Next Generation Optics for Datacenter, Access and Long-Haul

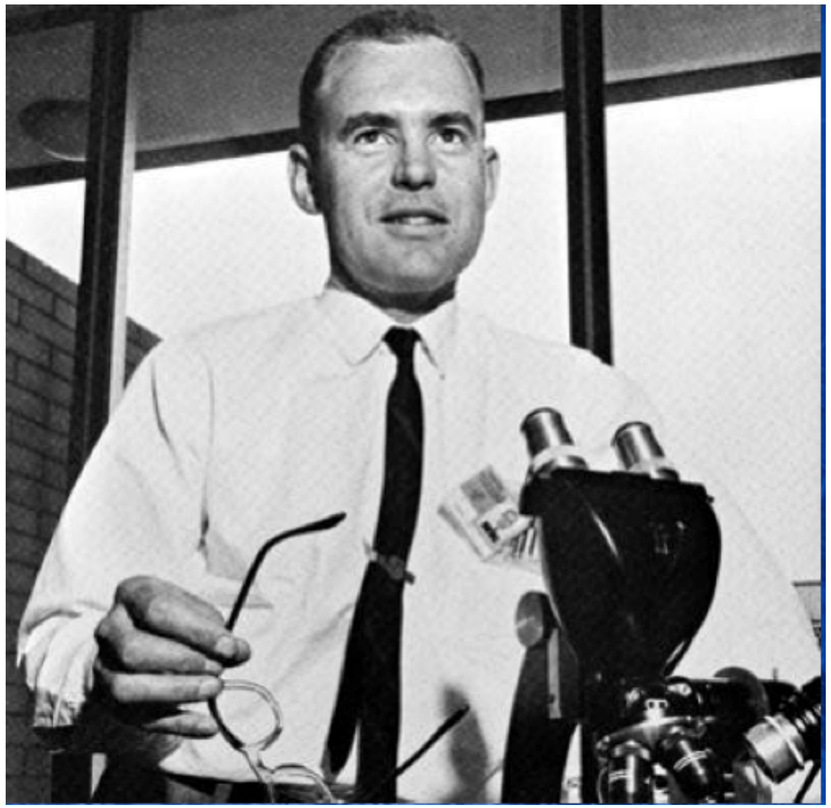
Andreas Bechtolsheim
Arista Networks, Inc

ARISTA

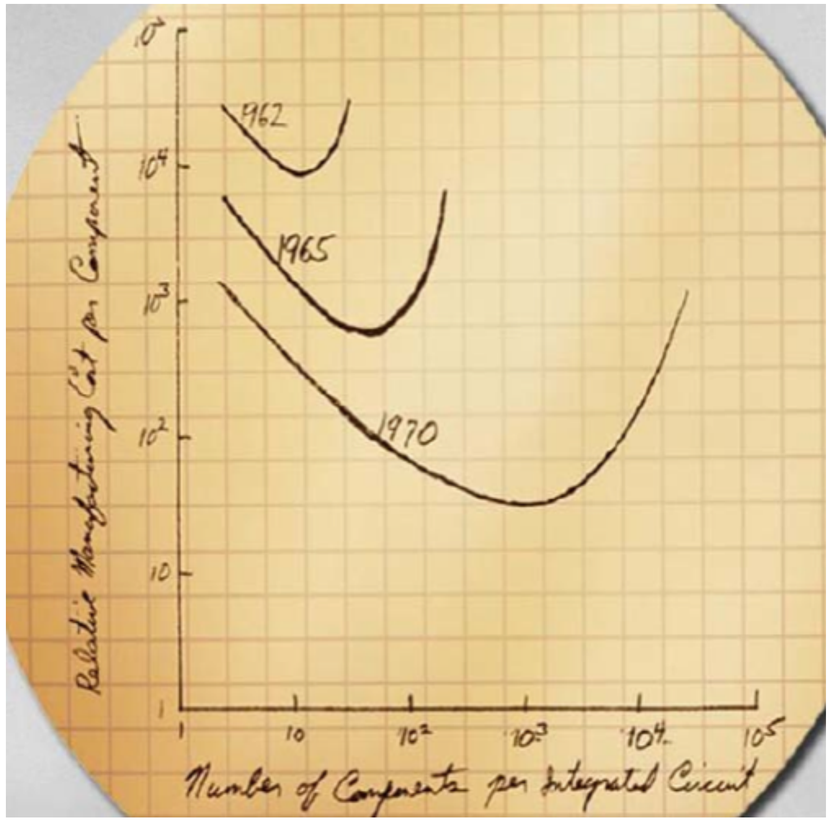
Agenda

1. Moore's Law
2. Serdes Roadmap
3. 400G and 800G Optics
3. 400G-ZR and 400G-ZR+
5. Optical Line Systems

Integrated Circuits and Moore’s Law

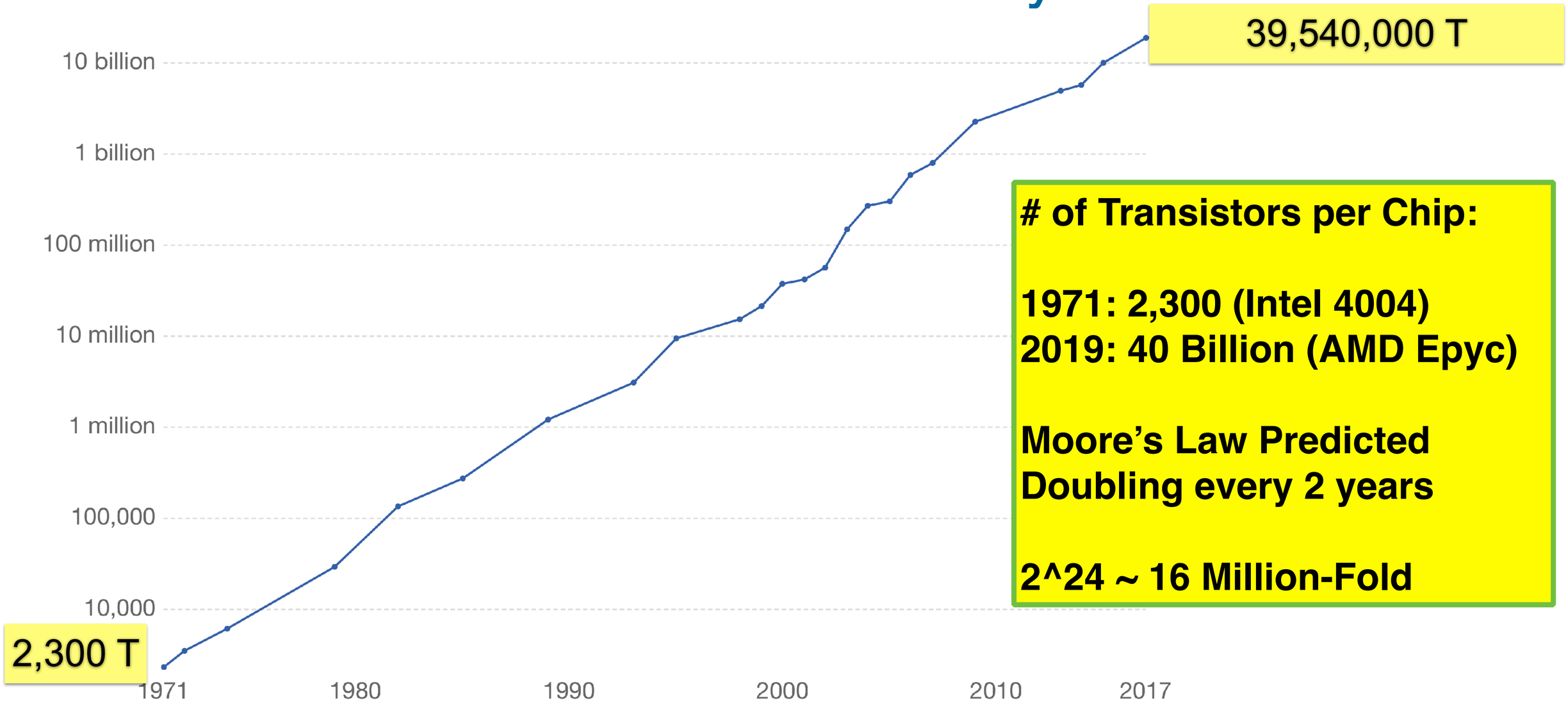


Dr Gorden Moore at Fairchild, 1966



Original Version of Moore's Law

Moore's Law 1971-2019: 2X every 2 Years



# of Transistors per Chip:

1971: 2,300 (Intel 4004)
2019: 40 Billion (AMD Epyc)

Moore's Law Predicted Doubling every 2 years

2^24 ~ 16 Million-Fold

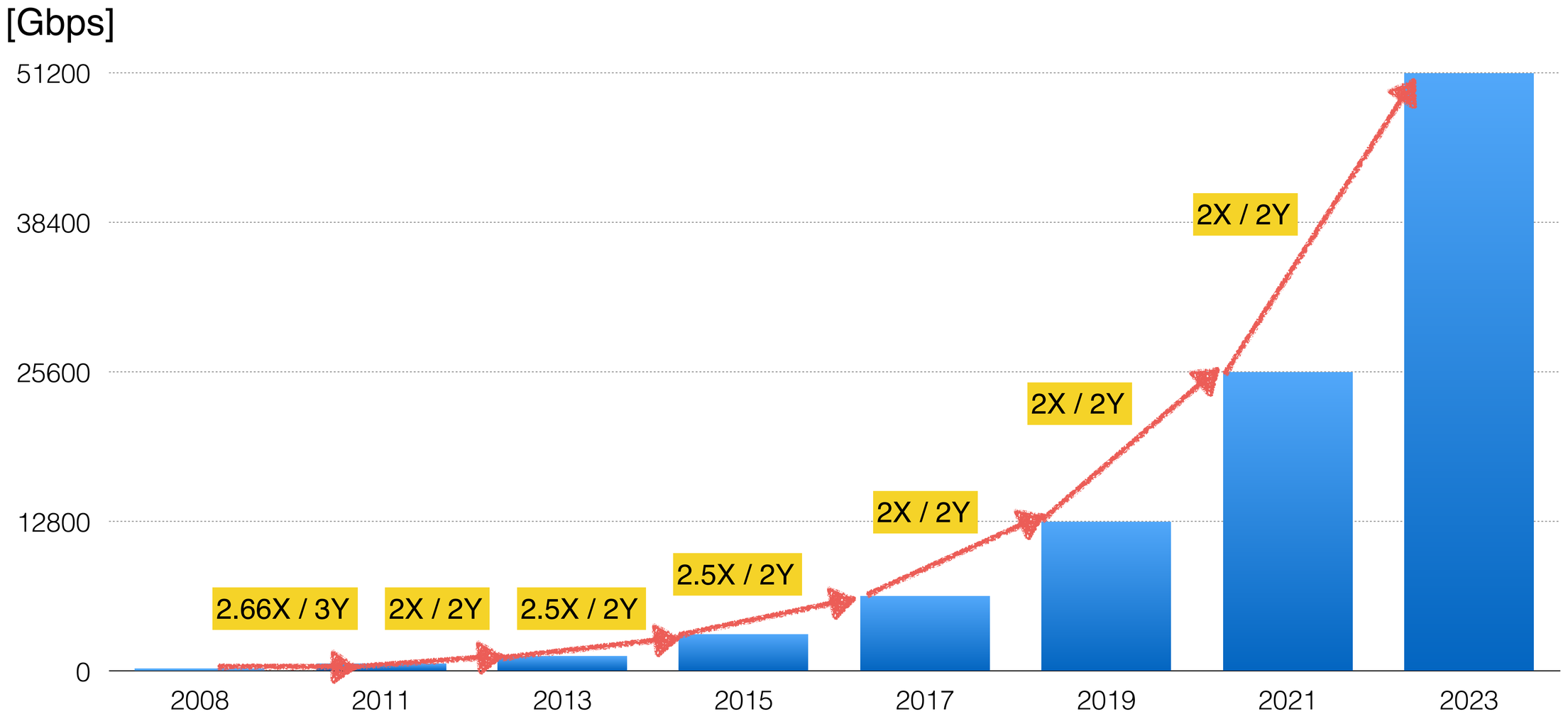
CMOS Logic Roadmap (IRDS 2020)

