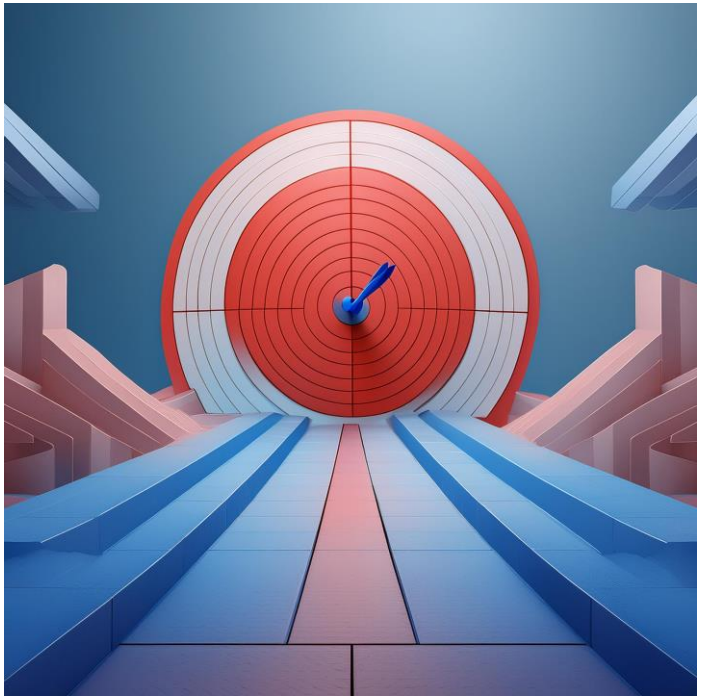
Maintain the same chassis and migrate each port to higher speed interfaces as needed

AI/ML Optics Connectivity Options for Compute



- A need for high-speed optics in compute to support high bandwidth GPU clusters
- Fewer servers per rack due to power and heat constraints (GPU)
- More inter-rack communication due to less servers in the rack
- Active Optical Cables (AOCs) and Pluggable Short Link (SL) optics for reaches of up to 30m
- 400G and 100G BIDI for reaches exceeding 30m
- Use of Direct Attach Cables (DAC) for smaller clusters

Key takeaways – client (IMDD) reaches



(AI generated pic)

400G & 800G Standards are mature. Ethernet interop is guaranteed regardless of form factor

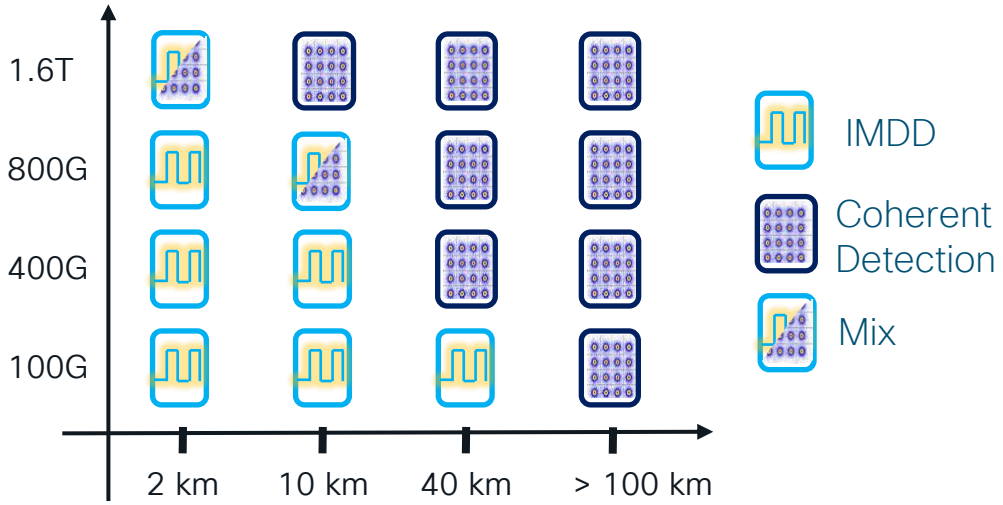
Optics and copper interconnect available in all reaches

Cisco has a wide variety of high-density transceiver and cabling breakout solutions

Sustainable and simplified deployment options are available

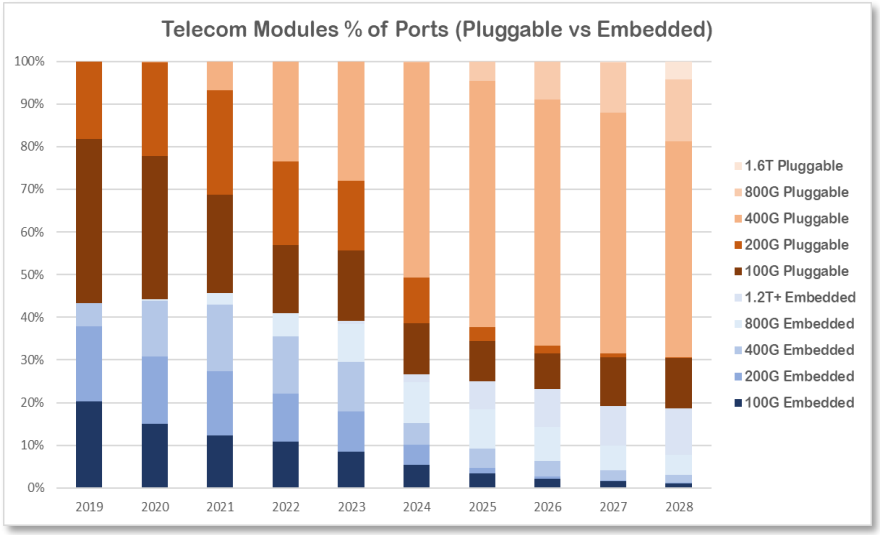
Coherent optics addresses higher reaches (and speeds)

- IMDD and Coherent technology will continue to be used
  - 800G, 1.6T and beyond
  - Coherent pushing towards shorter reaches. Not only in DWDM but also Grey applications
  - Focus shifting from performance enhancements ($$$) to interoperable interfaces and pluggables
- Coherent multivendor Interop more and more prevalent
  - 100G, 400ZR/ZR+, 800ZR/ZR+

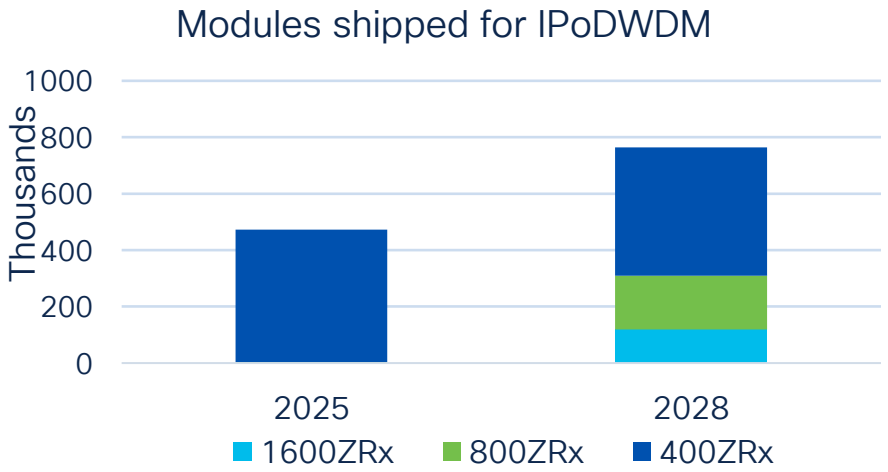


As speed increases, the crossover distance between IMDD & coherent decreases

The rise of “Coherent Pluggables”



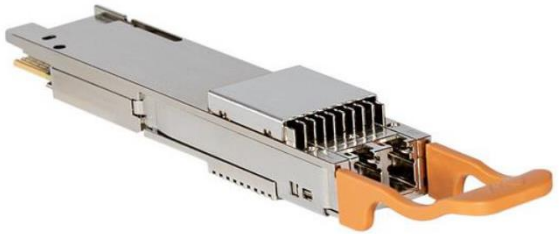
Today >70% of the coherent ports are pluggable coherent optics



Source: Cignal AI, IP-over-DWDM Pluggables Forecast, Aug. 2024

Cisco/Acacia is pioneer and market leader for coherent pluggables

Coherent pluggable brings some important levers



Drives Interoperability
MSAs/standards, best in breed adoption

01

Network Costs & Efficiencies
Capex costs breakdown, $/100G
PAYG models etc.

02

Power, Thermal & Space
Compact devices, Watt/100G, DC power and rack space optimization.

03

Network Architectures
Disaggregation, Optimize topologies, Spectrum efficiency

04

Approaches of Coherent Implementation

2014 → 2016 → 2020 → 2024

MSA Pluggable (aka DCO)

Industry's first pluggable coherent module (100G CFP)

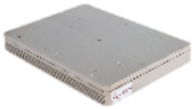
100G/200G CFP2-DCO

400G pluggable portfolio in QSFPDD, OSFP and CFP2

800ZR/ZR+ and 400ULH pluggables – QDD & OSFP

Deployed in router/switches and transport platforms (e.g., 8K, ASR9K, NCS, Nexus 9K etc.)

Multi-Haul Performance Optimized



400G multi-haul module (AC400)

1.2T Coherent Module (AC1200)

Industry's first 1.2T coherent multi-haul pluggable (CIM8)

Deployed in transport platforms (e.g., NCS 1004, 1014 etc.)

