Ethernet Alliance
TEF 2021: The Road Ahead

New Applications Driving Higher Bandwidths

TECHNOLOGY EXPLORATION FORUM

ethernet alliance

www.ethernetalliance.org

The Ethernet Alliance

Global Community of End Users, System Vendors, Component Suppliers & Academia

Our Mission

- **To promote** industry awareness, acceptance and advancement of technology and products based on, or dependent upon, both **existing and emerging IEEE 802 Ethernet standards** and their management.
- **To accelerate industry adoption** and remove barriers to market entry by providing a cohesive, market-responsive, industry voice.
- Provide resources to establish and **demonstrate multi-vendor interoperability.**

Ethernet Alliance Strategy

Expanding the Ethernet Ecosystem, Supporting Ethernet Development

- Facilitate interoperability testing
  - Industry Plug Fests supporting member and technology initiatives
- Interoperability Assurance
  - PoE Certification Program
- Collaborative Interaction with other Industry Organizations
  - Multiple SIGs, Applications and MSAs
  - Industry Consensus Building
- Global Outreach
  - Worldwide Membership
- Thought Leadership
  - EA Hosts Technology Exploration Forums (TEFs)
  - Technology and Standards incubation
- Promotion of Ethernet
  - Industry Analysts
  - Education
  - Marketing
    - *Trade shows & Panel Presentations*
    - *White Papers, Blogs & Social Media*

NEW APPLICATIONS DRIVING HIGHER BANDWIDTHS

Moderator – Nathan Tracy, *Ethernet Alliance Board Member & TE Connectivity*
Panelists –Brad Booth, *Microsoft,- "Paradigm Shift in Network Topologies"*
Rob Stone, *Facebook – "Co-Package Optics for Datacenters"*
Tad Hofmeister, *Google – "OIF Considerations for Beyond 400ZR"*

ethernet alliance

www.ethernetalliance.org

New Applications Driving Higher Bandwidth

Moderator: Nathan Tracy, *Ethernet Alliance Board Member & TE Connectivity*

ethernet alliance

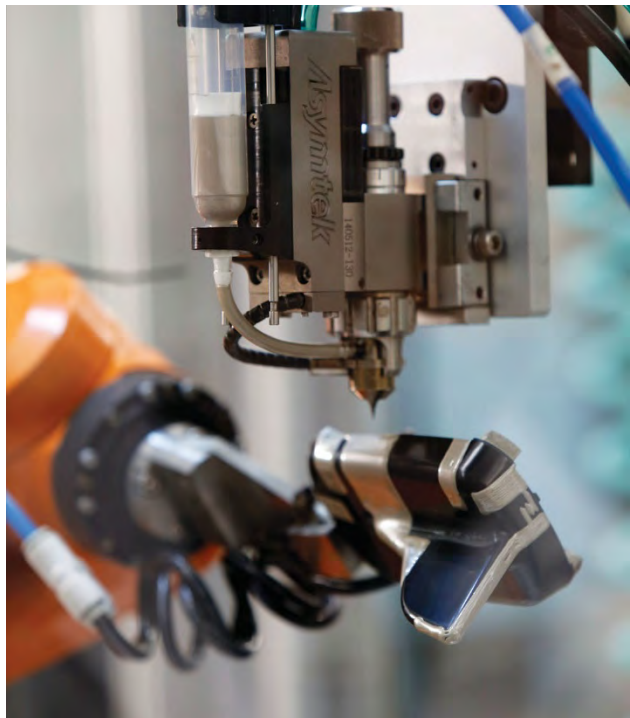
New Applications Driving Higher Bandwidth

or

There's a Bandwidth Problem? Where did that Come From?

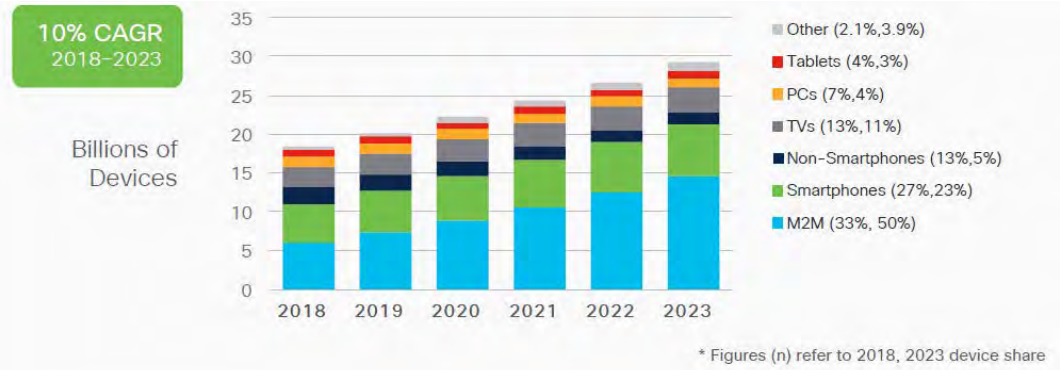
Nathan Tracy
Technologist
TE Connectivity
January 26, 2021