Table ES2 Overall Roadmap Technology Characteristics

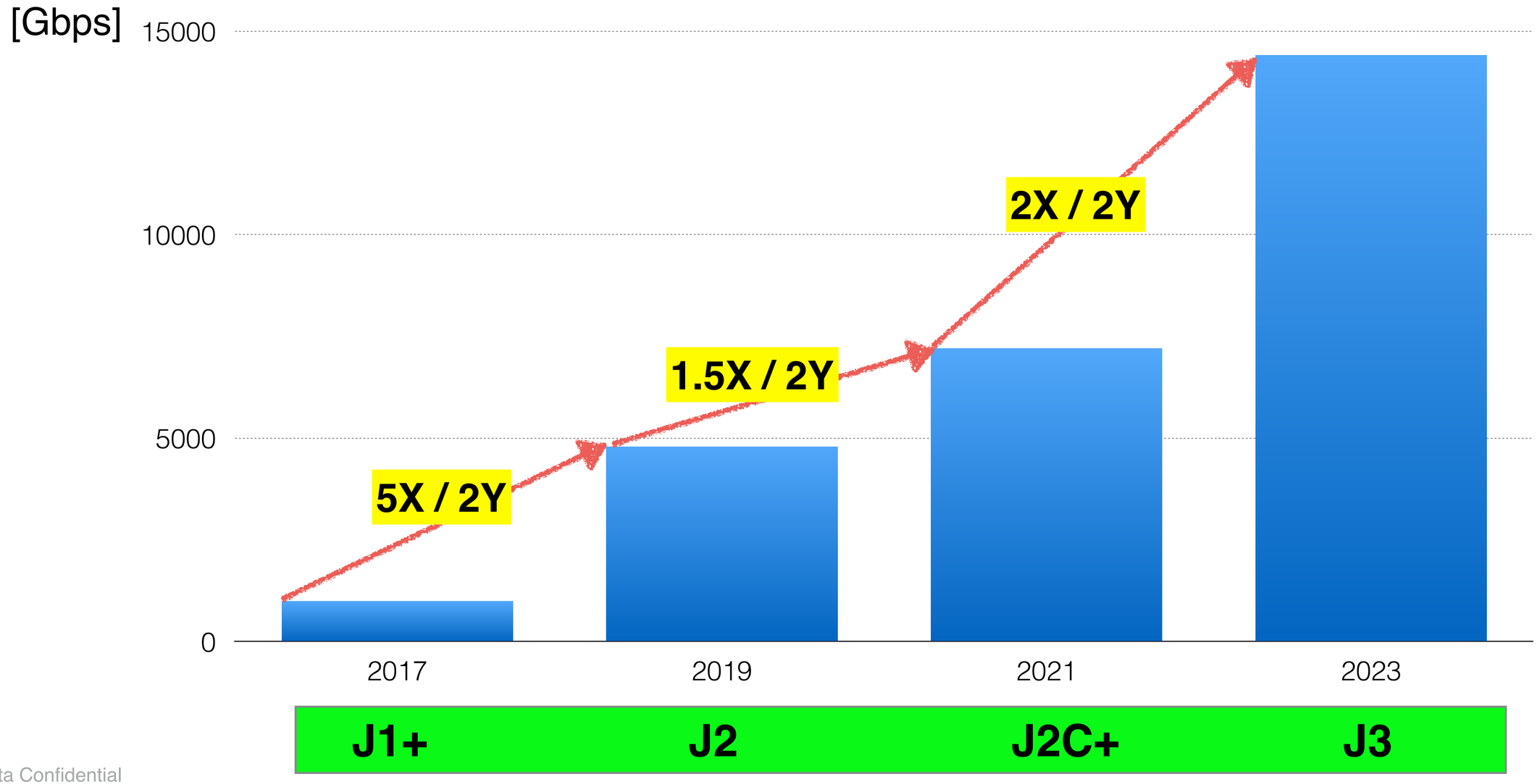
2020 IRDS Executive Summary Drivers-ORTC								
YEAR OF PRODUCTION	2019	2020	2022	2025	2028	2031	2032	2034
Logic device technology naming [4] NEW node definition	G54M38	G48M36	G45M24	G45M20	G40M16	G38M16T2	G38M16T3	G38M16T4
Logic industry "Node Range" Labeling (nm)	"7"	"5"	"3"	"2.1"	"1.5"	"1.0nm-eq"	"1.0nm-eq"	"0.7nm-eq"
Logic device structure options	FinFET	FinFET	FinFET LGAA	LGAA	LGAA VGAA	LGAA-3D VGAA	LGAA-3D VGAA	LGAA-3D VGAA
LOGIC CELL AND FUNCTIONAL FABRIC TARGETS								
Average Cell Width Scaling Factor Multiplier	1	0.9	0.9	0.9	0.9	0.9	0.9	0.9
LOGIC DEVICE GROUND RULES								
MPU/SoC M0 1/2 Pitch (nm) [1,2]	18	15	12	10.5	8	8	8	8
Physical Gate Length for HP Logic (nm) [3]	20	18	16	14	12	12	12	12
Lateral GAA (nanosheet) Minimum Thickness (nm)				7	6	5	5	5
Minimum Device Width (FinFET fin, nanosheet, SRAM) or Diameter (nm)	9	7	6	7	6	6	6	6
LOGIC DEVICE Electrical								
Vdd (V)	0.75	0.7	0.7	0.65	0.65	0.6	0.6	0.6
DRAM TECHNOLOGY								
DRAM Min half pitch (nm) [1]	18	17.5	17	14	11	8.4	8.4	7.7
DRAM Min Half Pitch (Calculated Half pitch) (nm) [1]	20.5	17.5	18.5	15	12	10	10	8.5
DRAM Cell Size Factor: aF^2 [4]	6	6	4	4	4	4	4	4
DRAM Gb/1chip target	8	8	16	16	32	32	32	32
NAND Flash								
Flash 2D NAND Flash uncontacted poly 1/2 pitch – F (nm) 2D [1][2]	15	15	15	15	15	15	15	15
Flash Product highest density (independent of 2D or 3D)	512G	1T	1T	1.5T	3T	4T	4T	4T+
Flash Product Maximum bit/cell (2D_3D) [6]	2_4	2_4	2_4	2_4	2_4	2_4	2_4	2_4
Flash 3D NAND Maximum Number of Memory Layers [6]	48-65	64-96	96-128	128-192	256-384	384-512	384-512	512+

https://irds.ieee.org/

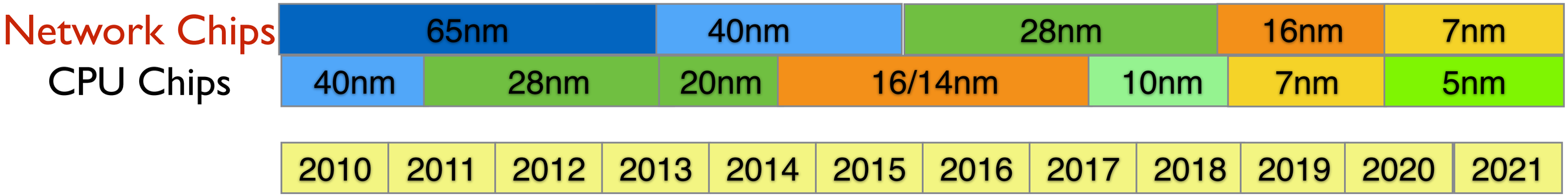
Merchant Switch Silicon Through 2023



Jericho VOQ Big Buffer Bandwidth Per Chip



Semiconductor Process Technology Roadmap



Network silicon traditionally lagged CPU process technology.
By catching up from behind, Network Silicon has been able to accelerate more quickly than Moore's law

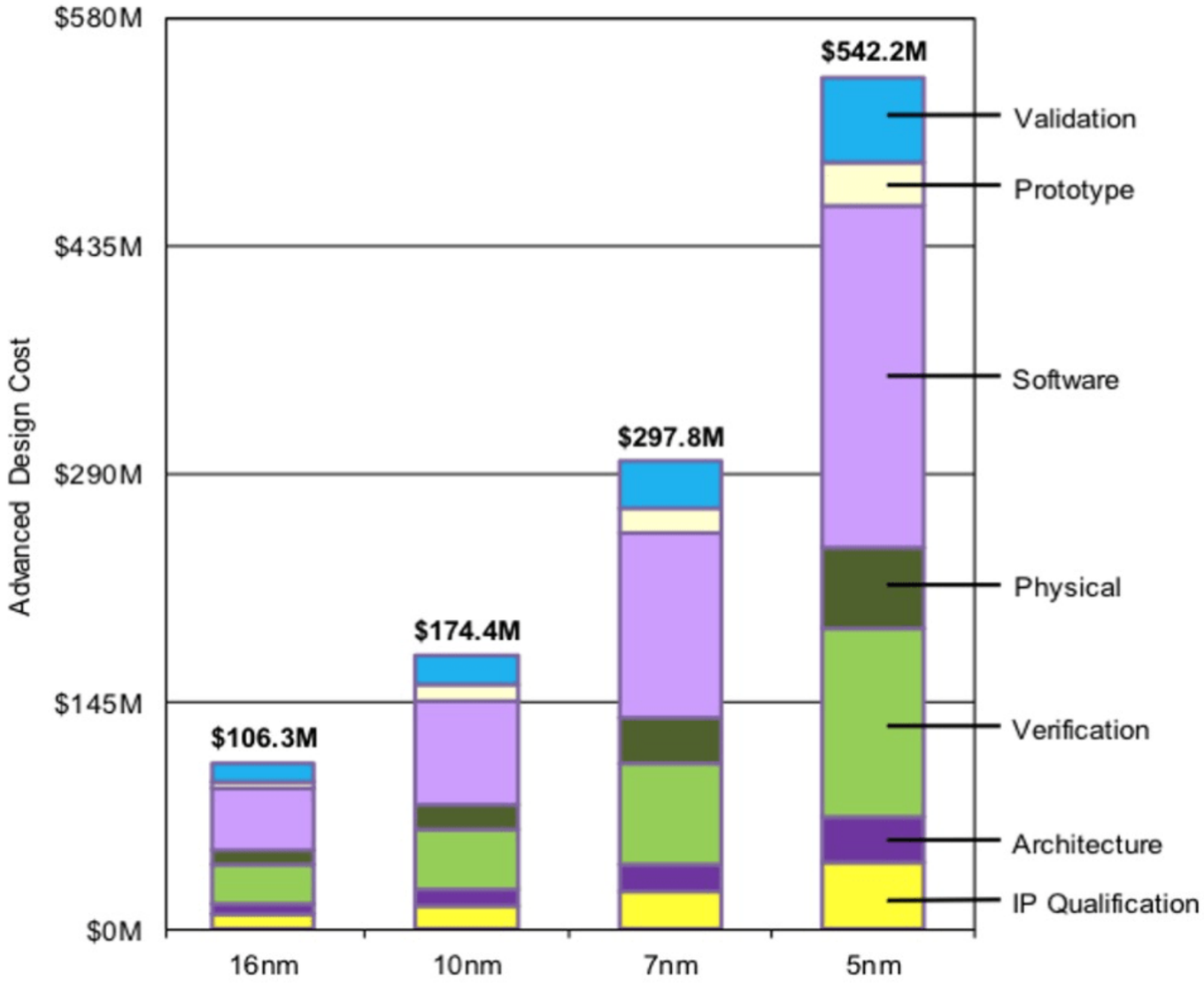
Process Technology Improvements

Improvement	16nm	7nm	5nm
Density	1	3,33	5
Speed @ IsoPower	1	1,35	1,55
Power @ IsoSpeed	1	0,55	0,45