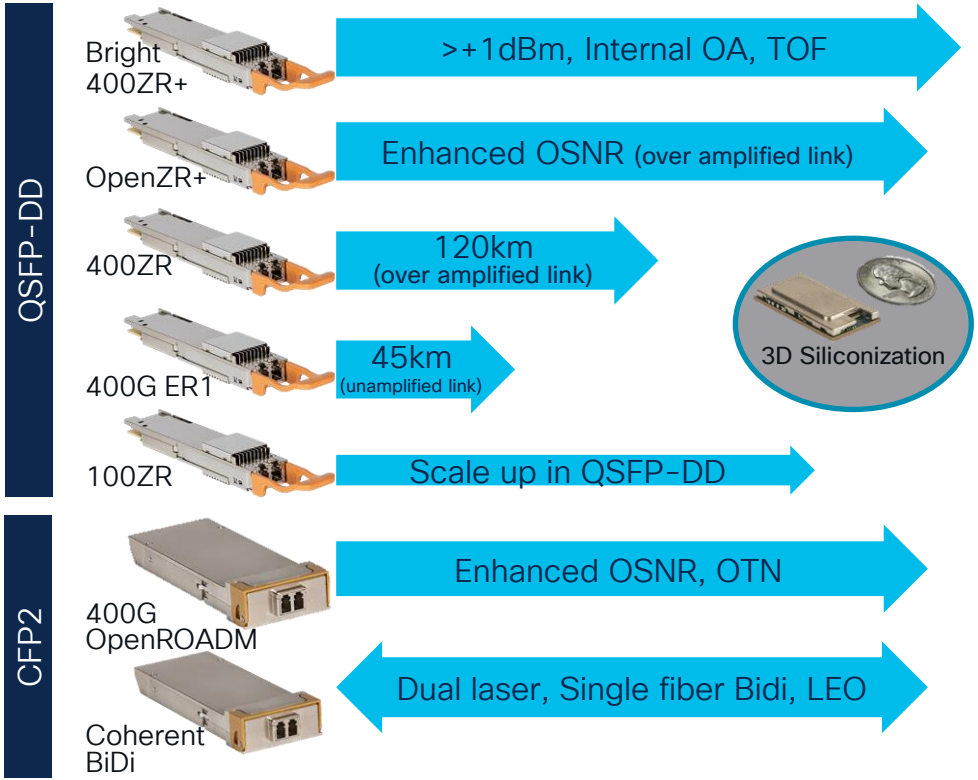
Comprehensive 400G MSA Pluggables Portfolio

Industry leading 400G pluggable deployments

Best in class quality

Standards compliant

OIF 400ZR · OpenROADM · OpenZR+ MSA · ITU · CableLabs · IEEE



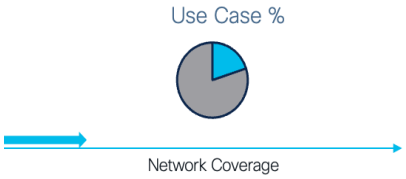
400G QSFP-DD Digital Coherent Optics Portfolio

ER1

Point to Point
Intra-data center, campus interconnect, core-to-edge router

Lowest Cost
Based on fixed laser with simple point-to-point connectivity

Short Reach
Up to 45KM for unamplified at 13dB

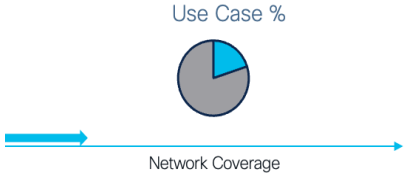


ZR

Point to Point
Web, Data Center Interconnect, Non-SP/SP router interconnect

Low Cost
Lowest cost 400G DCO option for very simple designs

Short Reach
Up to 120KM for P2P amplified links

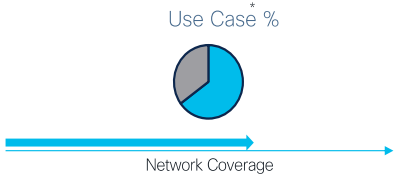


Bright ZR+

Transport Centric
Service Providers, Routed Optical Networking

High TX Power
+1dbm for difficult spans; interop with brownfield transponder & legacy line systems

Advanced Features
TOF, OTN and L1 Encryption features; integrated optical amplifier

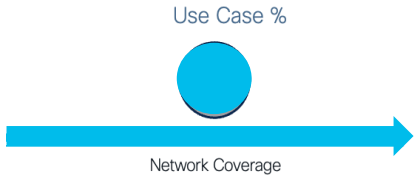


ULH

Transport Centric
Service Providers, Routed Optical Networking

Long haul optimized
400G with **RON Anywhere** up to 3000km, high Tx power

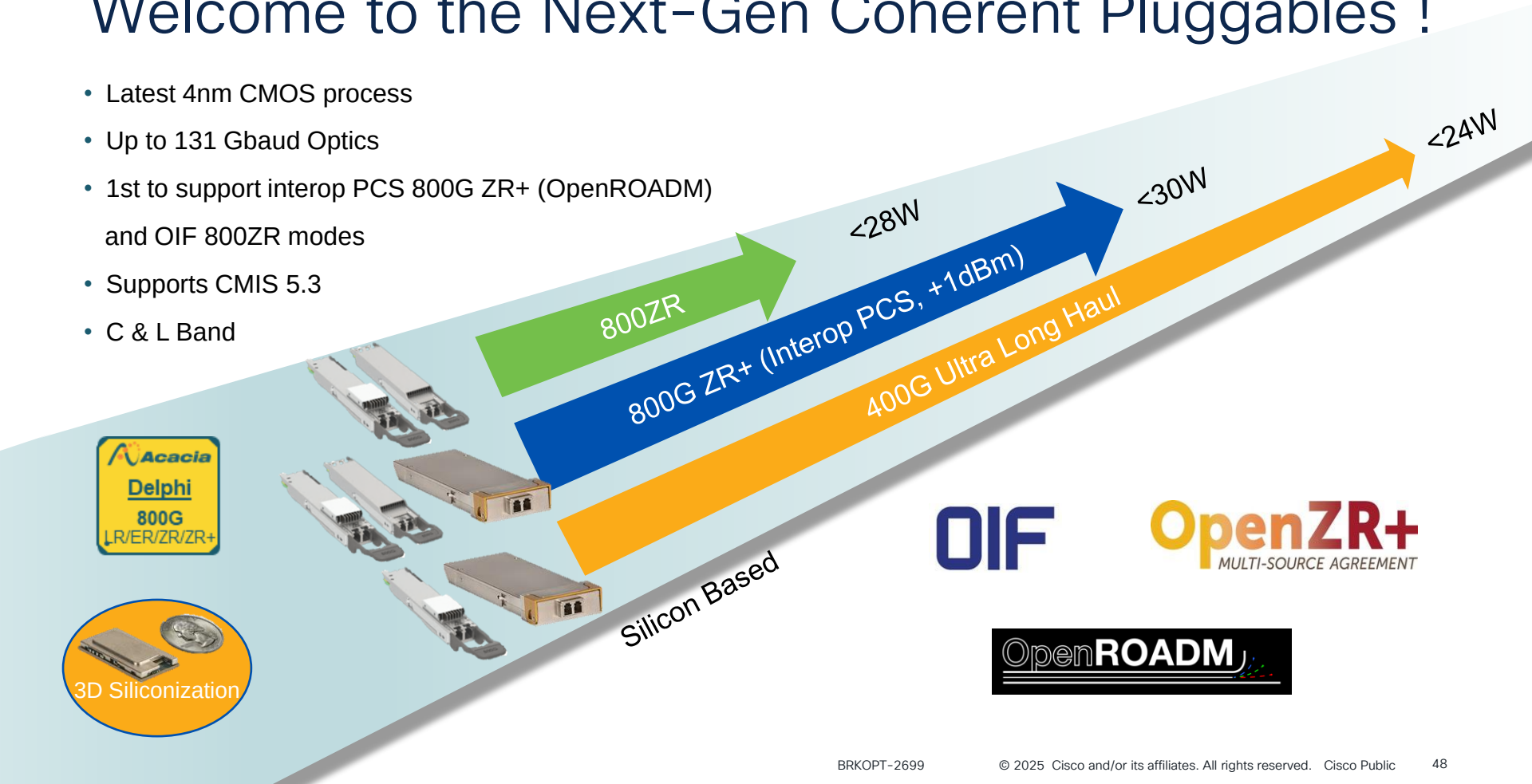
Power optimized
Designed to fit in all routers support 400G Bright ZR+ today



cisco Live!

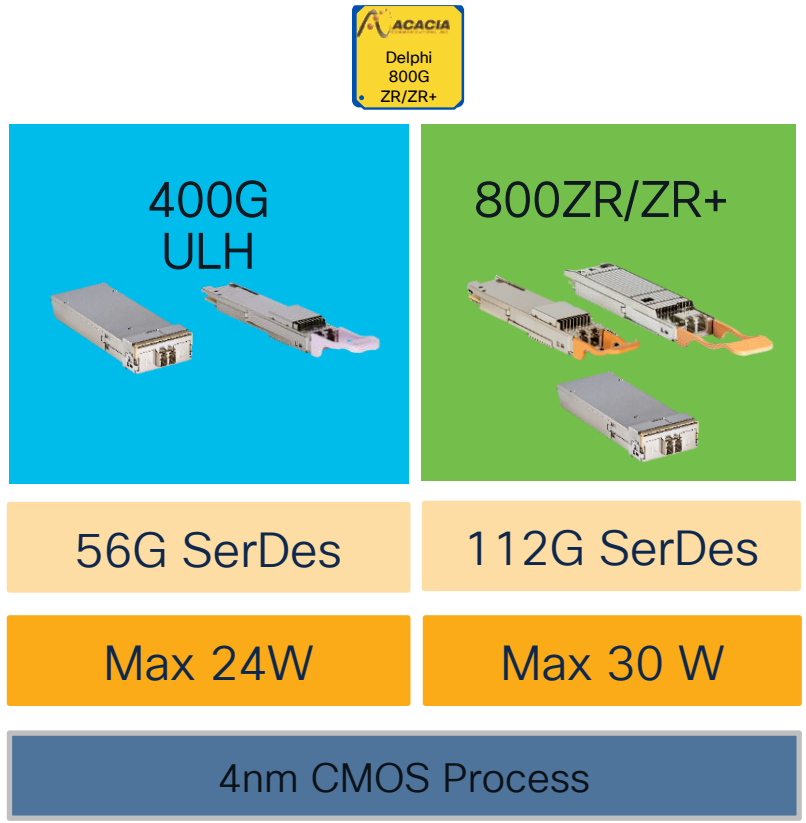
Welcome to the Next-Gen Coherent Pluggables !