EVERY CONNECTION COUNTS

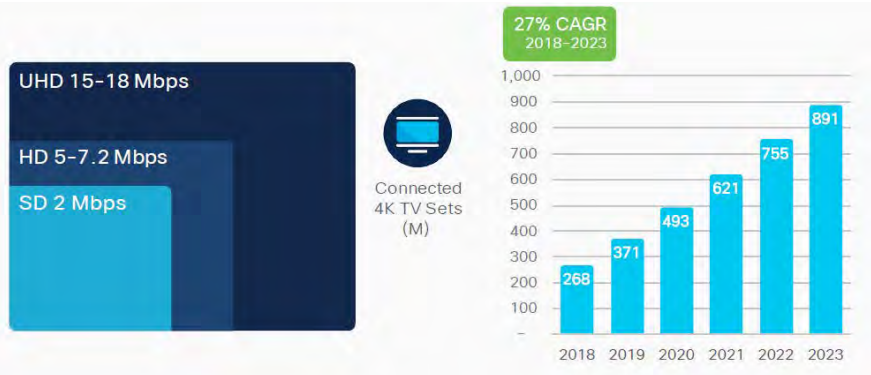


2018-2023 Internet User Trends

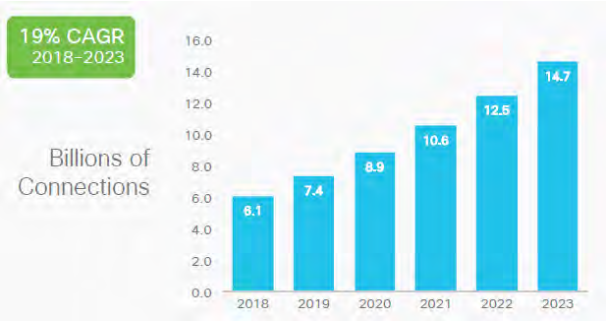
Devices and Connections: growing faster (10% CAGR) than population (1% CAGR) and users (6% CAGR)



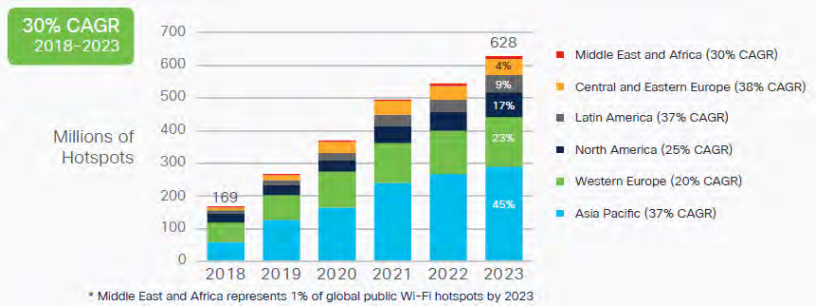
Video: Use of 4K TV's will grow from 33% in '18 to 66% in '23
4K bit rate is double HDTV and 9 times SDTV,



M2M Connections: 19% CAGR, 1.8 connections per capita, by 2023 will be 50% of all connections



WiFi: Speeds from mobile devices will be 3x faster from '18 to '23
Number of hot spots will grow 4x from '18 to '23