2019 Silicon | 2021 Silicon | 2022-23 Silicon

Each process generation enables more performance, better Power Efficiency, more buffers, bigger routing tables, etc

Chip Economics Only Work in Volume



Escalating Design Costs means that advanced process node chips only work in high volume

Moore's Law Summary

Moore's Law is not dead

Well-defined roadmap to 3nm and beyond

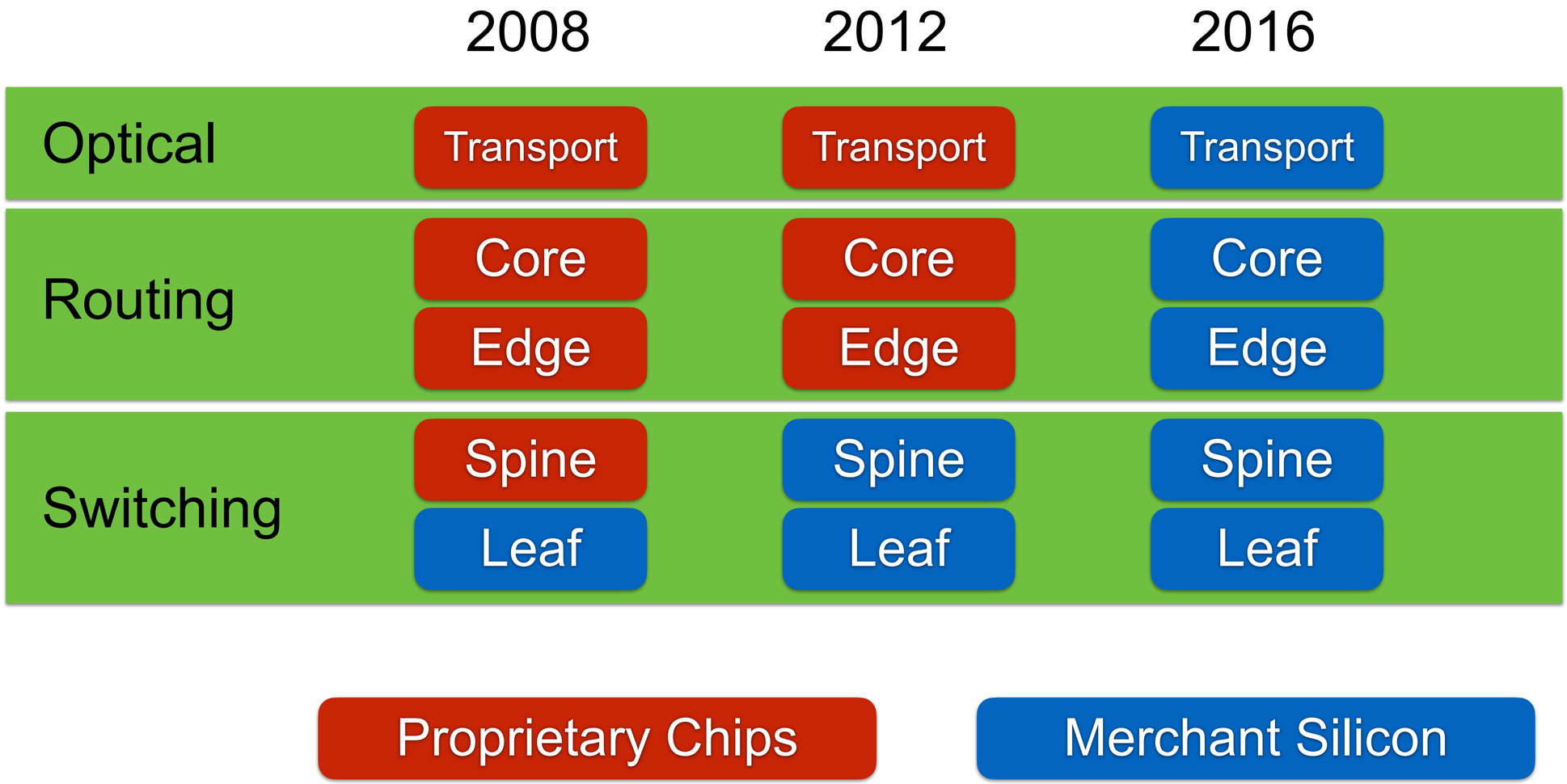
Chip Design Costs are Escalating

Only viable business model is merchant silicon

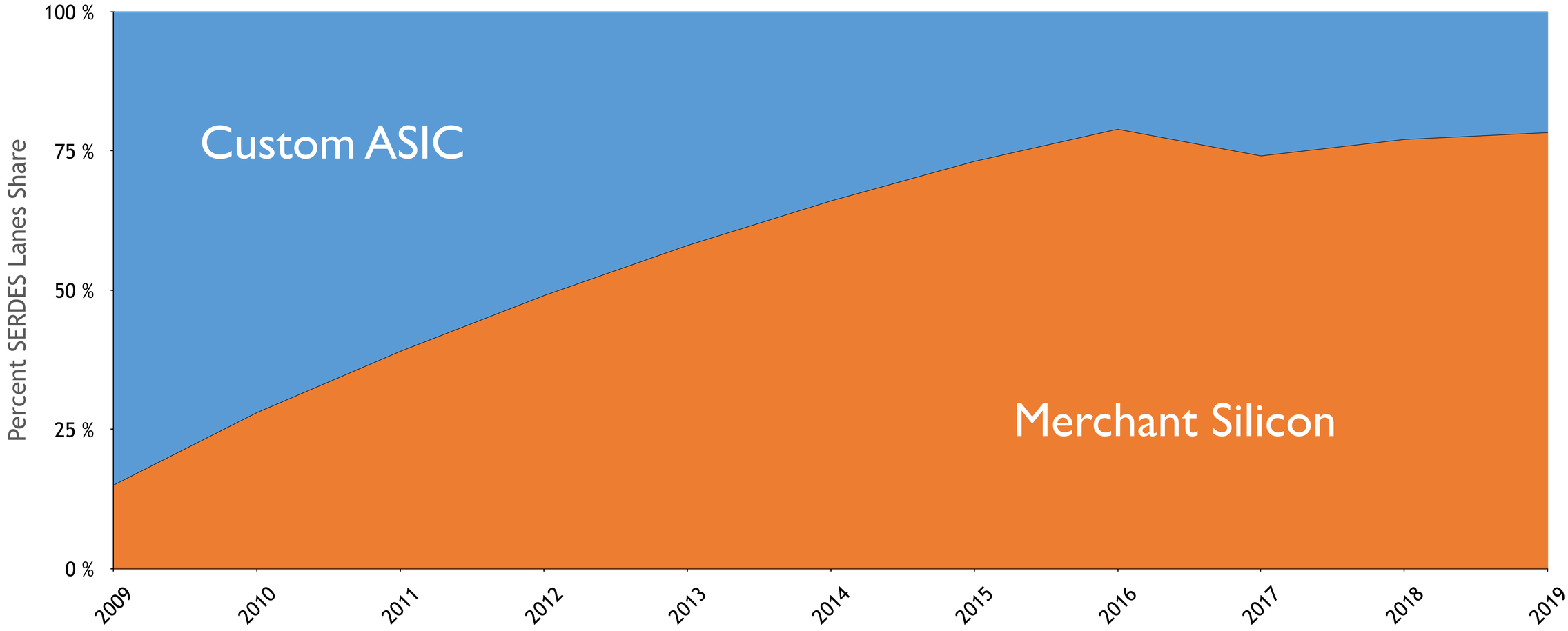
There will be fewer Chip Architectures

For Switches, Routers and Optics

The Rise of Merchant Silicon



Datacenter Switching Shift to Merchant Silicon



Source: 650 Group, December 2019

Long List of Merchant Switch Silicon Firsts

2008: First ultra-low latency 24-port 10G single chip
2010: First Large Buffer 10G Chip with VOQ Fabric
2011: First 64-port 10G single chip switch
2012: First 32-port 40G single chip
2013: First Large Buffer 40G Chip with VOQ Fabric
2015: First 32-port 100G single chip
2016: First Router 100G Chip with VOQ Fabric
2017: First 64-port 100G single chip
2018: First 32-port 400G single chip (128 port 100G)