- Latest 4nm CMOS process
- Up to 131 Gbaud Optics
- 1st to support interop PCS 800G ZR+ (OpenROADM) and OIF 800ZR modes
- Supports CMIS 5.3
- C & L Band

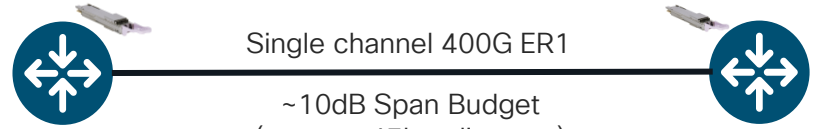


Next-Gen Coherent Optics RON building blocks

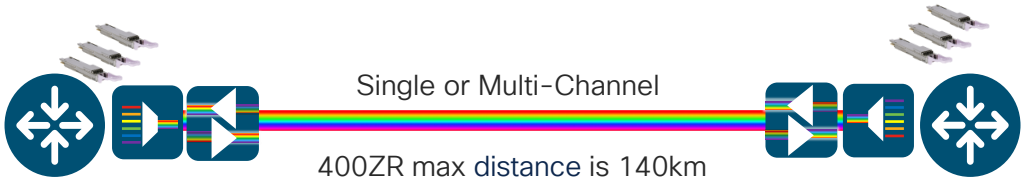
- Module Portfolio
  - 800ZR in QSFP-DD/OSFP
  - 800ZR+ in QSFP-DD/OSFP/CFP2
  - 400G Long-haul in QSFP-DD and CFP2
- Low Power 4nm DSP – Delphi
- Up to 130Gbd Optics
- 400G ULH Pluggables:
  - Designed for existing 400G platforms
  - Integrated EDFA and TOF
  - Flexible line side modes with PCS
- 800ZR/ZR+ Pluggables:
  - Comply to OIF 800ZR
  - Comply to 800G OpenROADM



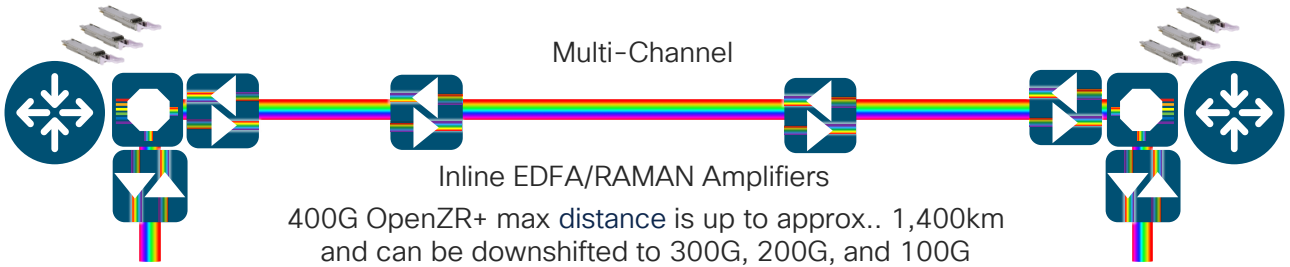
Coherent Pluggable Applications



Unamplified, Dark Fiber



Amplified, Point-to-point DWDM



Full featured DWDM
(ROADMs, ILAs)

Routed Optical Network (RON)

Use cases and deployment models

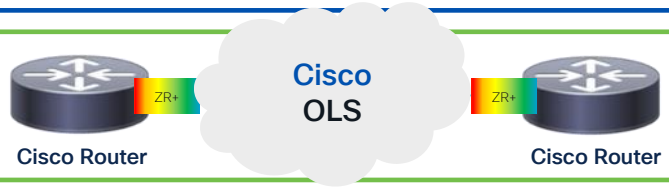
TCO* 46%
CapEx* 35%
OpEx* 57%

Different use cases and deployment models

1 QDD-ZR/ZR+ over third-party OLS

2 QDD-ZR over dark fiber or passive DWDM

3 QDD-ZR+ over Cisco OLS



>75% of the RON deployments are over third-party line system!

OLS: Open Line System

* ACG Research 2023

RON Deployments

348

Customers Engaged

300+

Customer deployments

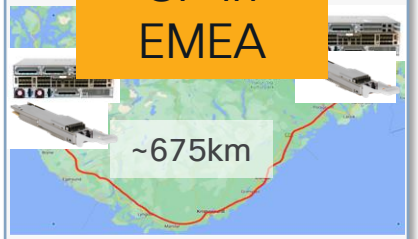
4x

Growth in deployments in FY24

Majority of the RON deployments are over 3rd party line system (Ciena, Nokia, Infinera etc.)

400G DCO Can Go the Distance

SP in EMEA

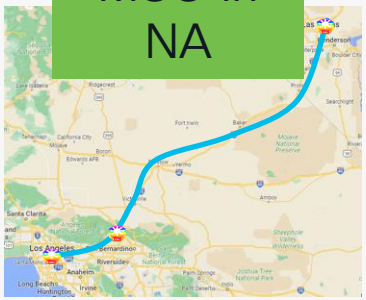


- Bright 400ZR+ QSFP-DD
- Alien λ over 3rd party line system
  - CDC ROADM, EDFA only
- >2.5dB margin
- Plug and play!

colt

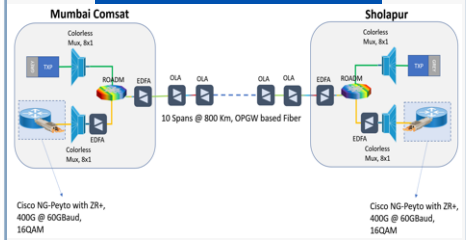
- Bright 400ZR+
- Across the English Channel with challenging high loss span
- Trial 1 at 430km; Trial 2 at 640km
- >2.5dB margin

MSO in NA



- Bright 400ZR+ with 8201
- 520km high loss E-LEAF fiber, span > 26.5dB
- Two different OLS vendors
- Error free with margin

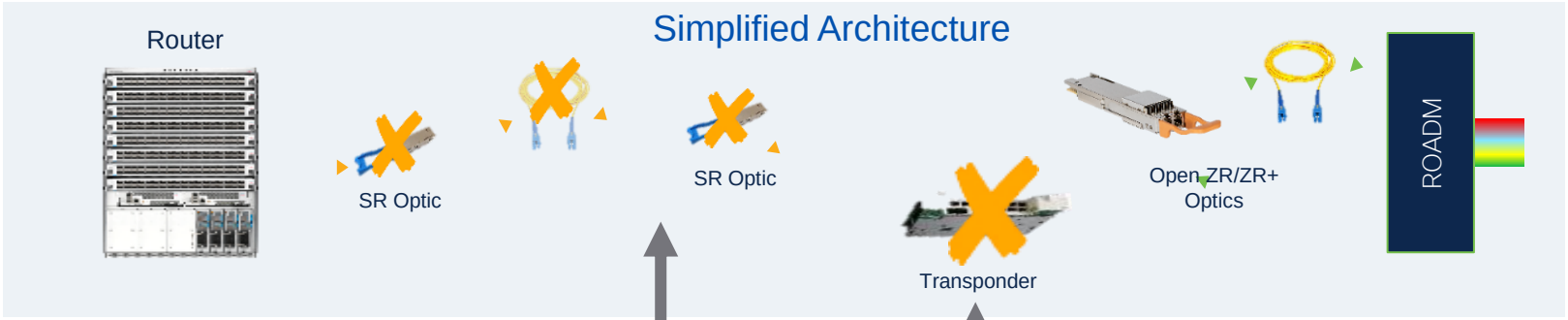
Tier 1 SP in India



- 400ZR+ with Cisco routers
- >800km longest route
- Core network over 3rd party ROADM OLS

Bright ZR+ has reach 600-800KM with good deployable margin over ROADM based OLS

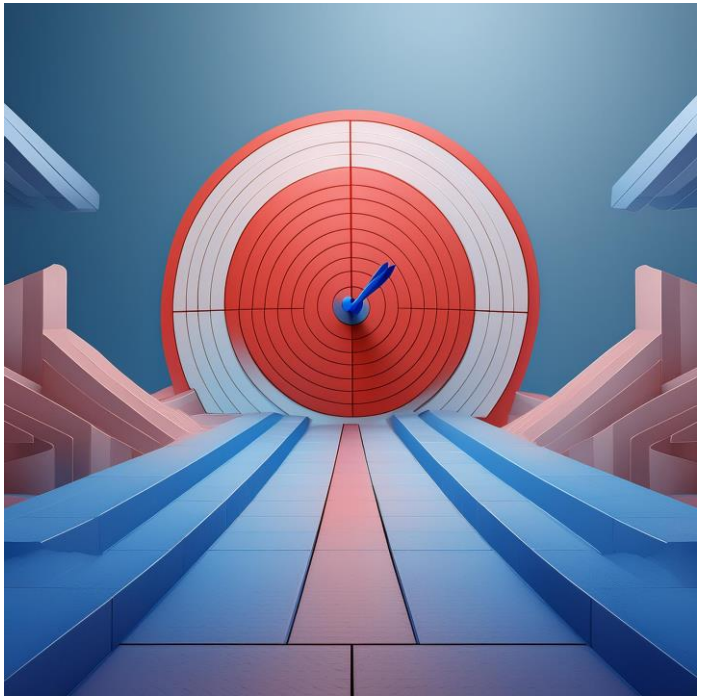
Summary – Coherent Optics enables IP-Optical Convergence