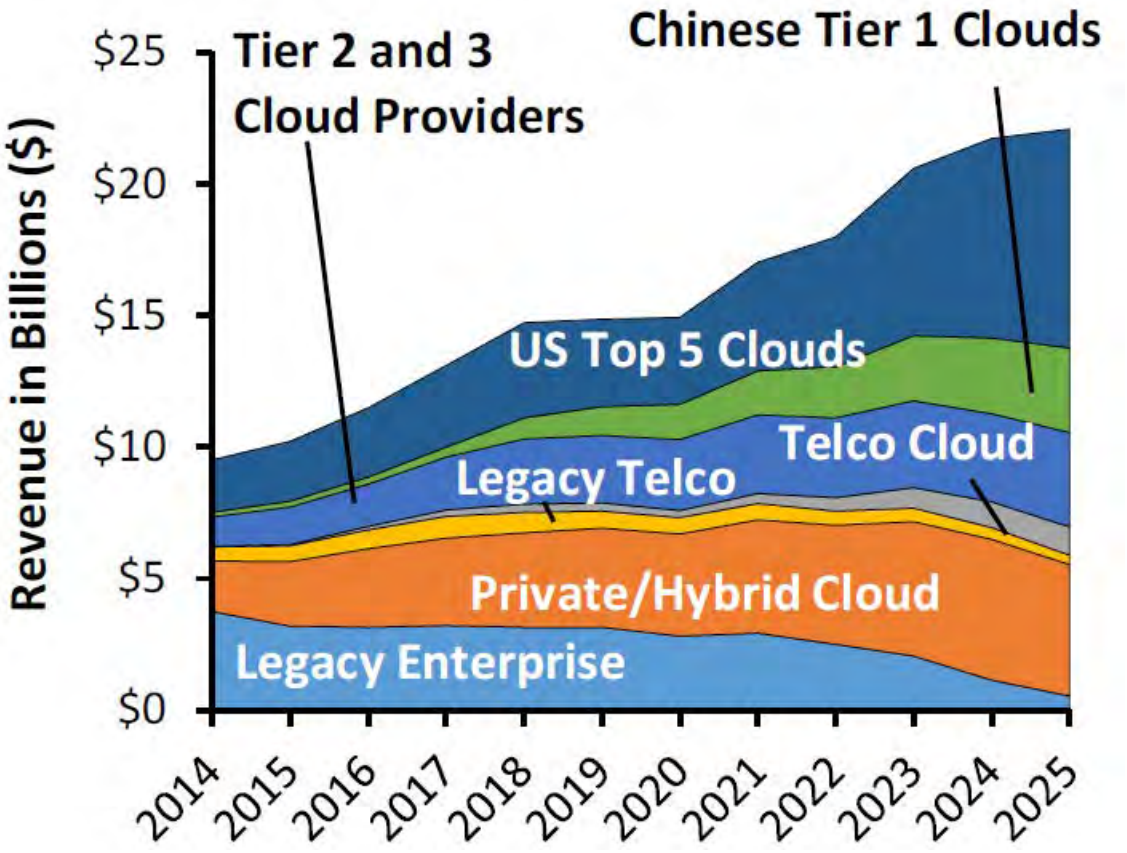
Mobility: >70% of the global population by '23,
5G will be >10% by '23 @13X faster than the avg mobile connection

Source: Cisco Annual Internet Report
https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

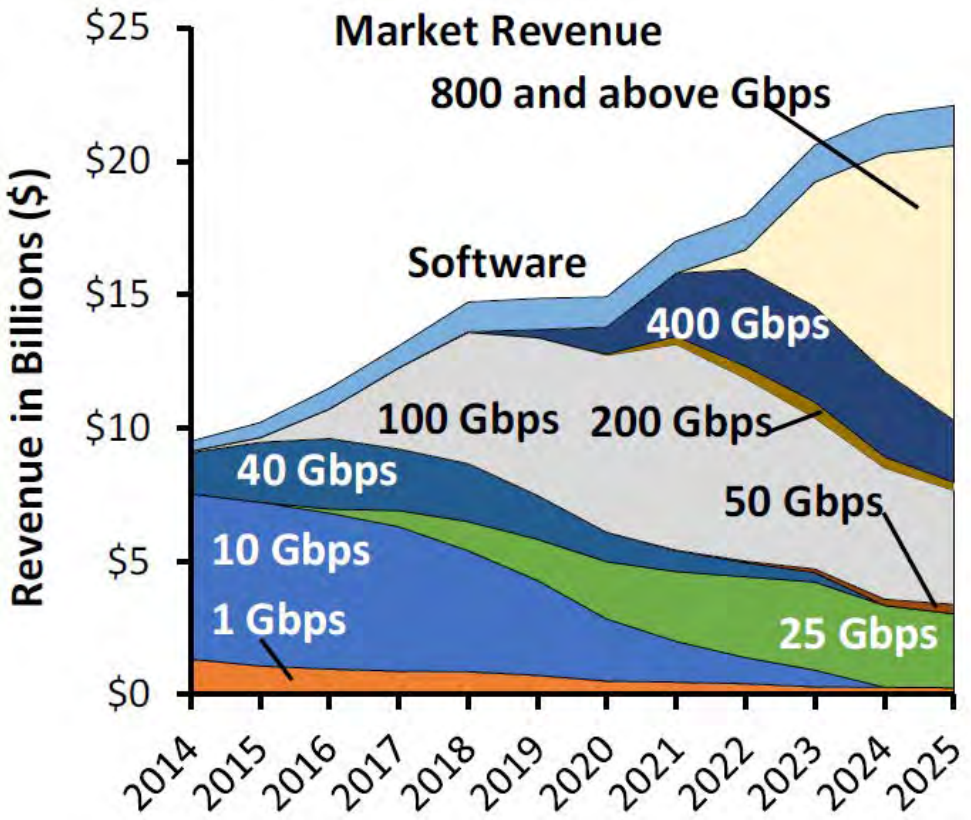
TECHNOLOGY EXPLORATION FORUM 2021 • THE ROAD AHEAD