Why Merchant Silicon Is Winning

Faster Time to Market

Rapid Evolution of Silicon Capabilities

Nobody wants Proprietary Networking Features

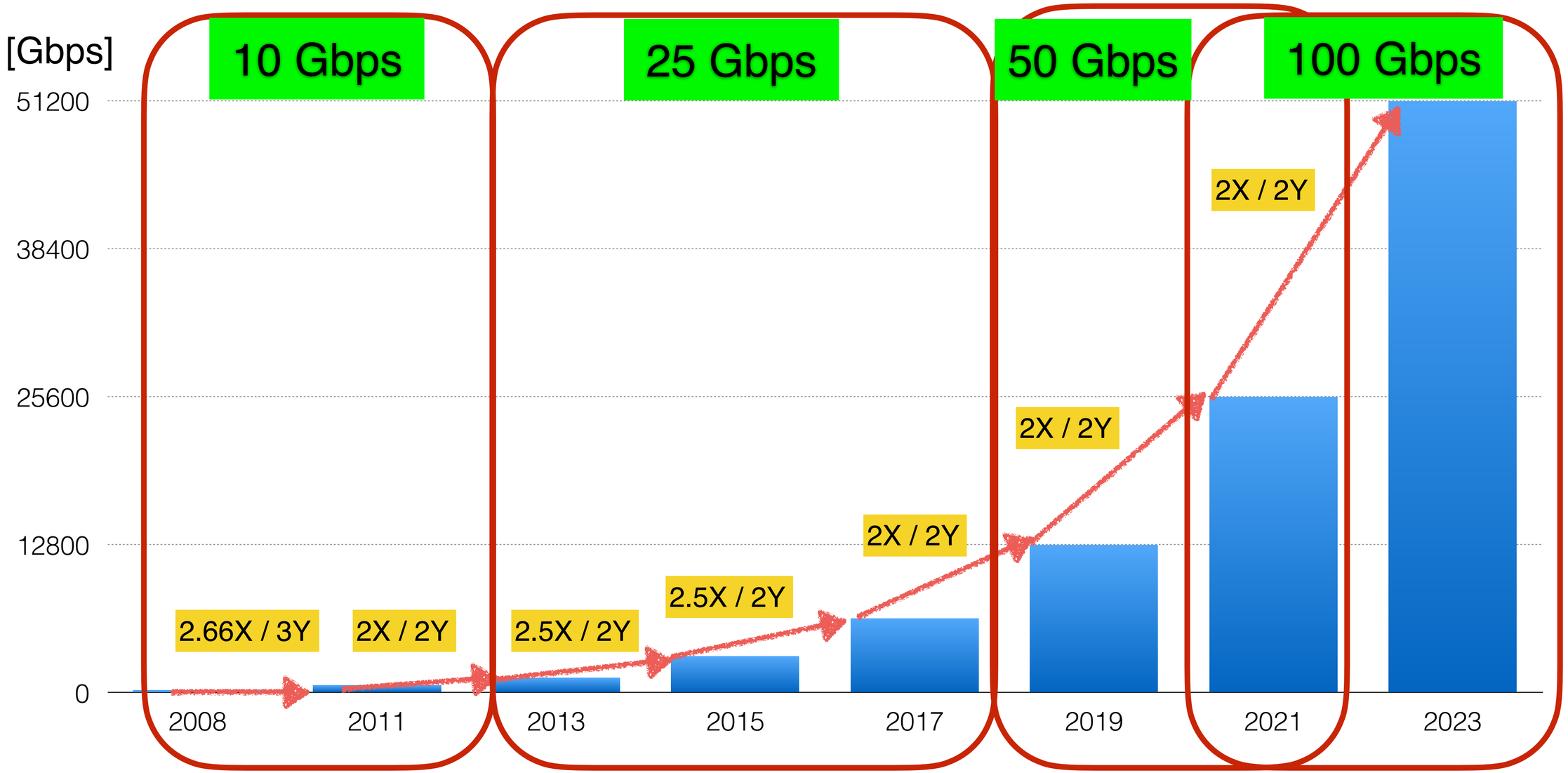
Volume Economics do not favor Custom Silicon

Serdes Roadmap

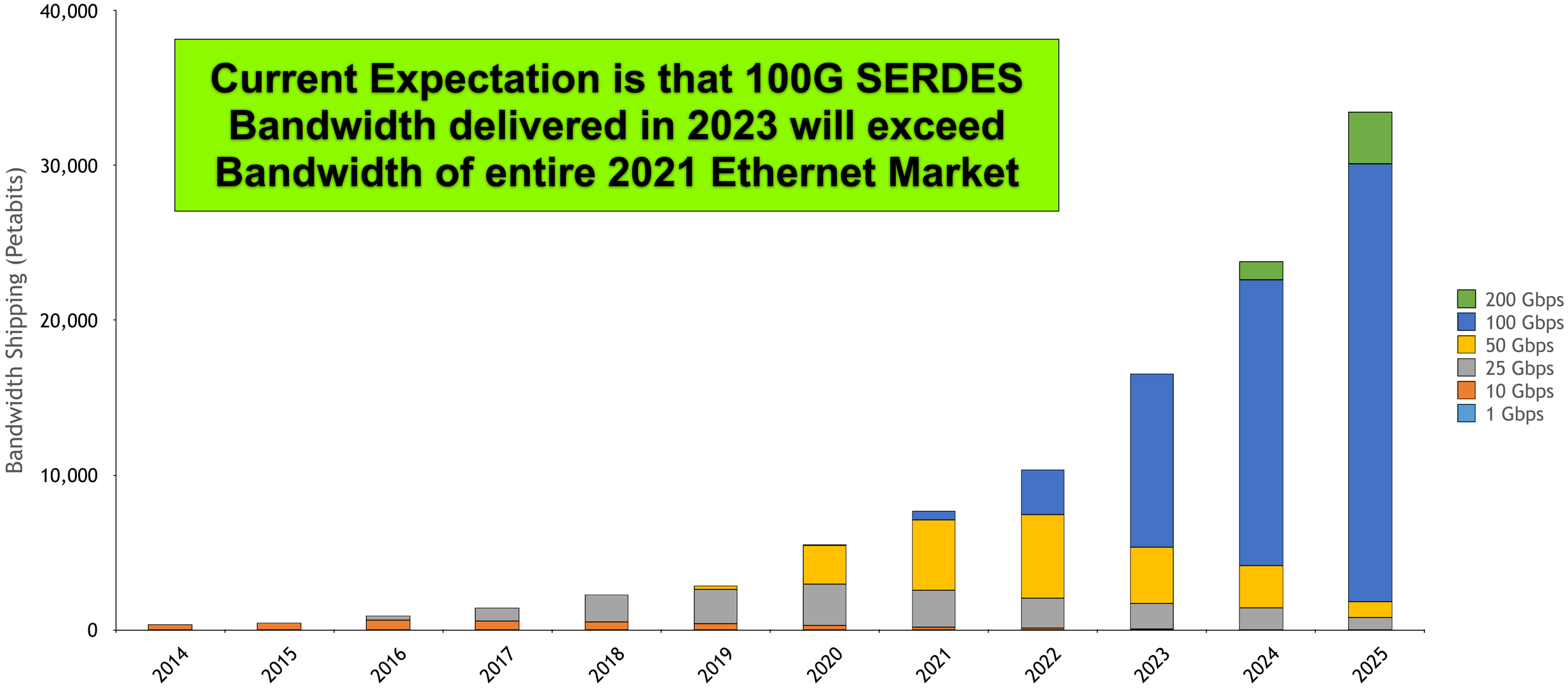
Ethernet Speed Transitions by Serdes Speed

Lane Speed	10Gbps	25Gbps	50Gbps	100Gbps	
1X	10G	25G	50G	100G	Server Interface
2X	—	50G	100G	200G	
4X	40G	100G	200G	400G	Leaf-Spine Interface
8X	—	—	400G	800G	
Availability	2011	2015	2019	2021	

Mapping Silicon to SerDes Transitions



Bandwidth By Switch Silicon SerDes Speed



SerDes Summary

Industry is in midsts of transition to 100G SerDes

All next-generation switch chips use 100G SerDes

100G Serdes Works Best with 100G Lambda Optics

Anything else requires gearboxes, adding cost and power

100G SerDes expected to drive very high volume

1 Billion Switch 100G Channels 2021 through 2031

100G Serdes Enables New Optics

100G-SFP: Densest, lowest cost 100G Optics

One Lane of 100G-Lambda (100G-DR, ER, LR, SR)

400G-QSFP: 4 lanes of 100G Lambda

Fully compatible with legacy QSFP form factor

800G-OSFP/DD: 8 lanes of 100G Lambda

Highest Density and Lowest cost per bit

Data Center Optics Technology Evolution

- NRZ for up to 100Gb/s I/O
- 50Gb/s PAM4 for 400Gb/s I/O
- 100Gb/s PAM4 for 800Gb/s I/O

Per lane speed (Gb/s)

100
50
25
10

Mod: 10G NRZ
Lane: 1
Power: <1W

56.5 x13.4 x 8.5
10G SFP+

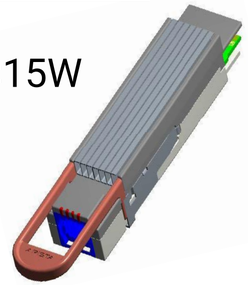
Mod: 10G NRZ
Lane: 4
Power:< 2.5W

40G QSFP+

Mod: 25G NRZ
Lane: 4
Power:<3.5W

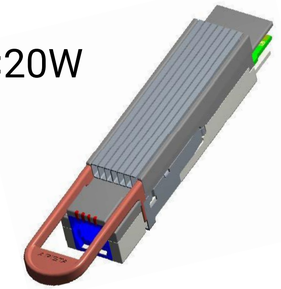
100G QSFP28

Mod: 50G PAM4
Lane: 8
Power:<15W



400G OSFP

Mod: 100G PAM4
Lane: 8
Power:<20W



800G OSFP

2007
2010
2014
2017
2020

Year

Google

Optics Roadmap

Optics for Different Distances

Speed	0-100m	0-2km	2-10km	10-40km	40-1000km
100G	100G-SR	100G-DR	100G-LR	100G-ER	N/A
400G	400G-SR4	400G-DR4	400G-LR4	400G-ER4	400G-ZR/ZR+
800G	800G-SR8	800G-DR8	400G-LR8	400G-ER8	800G-ZR/ZR+

Except for ZR/ZR+ all of these are 100G Lambda Optics

Next-Gen 100G Datacenter Optics Standards

Name	Fiber	Reach	Modulation
100G-ZR/ZR+	Duplex SMF	40km-1000km	16-QAM
100G-BiDi	Single SMF	40km	100G-PAM4
100G-ER	Duplex SMF	40km	100G-PAM4
100G-LR	Duplex SMF	10km	100G-PAM4
100G-DR	Duplex SMF	2km	100G-PAM4
100G-SR	Duplex MMF	50m	100G-PAM4

400G Datacenter Optics Standards

Name	Fiber	Reach	Modulation
400G-ZR/ZR+	Duplex SMF	40km-1000km	16-QAM
400G-ER4	Duplex SMF	40km	4x100G-PAM4
400G-LR4	Duplex SMF	10km	4x100G-PAM4
400G-FR4	Duplex SMF	2km	4x100G-PAM4
400G-DR4	8xSMF	2km	4x100G-PAM4
400G-SR4	8xMMF	50m	4x100G-PAM4

800G (Dual 400G) Optics Modules

Name	Fiber	Reach	Modulation
800G-ZR/ZR+	Duplex SMF	40km-100+km	16-QAM
800G-ER8	Duplex SMF	40km	8x100G-PAM4
800G-FR8/LR8	Duplex SMF	2km/10km	8x100G-PAM4
2x400G-FR4/LR4	2xDuplex	2km/10km	8x100G-PAM4
800G-DR8	16xSMF	2km	8x100G-PAM4
800G-SR8	16xMMF	50m	8x100G-PAM4

Why 800G Optics?