- Eliminate transponder and associated shelf/chassis
- Eliminate client optics

“Simplicity is the ultimate sophistication”
– Leonardo da Vinci

Key takeaways – coherent interfaces

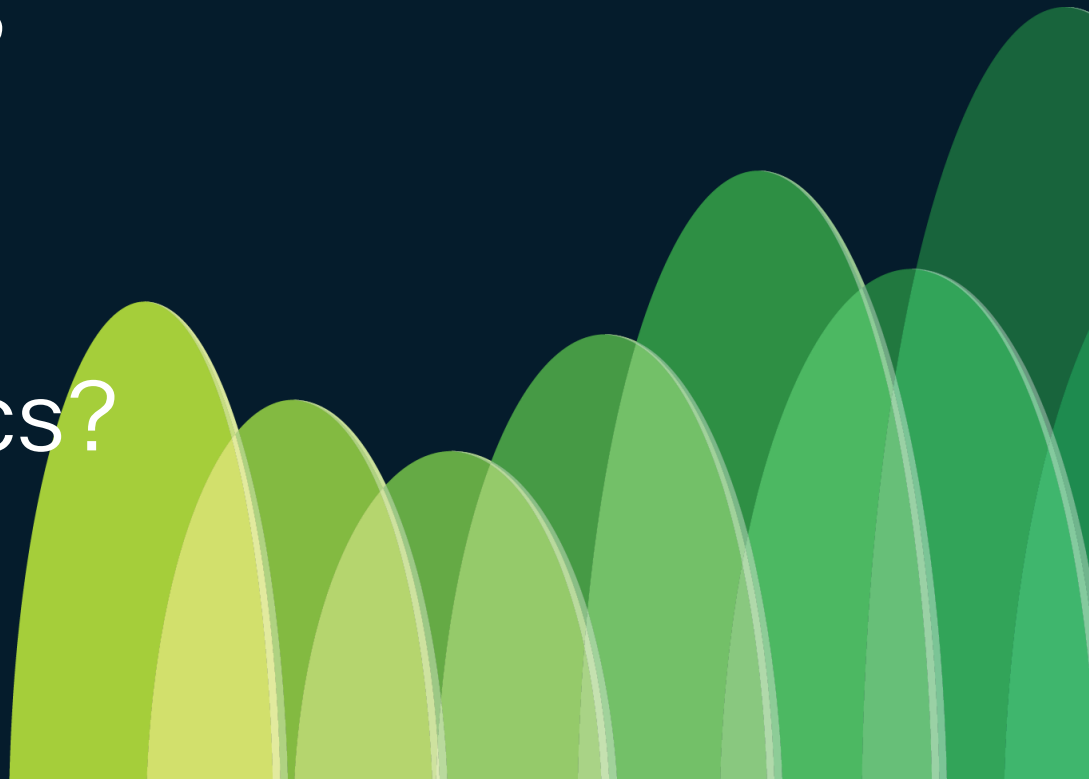


(AI generated pic)

- Pluggable solutions are leading option for coherent interfaces
- 400G coherent solutions (and standards) are mature with lots of deployment options
- 800G coherent solutions (and standards) available – interoperable PCS standardization improves performance
- Routed optical networking: improved sustainability, operational efficiency & cost efficiency

What is coming?

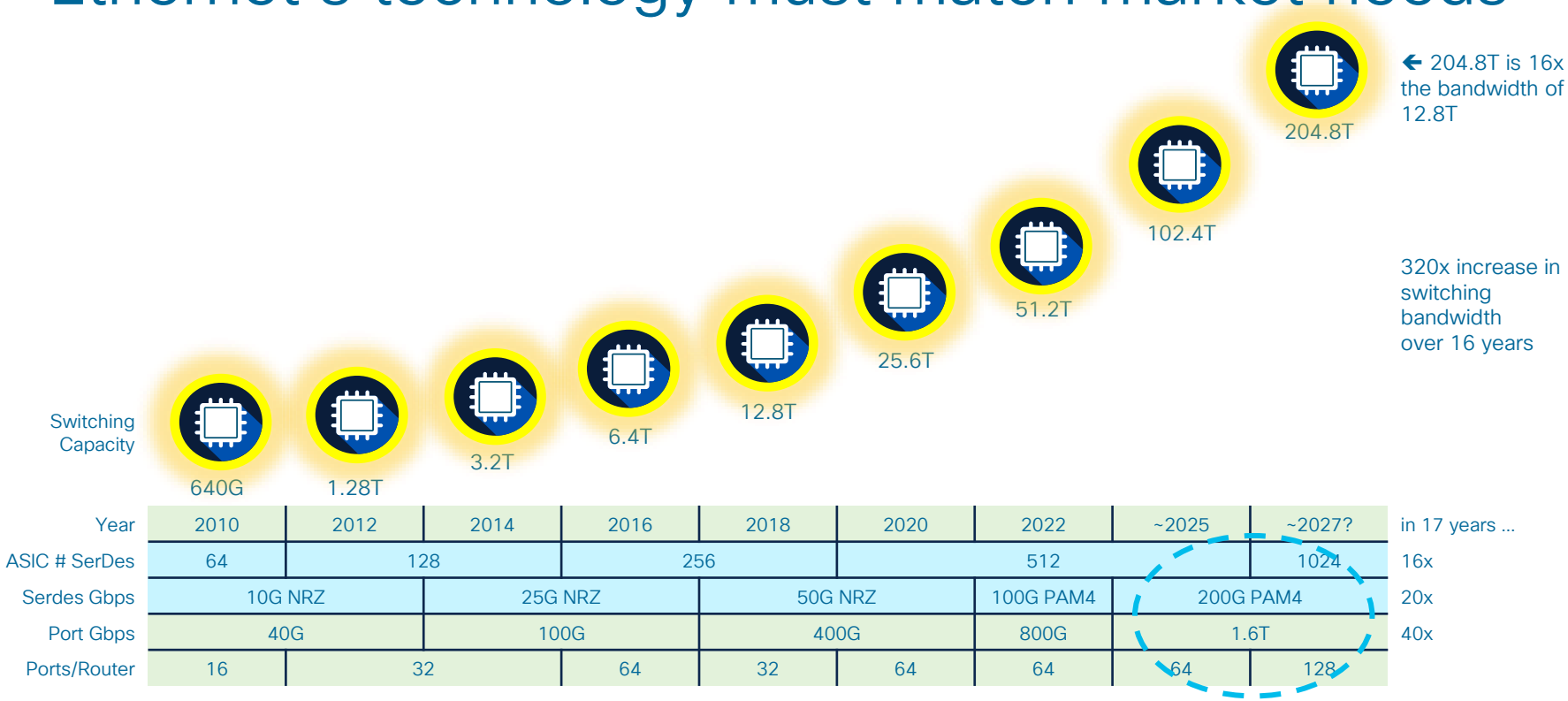
How does the AI/ML inflection point affect optics?



The function of pluggable optics

- The sole function of Optics is to extend the interfaces from one piece of equipment to another. The ASIC inside is driving the interface.
- Therefore, it is the ASIC capabilities and roadmap which primarily matter, and the role of optics is to keep up – without causing too many issues

Ethernet's technology must match market needs



ASIC density continues to redefine how products are built.
Gates & GHz. SerDes & Interconnect. Optics & wavelengths.

Credit: adapted from slide by Rakesh Chopra & John Chapman
https://blogs.cisco.com/sp/co-packaged-optics-and-an-open-ecosystem

Interconnects for an AI/ML world

AI/ML is a disruptive event for traditional networking

		Traditional Front-end DC	AI/ML Back-end DC
Rack Bandwidth (ToR/MoR)		3.2T-12.8T	>> 100T
Rack power		~10 kW	100 kW+
Packet Loss impact (reliability)		Low importance	Critical importance
Latency importance	Absolute	Low	Low
	Tail	-	High

Lots of interconnect
- → Speed matters
- → Power matters
- → Density matters

Massive rack density increase
- → Power matters
- → Copper cables matter
- → Density matters
- → Thermal solutions matter

Job completion time (JCT)
- →Link BER performance critical.
- →Tail latency most important