Ethernet Switch Growth Forecast

Data Center Ethernet Switch
Market Revenue By Application



Data Center Ethernet Switch
Market Revenue By Speed



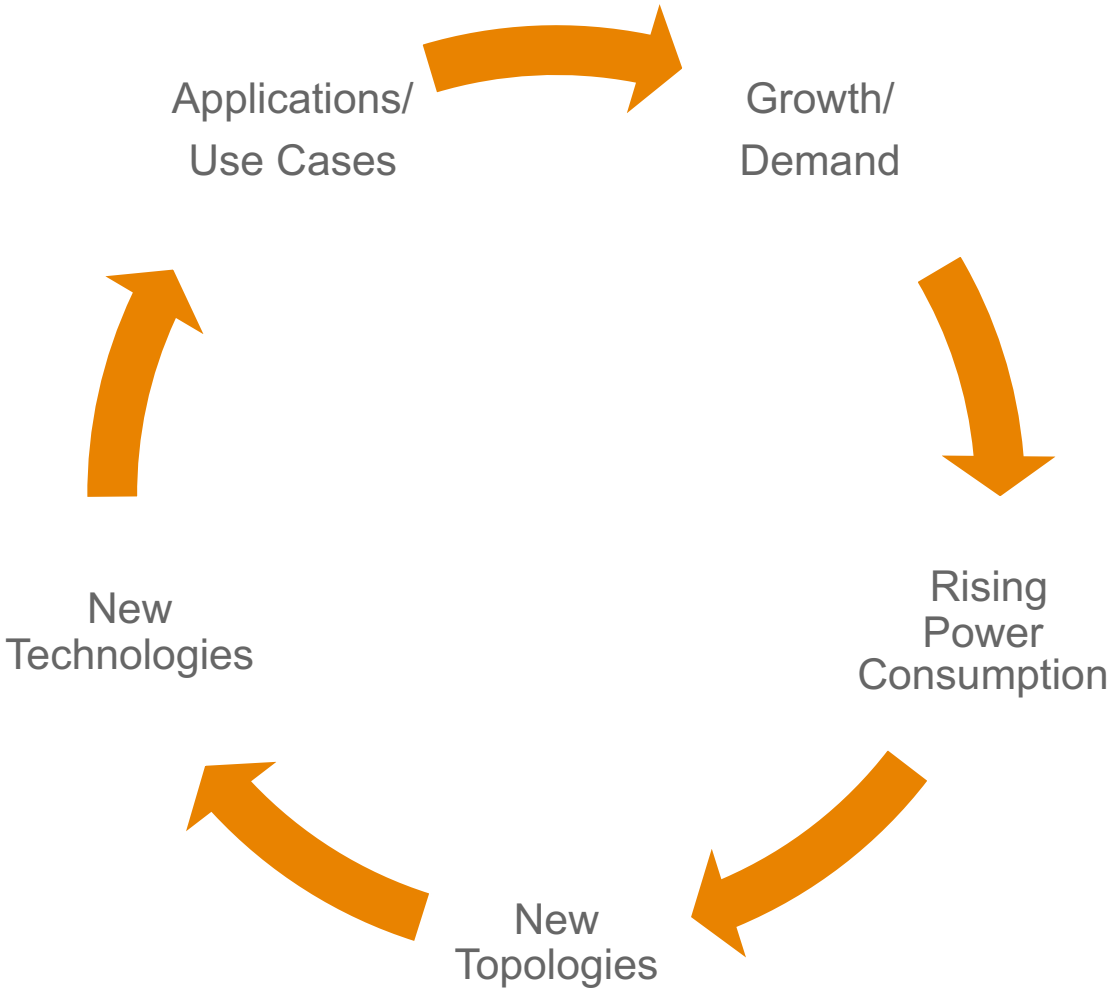
TECHNOLOGY EXPLORATION FORUM 2021 • THE ROAD AHEAD

650 GROUP MARKET INTELLIGENCE RESEARCH

December 2020 data

Dynamics Affecting Network Bandwidth

Where are we at on the “fly wheel”?