1. **Lower Cost Per Bit**

   Expectation is 33% to 40% lower than 400G (8x50G)

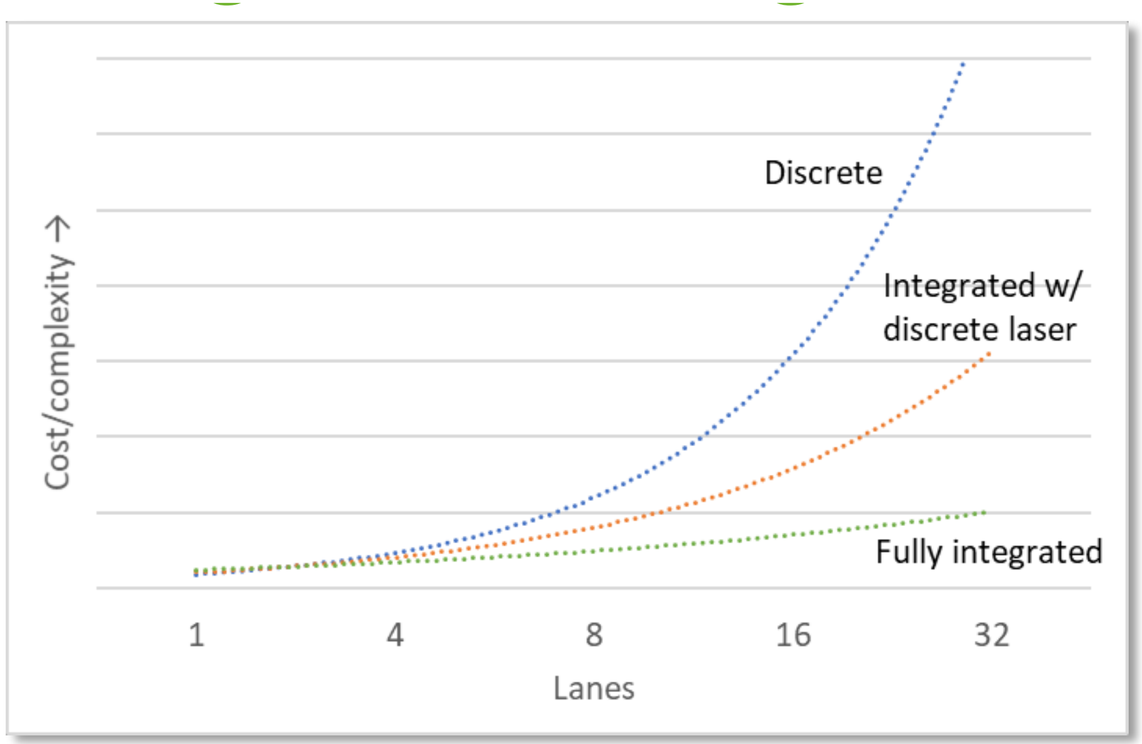
2. **Lower Structural Cost per System**

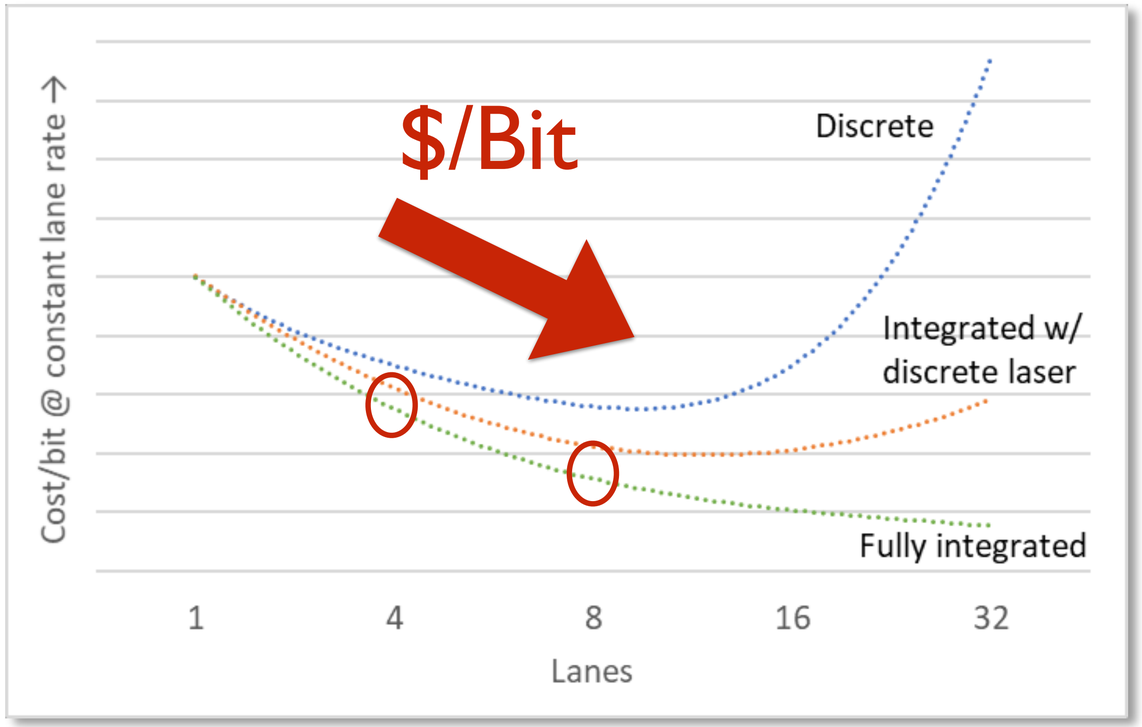
   Smaller Chassis, Fewer Connectors, fewer PCB, etc

3. **Fully Compatible with 400G Optics**

   Dual Fiber Connectors to match existing fiber cabling

Silicon Photonics Changes the Game





Source: Intel

8-Channel optics are expected to be 33%-40% lower per Bit

Future 51.2T Switch with 64 800G-OSFP

Supports 64 800G or 128 400G-FR4/LR4/DR4 in 2U
No co-packaged Optics Required

800G Optics and 800G Ethernet

1. **800G Optics Driven by 100G Serdes Ecosystem**

   800G Optics will be in the market and ramping in 2021

2. **800G Ethernet spec was completed April 2020**

   Work done by Ethernet Technology Alliance

3. **Expect first 800G Switches and Routers in 2023**

   First time in history optics will be ready before 800G MAC

Backward Interoperability

Three Levels of Interoperability with Installed base

1. Optical: Addressed with Multi-speed Optics
2. Physical: Addressed with compatible form factors
3. Fiber Connectors: Support existing Connectors

Optical Interoperability can be achieved with Multi-speed Optics across Different Form Factors

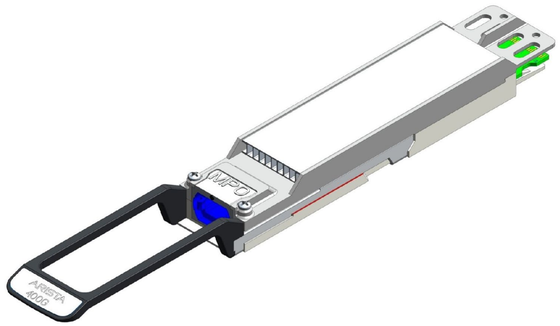
Optical Interoperability across Optics Form Factors

1st Generation
QSFP-100G-DR
Higher power
Higher cost

Optically
Interoperable

2nd Generation
SFP-100G-DR
Lower power
Lower cost

Multi-speed 400G-FR4 Optics



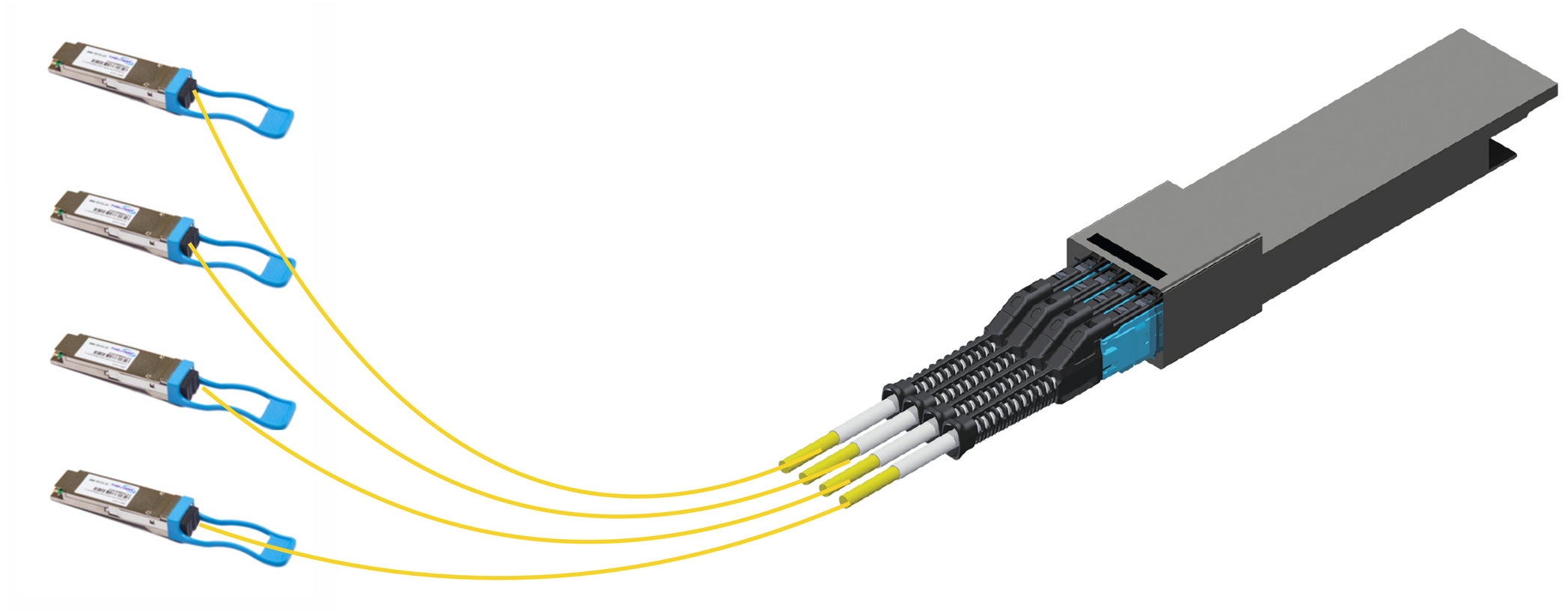
400G-FR4

200G-FR4

100G-CWDM4

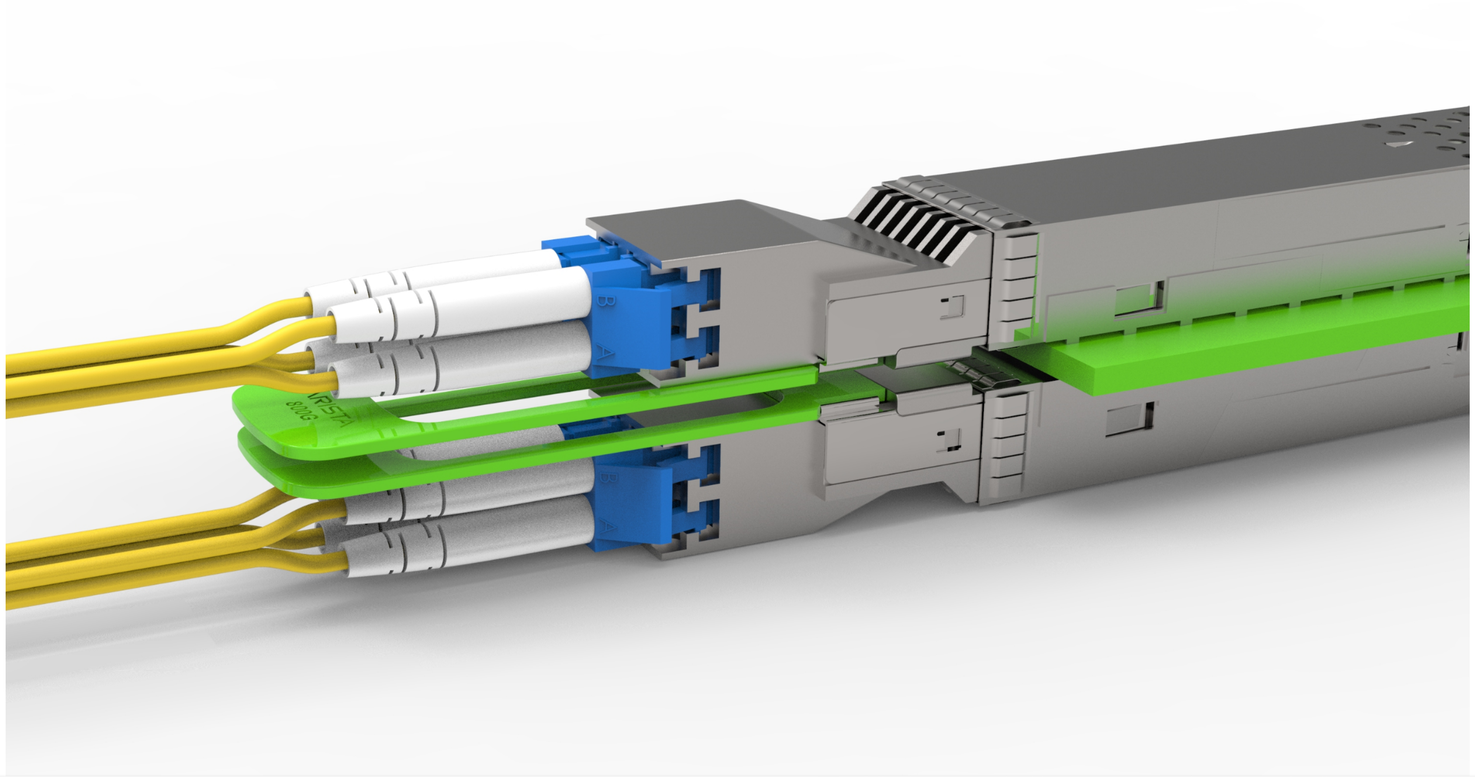
Multi-speed 400G-FR4 can interoperate with installed base of 200G-FR4 and 100G-CWDM4