Priorities for interconnect in an AI/ML world

Power matters

Lower W/Gb

Copper Cables

Linear pluggable

Speed matters

Interfaces
800 GbE → 1.6 TbE

SERDES/ λ
100 Gb/s → 200 Gb/s

Solution integration* matters

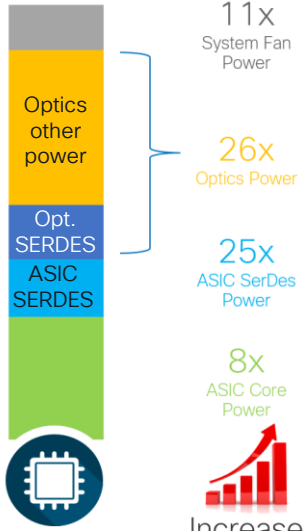
*compute plus networking

Rack density

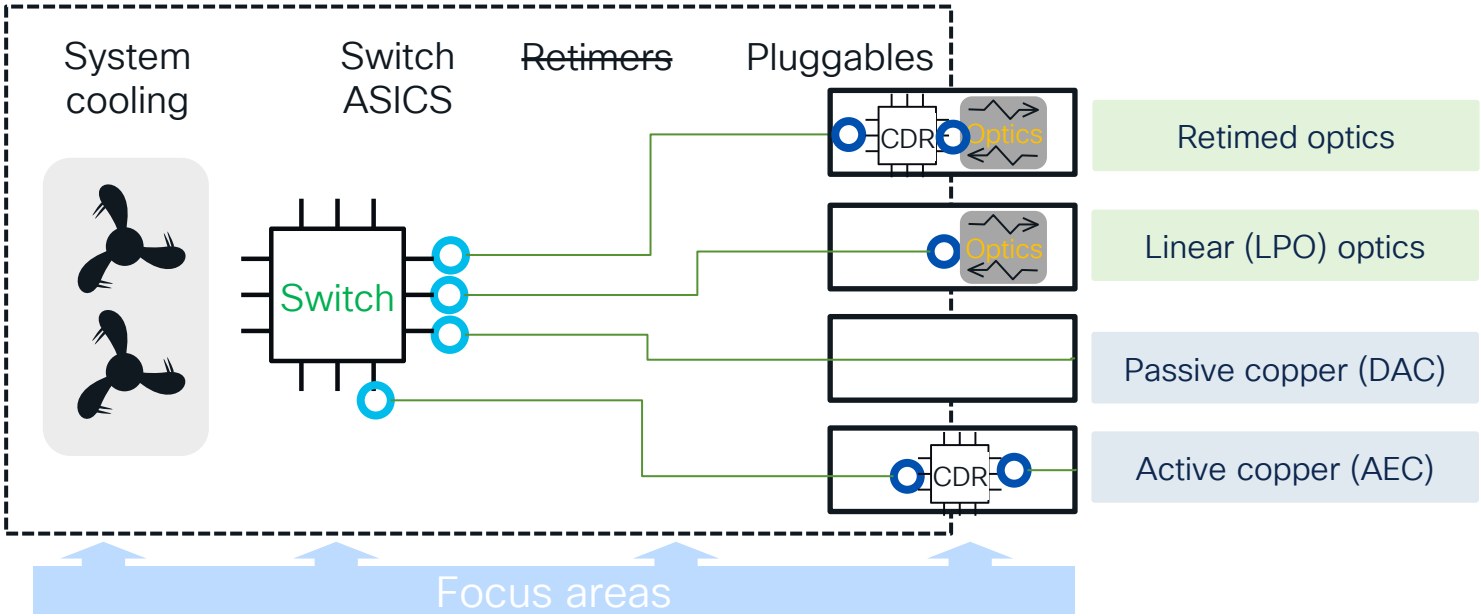
Switch density

Liquid cooling

Interconnect and power



Power increase for 80x bandwidth increase



- Air vs liquid cooling
- Highest capacity
- Serdes capability
- Optimized board designs
- Retimed optics
- Linear optics (new)
- Copper cables

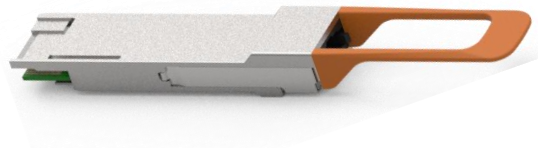
High-Density Interconnect options: Power vs. Reach Tradeoff

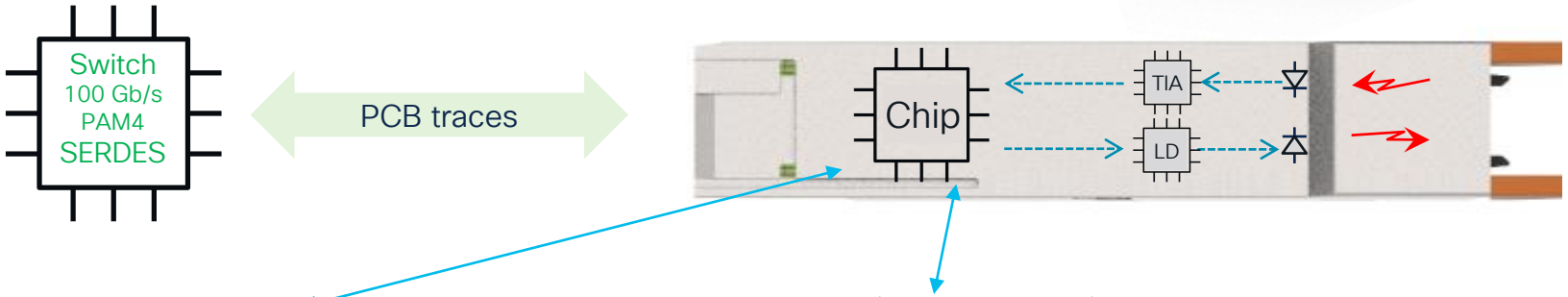
LPO – Linear Pluggable Optics
- fully linear optics (Tx & Rx)

LRO – Linear Receive Optics
- DSP on transmit side

Linear Optics
- very attractive as a solution to help with the AI/ML deployment power challenge
- But it needs an interoperable ecosystem – which is active work in progress.

What is Linear pluggable optics (LPO)?





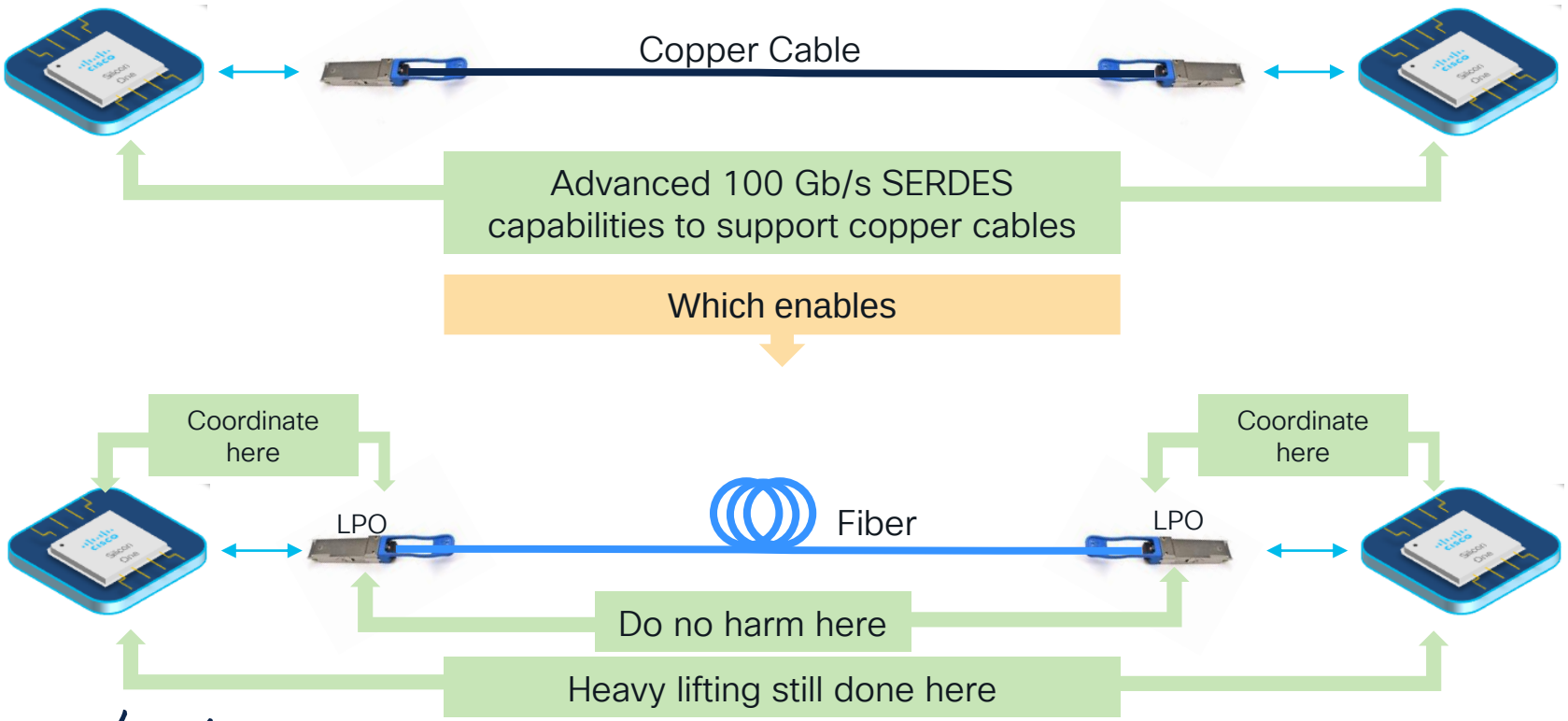
Current (Retimed)

- DSP has full digital equalization for both electrical & optical signals
- Enables broad interoperability
  - Host/Port/Module
- Full telemetry & loopbacks possible
- FEC Monitoring or Partitioning possible
- But this adds power

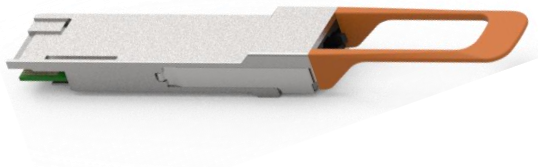
LPO (Non-retimed)

- Power reduction due to absence of digital equalization
- Performance is based on quality of every component in link since they are concatenated
  - Optics, Serdes, PCB, connectors
  - Varies port to port
  - Some linear gain and equalization in module
- Interoperability becomes more challenging
- Loss of some telemetry, monitoring or loopback

Why does linear optics work?



How does LPO work?



ASIC Serdes are the key

This is the big change!

- Very capable ASIC serdes Tx and Rx equalization capabilities
- Serdes are designed and capable to drive linear CR channel (copper cable)
- Cable (CR) channels are specified (IEEE) with lower host loss ports
- LPO modules need to work in every switch port, so they need to compensate for the extra host loss in those high loss ports
- LPO modules have limited linear gain and equalization that can compensate for host channel
- Host and module handshake to adjust equalizer settings
- Linear sub-components are key – SiPhotonics helps.

Cisco Silicon One G200: Retimed vs. LPO System Power Comparison

Cisco demonstrated LPO operation at OFC2024

<u>Set-up:</u>

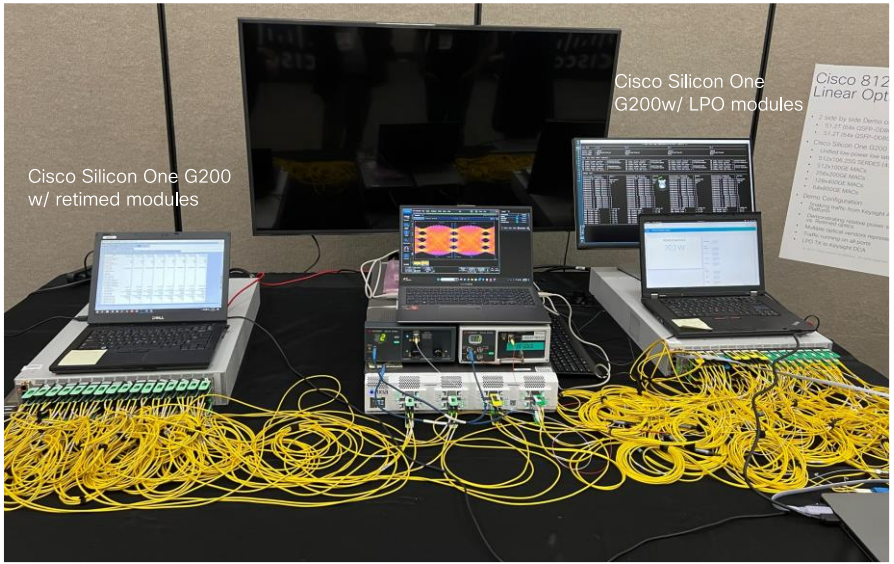
Two identical Silicon One 51.2Tb 64-port G200-based switches:

- 100% retimed optics
- 100% LPO optics

<u>Result:</u>

Both switches ran full traffic on all ports

Overall power reduction: ~ 700W



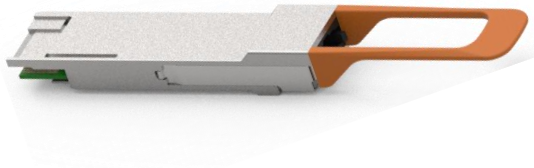
Linear (LPO) optics: Interoperability and ecosystem

LPO-MSA

- Recently formed Multisource Agreement (MSA) group with expertise from across the industry
- Creating specifications that ensures multi-vendor interoperability
- 100 Gb/s Specification publication soon
- Interop event confirms multi-vendor interoperation feasibility

www.lpo-msa.org

What is Half-linear (LRO)

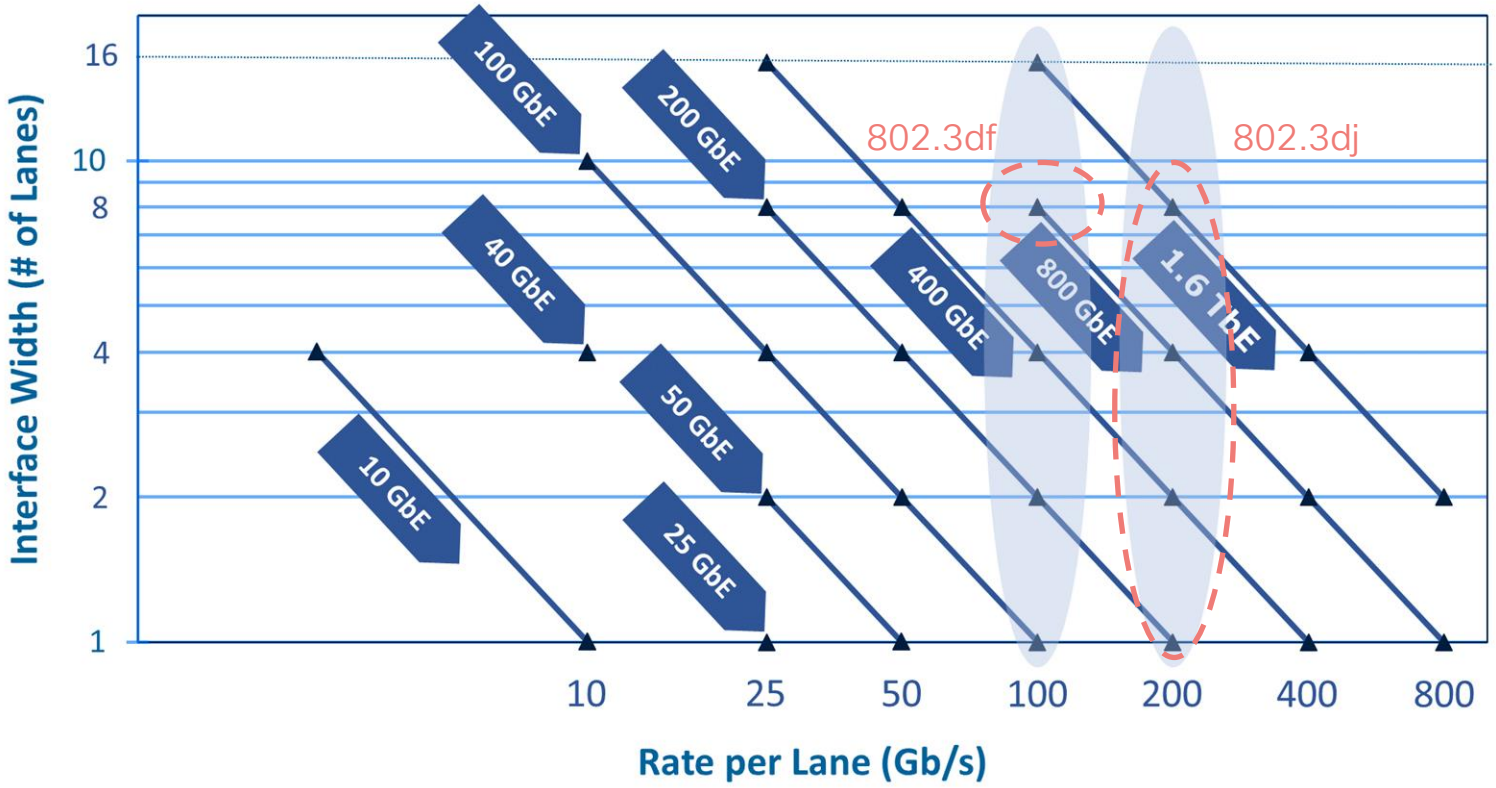


Half Linear (LRO – linear receive optics, aka TRO – transmit retimed optics)

- Pitched as a way of halving the problem. Lower power since half of DSP used.
- Reduces ability of host Tx ASIC serdes' to optimize the full end-to-end link performance.
- Measurements so far show that link performance is ~1-2 orders of magnitude worse than LPO.

For AI/ML link performance (BER margin) is very important. Any bit error can force a packet retransmit which delays the AI/ML job completion time.

Next generation Ethernet's High-Speed Solutions – reuse, reuse, reuse



IEEE 802.3dj objectives: set by assessing market requirement and technology potential

Key themes to notice:

- 200 GbE, 400GbE, 800GbE and 1.6TbE Ethernet interfaces (N x 200Gb/s)
  - Consideration of Breakout use-case
- Electrical interfaces based on 200 Gb/s
  - Copper Cables
  - Backplanes
  - Chip to module (AUI) interfaces and Chip to chip interfaces
- Optical Interfaces
  - Only single mode fiber (Duplex and Parallel)
  - 500m to 10km using 200 Gb/s per lane IMDD technology
  - 10km to 40 km using 800 Gb/s per lane (16QAM) coherent technology

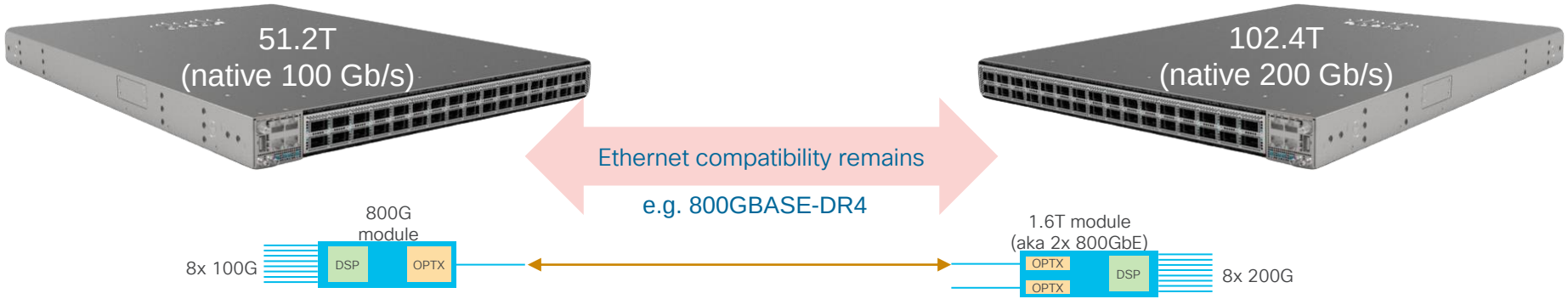
What Will P802.3dj provide? Flexibility of implementations

- Copper cables
- Multimode Fiber – 100m (future project)
- Single Mode Fiber inside DC – 500m & 2km
- Single Mode Fiber Campus – 10 km
- Outside plant, DCI – up to 40 km

Compatibility with known implementations was a strong factor in P802.3dj decisions

- Compatible with current CMOS
- Compatible with popular system configs or pluggable modules

IEEE P802.3dj will maintain Ethernet's backwards compatibility

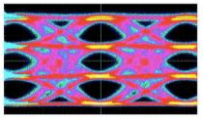


P802.3dj's logic and architecture ensures compatibility with "legacy" systems and implementations.