400G DR4 with Four Duplex 100G-DR Fibers



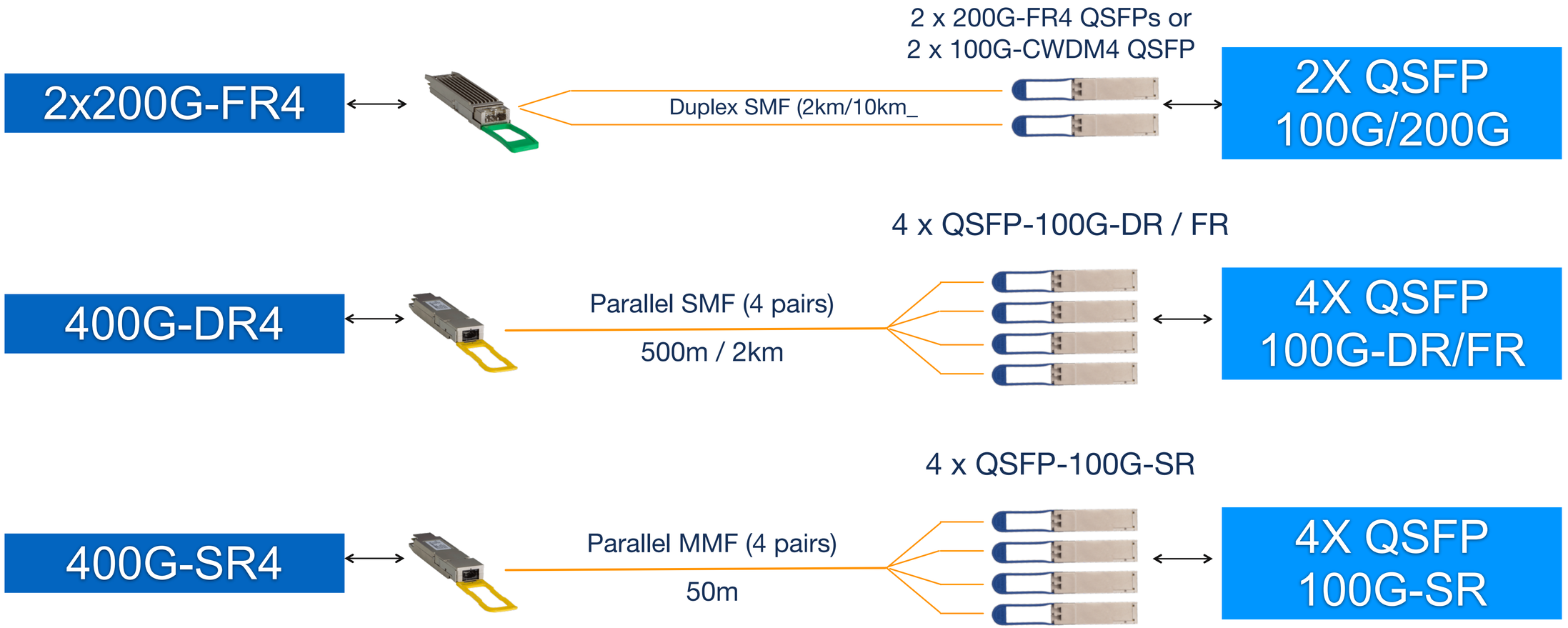
New SN/MDC connector avoids MPO/MTP Splitter Cable

Dual LC Connector for 2x400G-FR4



2 LC Connectors to support Dual 400G-FR4 in OSFP

400G Breakout Options: 2x200G and 4x100G



Datacenter Optics Summary

1. **All Future Switches will use 100G SerDes**

   => Directly compatible with 100G Lambda Optics

2. **100G Serdes enables new Optics Modules**

   => 100G-SFP, 400G-QSFP and 800G-OSFP/DD

3. **Multi-speed Optics are Backwards Compatible**

   => Can be configured in software for legacy speeds

200G Serdes Roadmap

1. **Early discussion on 200G SerDes**

   => Will require new FEC, new optics ecosystem

2. **First Focus is on 800G (4x200G) Optics**

   => First samples expected in 2024

3. **4x200G Optics tied to Adoption of 800G Ethernet**

   => 400G expected to remain with 100G Lambda Optics

Except First 200G Serdes on 102.4T Switch Chip in 2024

Extended-Reach Optics: 10-40km and 40-100km+

Two Main Categories

100G-ER
400G-ER4
800G-ER8

Up to 800G/Fiber
Up to 40km Reach
No Amplifcation

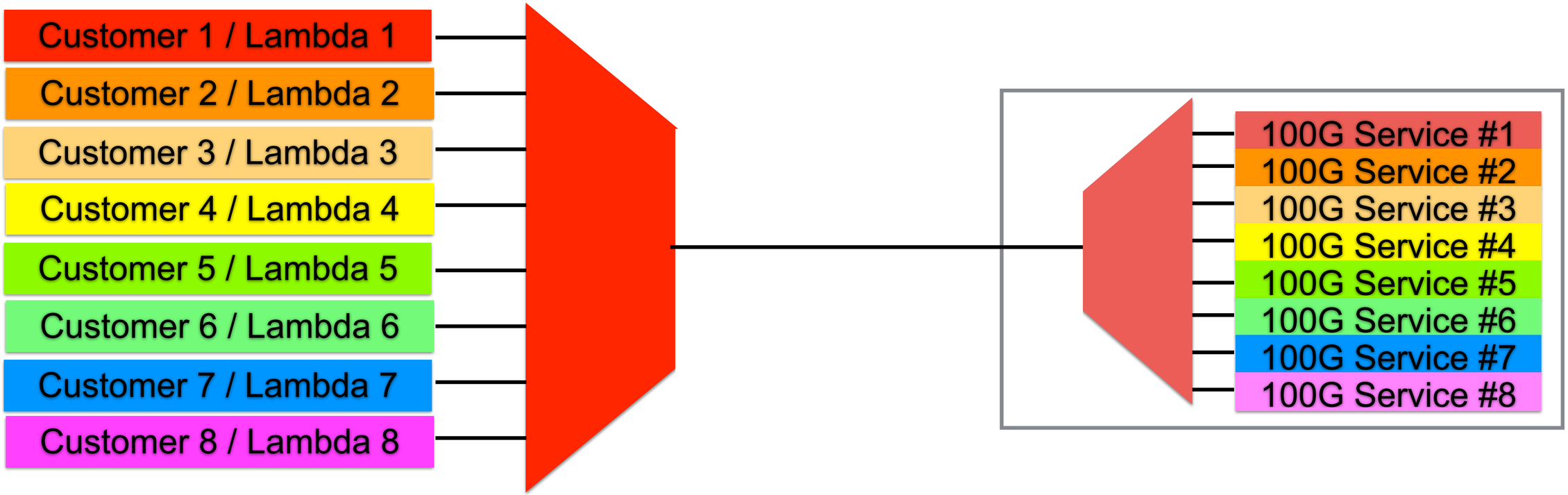
400G-ZR
400G-ZR+
400G-ZR++

Up to 25.6T/Fiber
100km+ Reach
External Amplification

100G-ER/400G-ER4 10-40km Optics

1. **400G-ZR is overkill for many Access Use Cases**
   100G-ER better match for native 100G Speeds
2. **100G-ER/400G-ER4 do not require optical amp**
   Simpler, more reliable, lower power
3. **100G-ER can support up to 8 Colors DWDM**
   Up to 800G with 100G over duplex fiber

8x100G-ER Aggregation Use Case

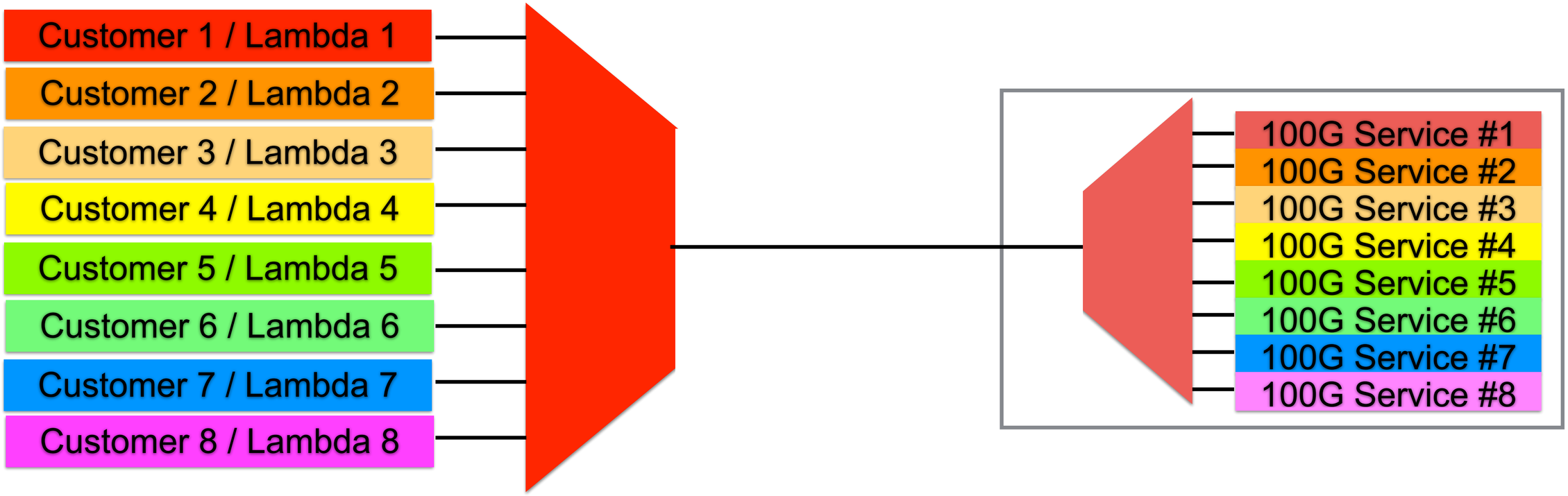


8 Traffic Sources @ 100 Gbps each

Up to 40km Reach

800G-ER8 Module in Aggregation Router

8x100G-ER Aggregation Use Case



Passive Optical Mix replaces Field Aggregation Router

100G-ER/400G-ER4 Summary

1. **100G-ER in QSFP Form Factor will be first**

   Supports up to 8 Lambda over duplex fiber

2. **400G-ER4 in OSFP/DD will be second**

   Supports 4 Lambda over duplex fiber

3. **800G-ER8 in OSFP/DD will be third**

   Supports 8 Lambda over duplex fiber

100G ER1-30/40 QSFP28 :