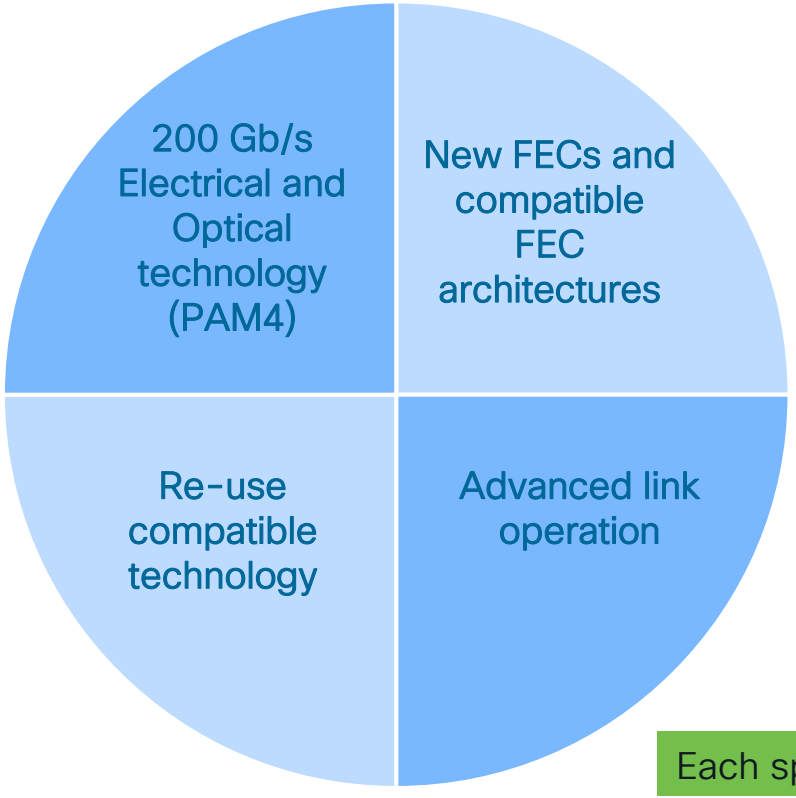
Any 800 GbE capable port will operate with any other 800 GbE capable port:

- As long as the same PMD interface is used
- For example, 800GBASE-FR4 in both ports

Key elements of 200 Gb/s technology



Compatible with current CMOS, SiPhotonics, EML, Cable, Fiber technology



Multiple FEC schemes for different reaches. Common architecture to enable compatible implementations

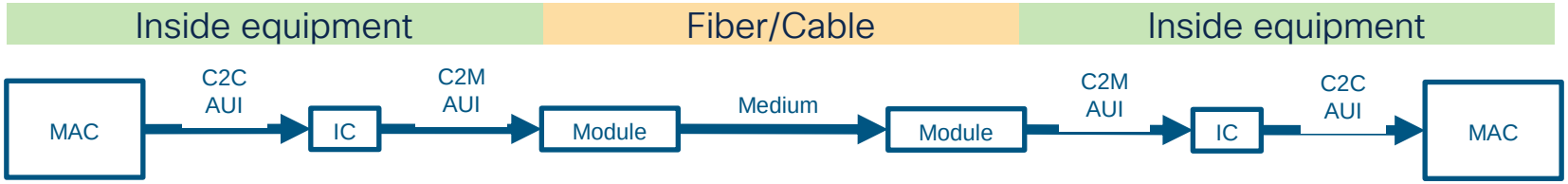
Implementations can use same technology for 8x, 4x, 2x or 1x variants

New (optional) Optical Link Training and proposed Auto-negotiation

Each speed step is more challenging!

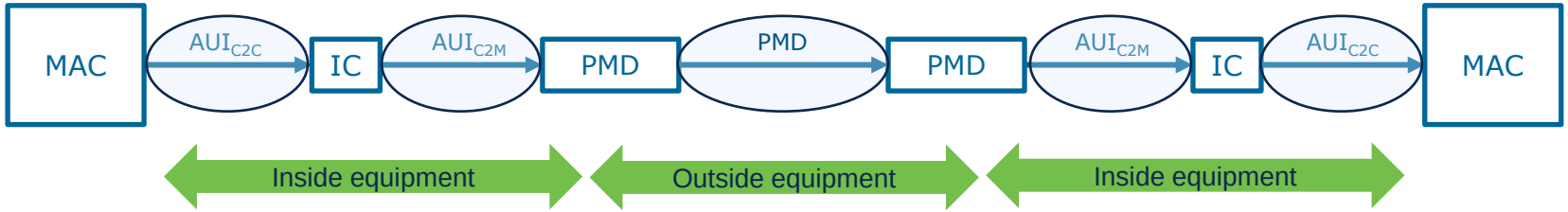
BER = bit error ratio
C2M = chip-to-module
C2C = chip-to-chip

P802.3dj: Multiple FEC Schemes Needed



		PMD Reaches
End-to-end	**FEC A corrects all errors** RS(544, 514)	0-500m
Concatenated	**Inner FEC (FEC I) corrects most errors here** **Outer FEC (FEC A) corrects all other errors** RS(544, 514) + BCH(128,120) or RS(544, 514) + BCH(126,110)	2km – 10km
Segmented	**FEC A corrects errors here** RS(544, 514) / **FEC B corrects errors here** oFEC / **FEC A corrects errors here** RS(544, 514)	> 10km

Making 200 Gb/s easier to use



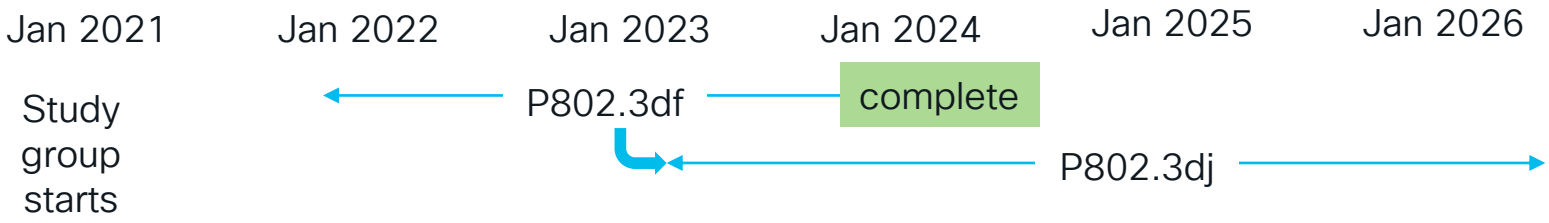
1. FEC ← Integrated host FEC for all variants → FEC

2. Extra FEC for different PMDs as needed — FEC ↔ FEC

3. Auto-negotiation between devices

4. NEW — Electrical and optical link training between devices

When will IEEE P802.3dj be done?



Status:

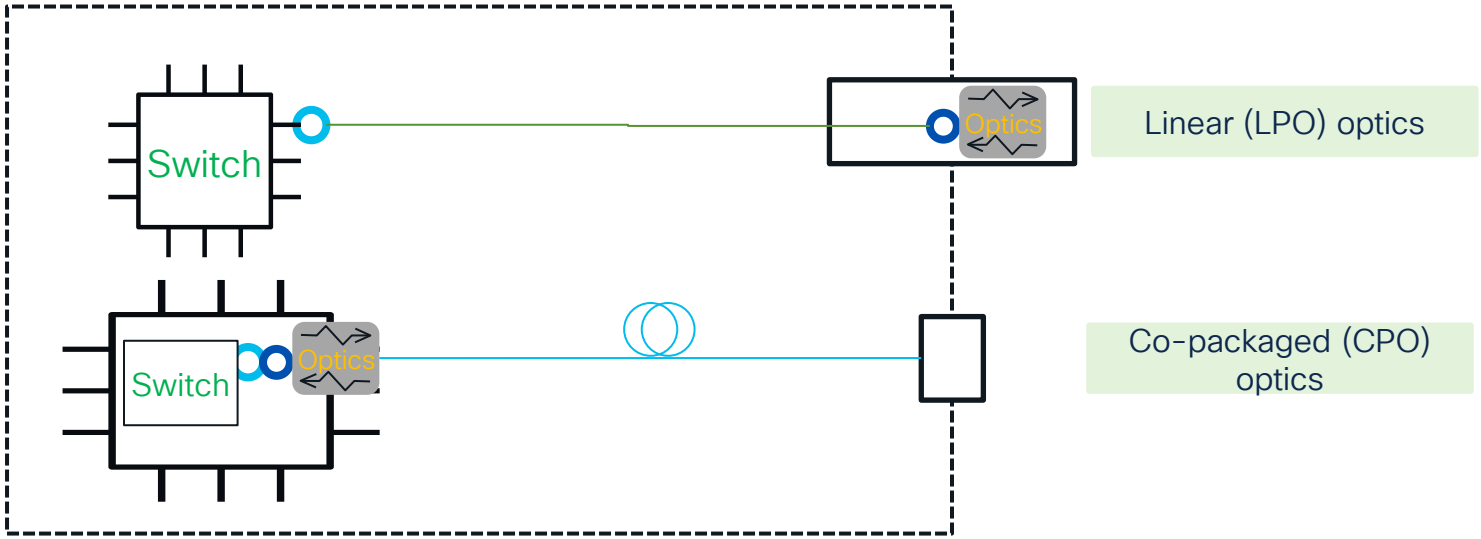
- Complete set of baselines adopted. Deep into TF review. Last TBDs remaining
- Refinement underway
- Working towards a technically complete document (known as Draft 2.0). Target Q1 2025
- Formal Standard publication target is 2H-2026

Products:

- Technology demos already happening
- Product development, in progress while specifications mature, does not wait for Standards publication.

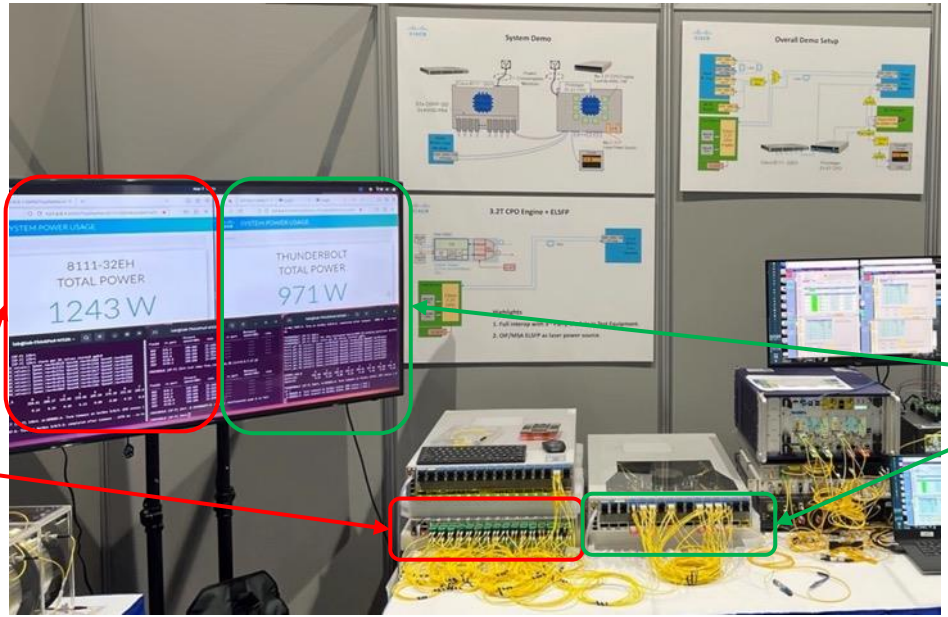
CPO & LPO

Co-packaged optics (CPO) and Linear Pluggable Optics (LPO) are two implementation variants of the same idea – reduce ASIC to optics power/DSP



Cisco Silicon One: OFC 2023 Demo – 25.6T
Co-packaged Optics vs Retimed

CPO system achieved ~ 22% total system power reduction



25.6T Ethernet Switch
64 ports @ 400G QSFP-DD

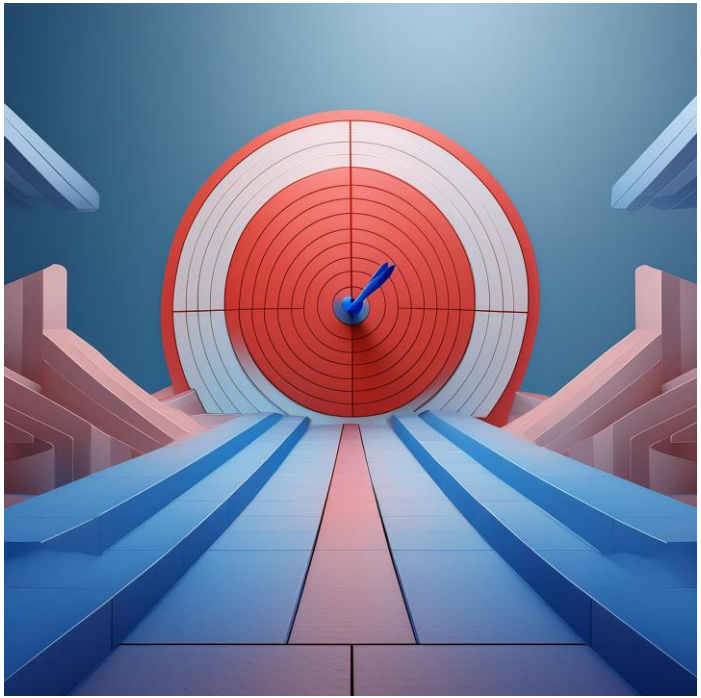
25.6T Ethernet Switch
Enabled by CPO and using ELSFP external light sources

Greater density requires advanced cooling

If rack capacities grow from ~10kW → >100kW something needs to remove the heat. Two solutions under development:

- Cold plate liquid cooling
  - Liquid circulated through pipes and cold plates which are attached to key components (ASIC, optics) to efficiently remove heat
- Immersion liquid cooling
  - Equipment is immersed in special non-conductive liquids to very efficiently remove heat

Key takeaways



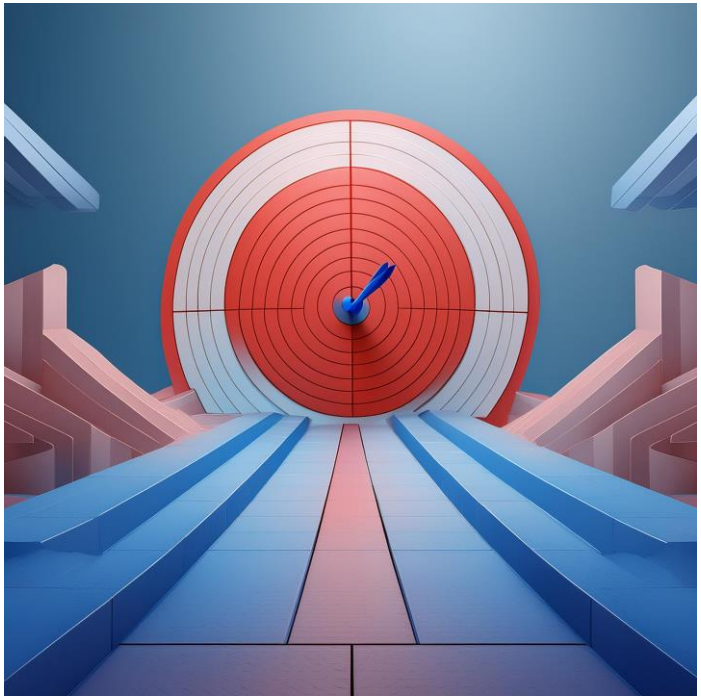
(AI generated pic)

Requirements around AI/ML are driving interconnect right now

- Faster speeds – 800G → 1.6T → …
- Lowering power is key
  - Copper cables still live @ 200 Gb/s
  - Novel implementations like LPO / CPO
  - Faster speeds yield better energy efficiency (W/Gb)
- Advanced cooling strategies are being considered
  - Liquid cooling and Immersion cooling
- AI networking will dominate the future optics design priorities for data centers

Wrap up

Key takeaways – thank you for your attention



(AI generated pic)

IMDD: 400G is mature & 1st gen 800G uses same tech. Breakout solutions offer efficiency

Coherent: 400G is mature, 800G available. Pluggable & RON adoption are growing

Optics and Interconnect are responding to demands from AI/ML growth

New speeds: 800G, 1.6T
New approaches: Linear pluggable (LPO), Co-packaged optics (CPO), Liquid cooling

Webex App

Questions?

Use the Webex app to chat with the speaker after the session

How

1. Find this session in the Cisco Events mobile app
2. Click “Join the Discussion”
3. Install the Webex app or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until February 28, 2025.

Fill Out Your Session Surveys

Participants who fill out a minimum of 4 session surveys and the overall event survey will get a unique Cisco Live t-shirt.

(from 11:30 on Thursday, while supplies last)

All surveys can be taken in the Cisco Events mobile app or by logging in to the Session Catalog and clicking the ‘Participant Dashboard’

Content Catalog

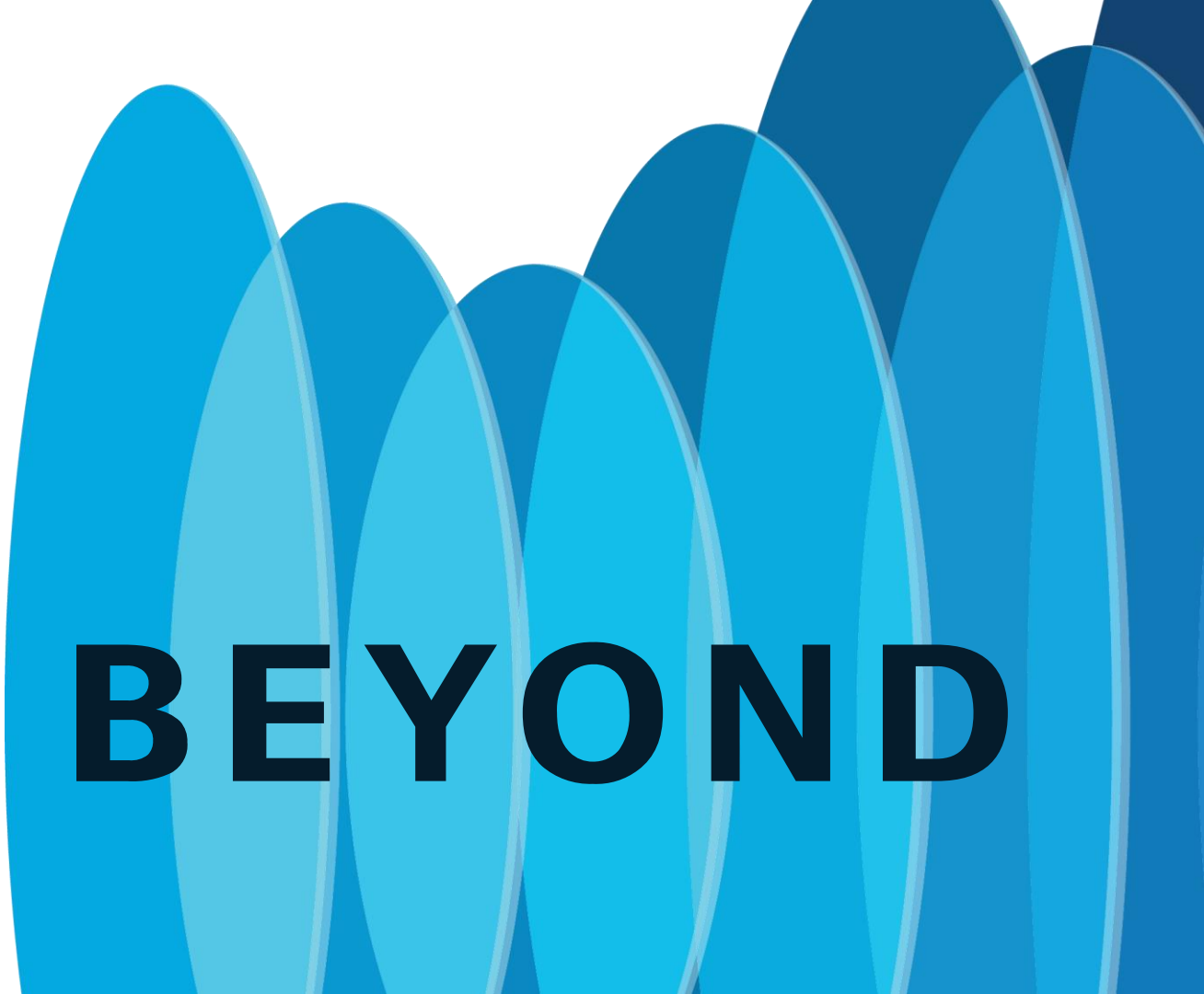
Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at ciscolive.com/on-demand. Sessions from this event will be available from March 3.

Contact me at:BRKOPT-2699: 400G, 800G, and Terabit Pluggable Optics: What You Need to Know


cisco
Thank you
cisco Live!

cisco Live!



GO BEYOND