100G-LR1-20, 100G-ER1-30 and 100G-ER1-40 Technical Specifications Rev 0.1

100G-LR1-20, 100G-ER1-30 and 100G-ER1-40 Technical Specifications

100G Lambda MSA

- 100G Lambda MSA Compliant: Baseline spec. accepted.
  - Tx wavelength: 1309+/-1nm
- QSFP28: 4x25Gbps Host Interface, replace ER4/ER4-Lite
- Support PtP 30/40km Reach defined by MSA
- Power dissipation < 3.5W
- Key Components:
  - Mature 53Gbaud EML
  - 53GBaud Ge/Si APD
  - DSP
    - Strong equalizer capability
    - Option with Build-In Strong FEC

400G-ER4 Power and Cost Comparison

Power Dissipation Considerations

	400G ZR (Coherent)	400G ER8	Proposed 400G ER4-30
Laser/Modulator	6W	4W	3W
Receiver	3W	3W	2W
DSP	8W (7nm DSP)	5W (16nm DSP)	4W (7nm DSP)
Others	1W	1W	1W
Total	~17-18W (target)	13W (<14W)	~10W

Relative Cost Estimates:

	400G ZR (Coherent)	400G ER8	Proposed 400G ER4-30
Laser/modulator	$$$$$$$	$$$$$$	$$$
Receiver	$$$$	$$$$	$$$
DSP	$$$$$$	$$	$$
Others	$$	$	$
Total	$$$$$$$$$$$$$$$$$$$$(19$)	$$$$$$$$$$$$ (13$)	$$$$$$$$$ (9$)

400G-ER4/100G-ER Positioning

1. **Lowest Power and Lowest cost for 10-40km**

   About 50% of power and cost compared to 400G-ZR

2. **Up to 800G (8 100G Lambda) over duplex fiber**

   In contrast, 400G-ZR supports 64 Lambda or 25.6T

3. **No Amplification required for 10-40km**

   400G-ZR requires amplified optical line system

400G-ZR and 400G-ZR+ Pluggable DCO Optics

400G-ZR/ZR+ Optics

Multi-Vendor DWDM Standard

Pluggable Small Form Factor

Same Density as Client Optics

Low Power Consumption (20W)

Revolutionary Price Performance

Use cases from 10km to 1000km and Beyond

400G-ZR/ZR+ Enables IPoDWDM

1. **IPoDWDM Concept has Existed for Years**

   Traditionally with 70% to 90% density penalty, specialized line cards and single vendor line cards

2. **400G-ZR/ZR+ Finally Gets this Right**

   Multi-vendor interoperable standards based solution with same high density client optics form factor

Order of Magnitude Cost-Reduction with 400G-ZR/ZR+

Coherent Pluggable to Transform Optical Transport Market by 2024

BOSTON (September 24, 2020) – Adoption of coherent pluggable optics will accelerate in 2021 as new technologies reach the market, according to the latest Transport Applications Report from networking component and equipment market research firm Cignal AI. Fourth-generation coherent will extend beyond 400ZR to include higher performance solutions generally referred to as 400ZR+ and lower speed 100Gbps (100ZR) targeted at the metro edge.

“Standardized pluggable coherent optics, coupled with open line systems and network control, will represent a major change in network design,” said Andrew Schmitt, Directing Analyst for Cignal AI. “This shift will begin in earnest in 2022.”

World's First 400G-ZR IPoDWDM Network

Link-up January 2020 at Arista Networks

Three Types of 400G-ZR

1. **Mainstream 400G-ZR (80-120km, amplified)**

   Application Code 0x01, High volume DCI use case

2. **Grey Optics 400G-ZR- (40km pt-pt, non-amplified)**

   Application Code 0x02, Competes with 400G-ER4

3. **Enhanced 400G-ZR+ (300-600km, Open-ZR+)**

   Metro Reach with enhanced FEC

400G-ZR Optical Specs

Optics Module	Min Output Power	Max Output Power	Min Rx Sensitivity	Link Budget non-amplified
400G-ZR 0x01	-10 dBm	-6 dBm	-12 dBm	2 dB
400G-ZR 0x02	-9 dBm	0 dBm	-20 dBm	11 dB

400G-ZR App Codes 0x01 and 0x02 are not the same Module

0x02 is designed with grey optics, higher TX power and higher RX sensitivity to achieve 40km reach with 0.25 db/km fiber.

0x02 overlaps with 400G-ER4 for the 40km reach application.

400G-ZR/ZR+/ZR- Summary

1. **Industry Focus is to get 400G-ZR into production**

   Largest immediate market opportunity

2. **Next Focus is 400G-ZR+ (Open-ZR+ Standard)**

   Same DSP silicon but more validation work needed

3. **400G-ZR- Grey Optics will be there (40km pt-pt)**

   Higher cost and higher power than 400G-ER4 optics

400G-ZR+ Status

1. **Open ZR+ Specification nearing completion**

   Based on OFEC (driven by Acacia and NEL DSP)

2. **Current 400G-ZR DSP Silicon is Open-ZR+ Capable**

   However power so far is significantly above target

3. **Expect 400G-ZR+ Availability in 2H2021**

   Compared to Q1’2021 for 400G-ZR

Roadmap to 800G-ZR Optics

Double the Density

Lower Cost per Bit

Lower System Cost

Better OSNR Performance

Compatible with 400G-ZR

Primary Target is 80km DCI Use Case

Arista 400G-ZR Support

Arista will qualify all leading optics vendors
Both OSFP and QSFP-DD Form Factors
Arista EOS supports OpenConfig with gRPC
Full Set of Telemetry Performance Monitoring
IPoDWDM eliminates separate Optical Shelves

400G-ZR will transform the Optical Transport Market

Optical Line Systems for 400G-ZR and 400G-ZR+

Background

400G-ZR/ZR+ Standard has unique Characteristics

Single 400G Data rate
75 GHz channel spacing with ~ 60 Gbaud 16QAM
-6 to -10dBm minimum TX power
-12 dBm RX sensitivity

400G-ZR requires new OLS. Deploying 400G-ZR with brownfield OLS is beyond the scope of the standard

Enhanced 400G-ZR+ (2H2021)

Enhanced 400G-ZR+ SKUs with higher Output Power

0 dBm minimum TX power
-12 dBm RX sensitivity
Enhanced FEC modes to support 400G over 1000km
200G-16QAM Mode with 30 Gbaud to fit 50 GHz grids

Enhanced 400G-ZR SKUs can be used with brownfield optical line system deployments

Primary Use Case for 400G-ZR is DCI

1. **Traditional Optical Line systems are overkill for DCI**

   Require significant rack space and power

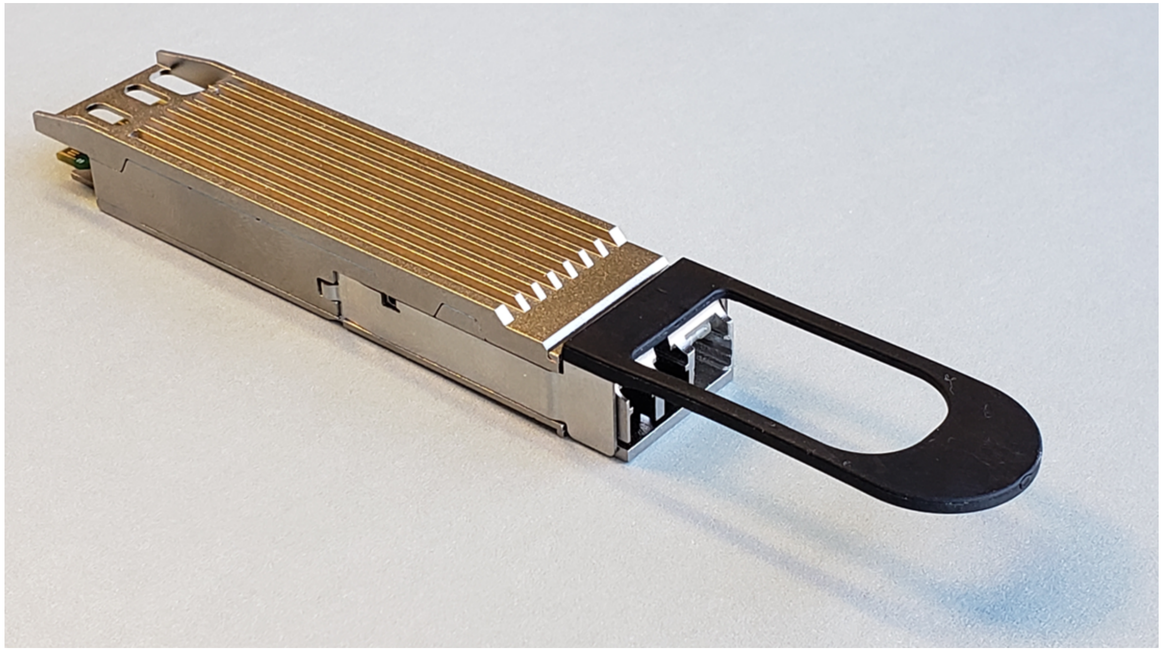
2. **Traditional OLS are complicated to configure**

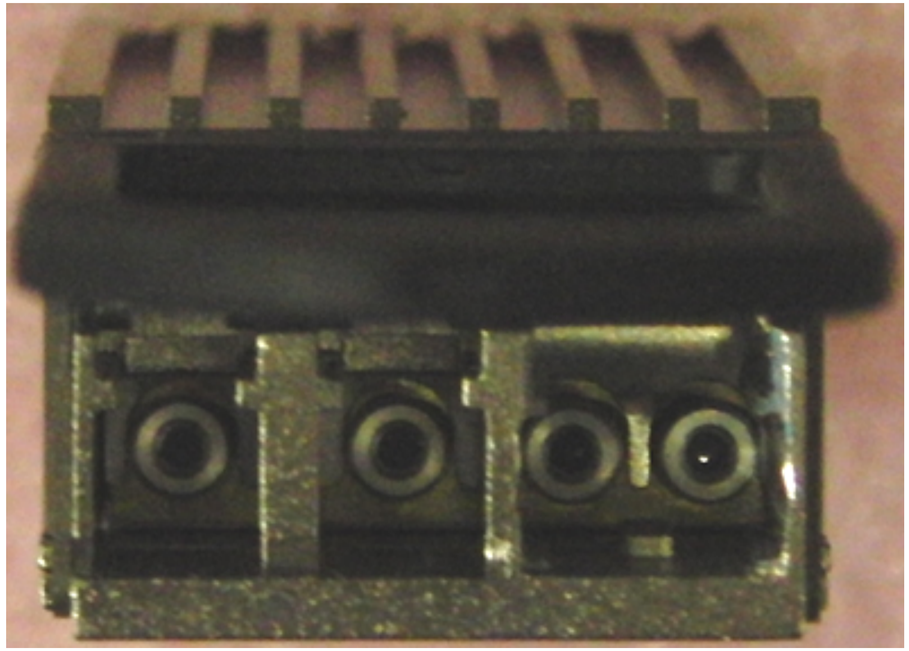
   Require optical expertise, separate management stack

3. **A Much Simpler Solution is Needed**

   Something mere mortals can use

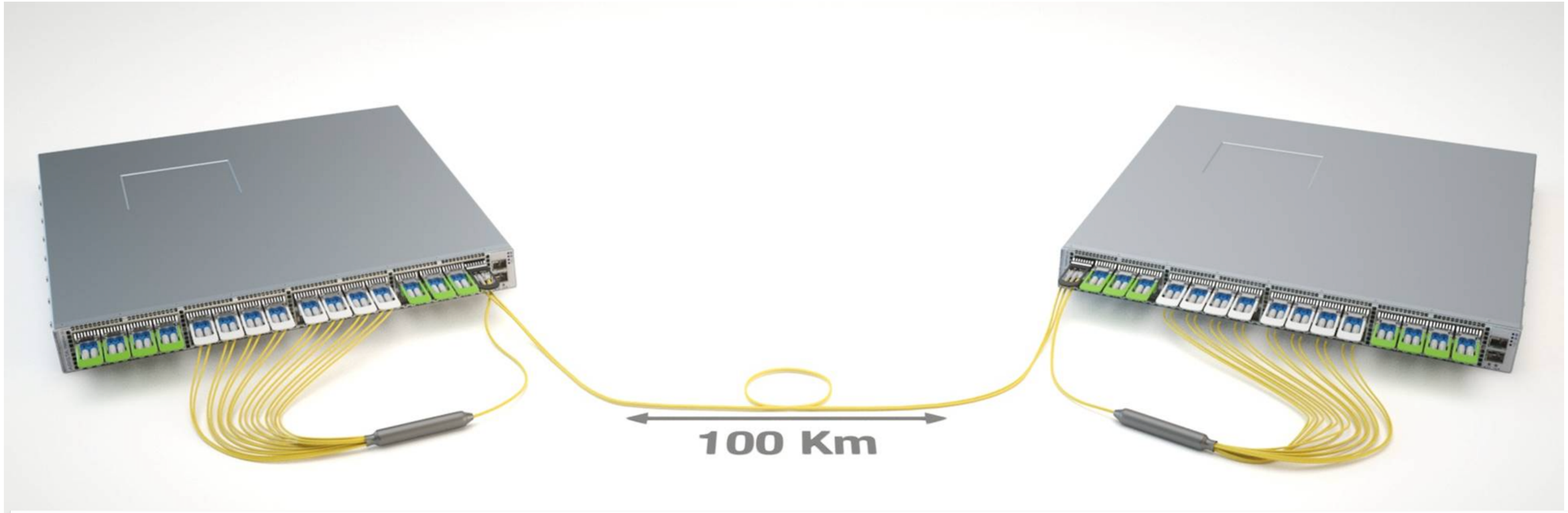
The Arista OSFP Line System (OSFP-LS)





8:1 Colorless MUX + Duplex Amplifier in OSFP Form Factor

World's First Integrated Optical Line System

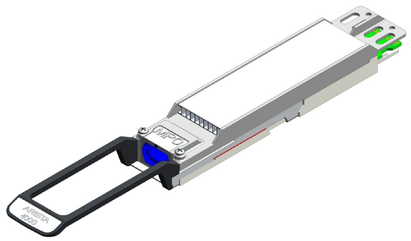


Plug-and-play Optical Line System for mere Mortals

Arista's Optical Line System Replaces This

+

=



8:1 DWDM MUX + Duplex Amplifier in OSFP Form Factor

Optical Line System Summary

1. **Using vanilla 400G-ZR with brownfield OLS**

   This is outside of the scope of the standard

2. **Using enhanced 400G-ZR with brownfield OLS**

   This will be possible with legacy OLS vendor support

3. **Using vanilla 400G-ZR with Arista OLS module**

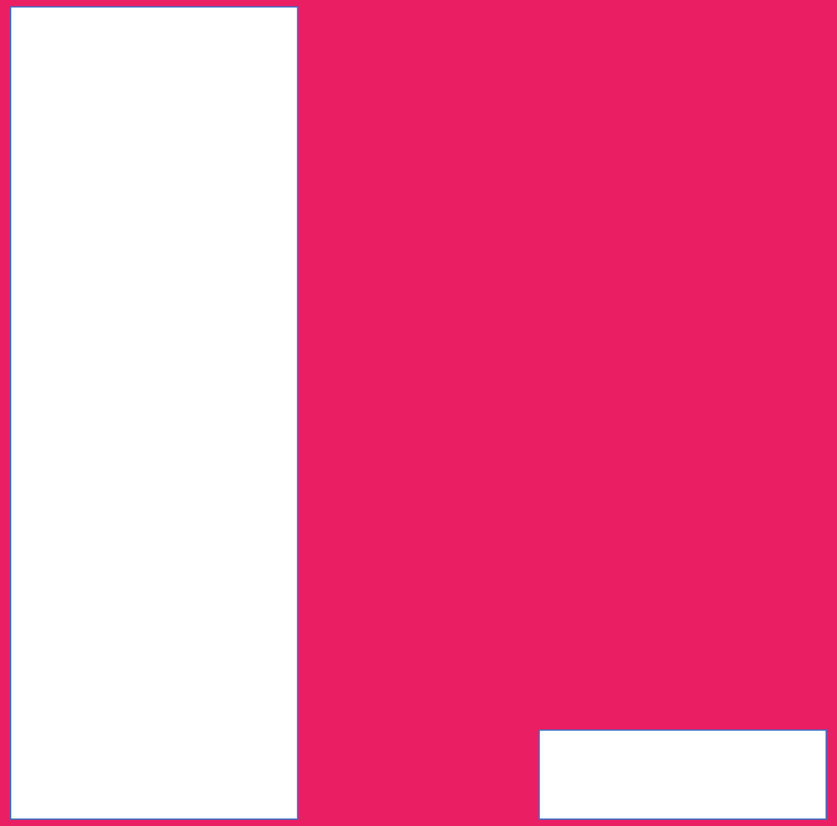
   Plug and play solution for mere mortals

Summary

400G Router Price per Port

10X Improvement in Price-Per Port with with Merchant Silicon Routers compared to legacy Router Price Points

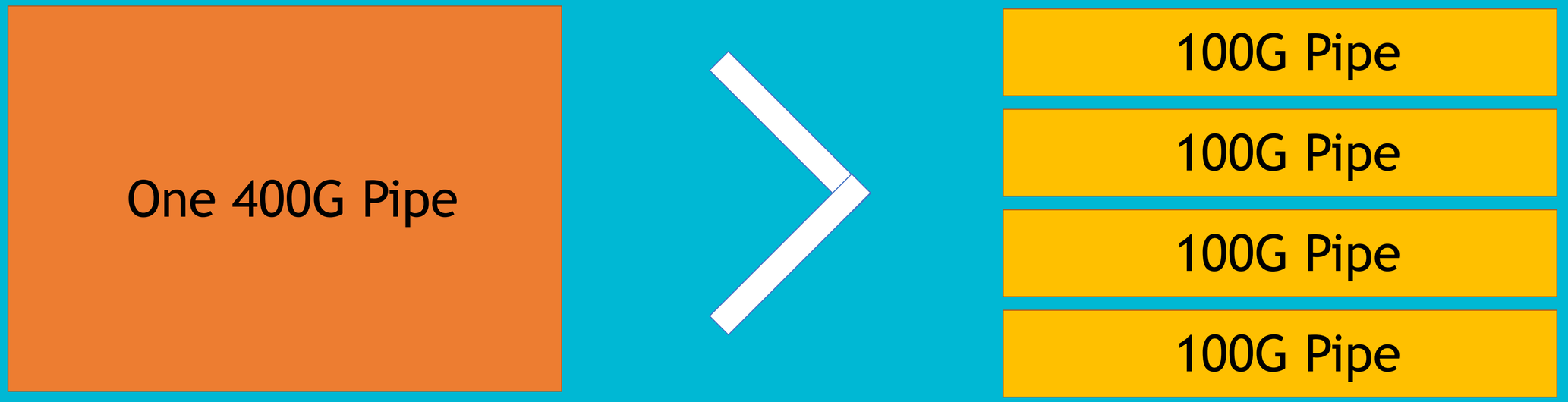
400G DWDM Price Per Bandwidth



Order of Magnitude Cost-Reduction with 400G-ZR/ZR+ compared to legacy Optical Transport Price Points

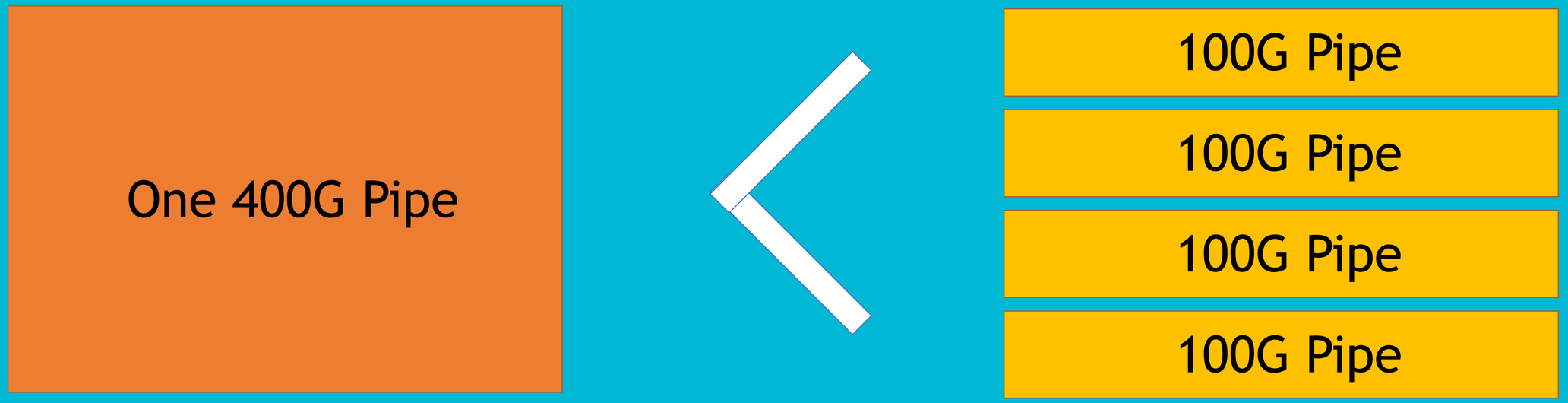
Fatter Pipes are Easier to Manage

Fatter Pipes are more efficient and easier to manage than equivalent bandwidth with smaller pipes



Fatter Pipes are Lower Cost per Bandwidth

400G is fundamentally lower cost than 4x100G



400G Summary

Order of Magnitude Improvement in Price-Performance in Equipment and Optical Transmission Costs

Enables Fundamentally more cost-effective Networks that are also more efficient and easier to manage

Timeline: Field Trials late 2020, Production 2021