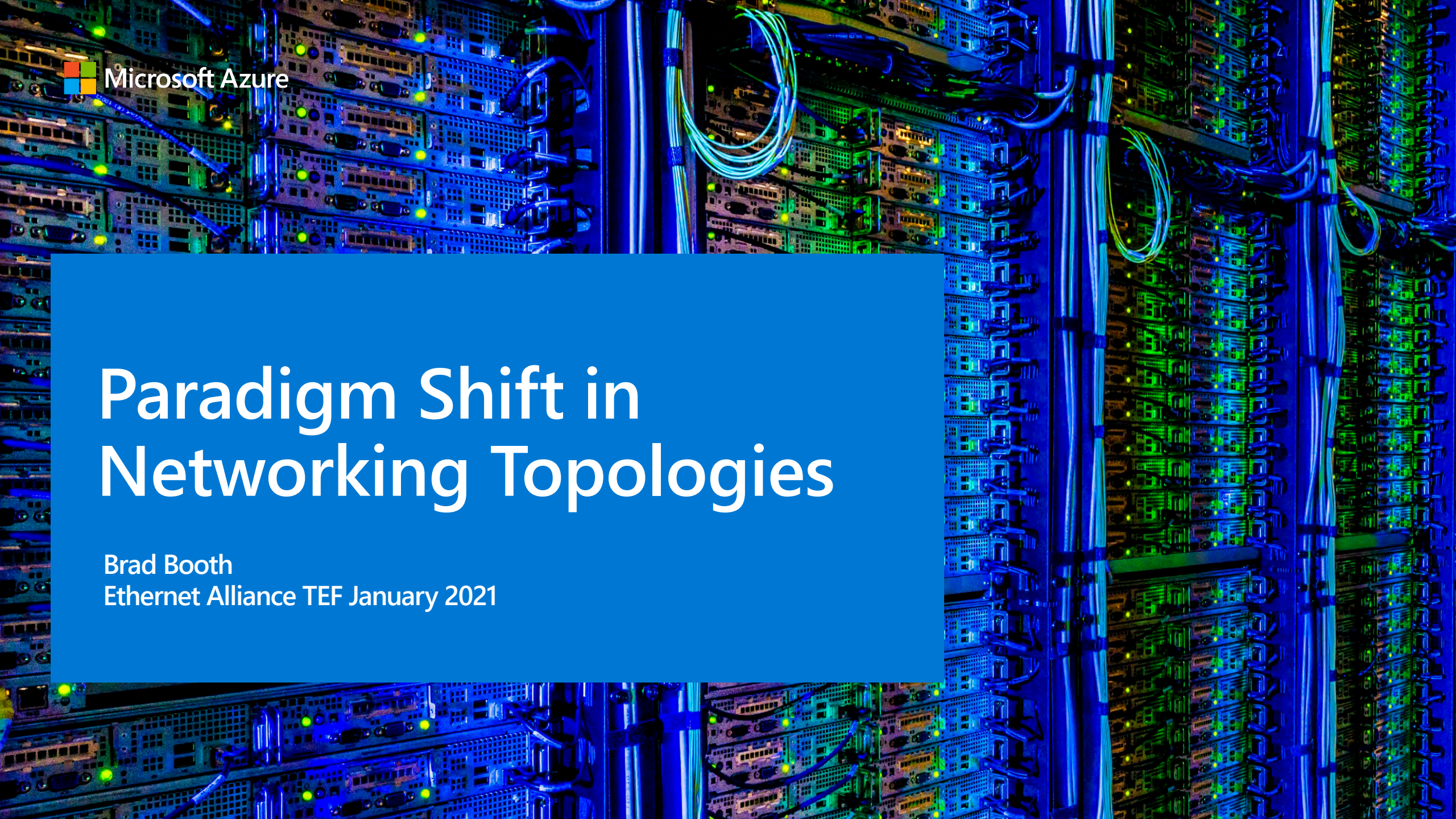


TECHNOLOGY EXPLORATION FORUM 2021 • THE ROAD AHEAD

Paradigm Shift in Network Topologies

Brad Booth, *Microsoft*

ethernet alliance



Paradigm Shift in Networking Topologies

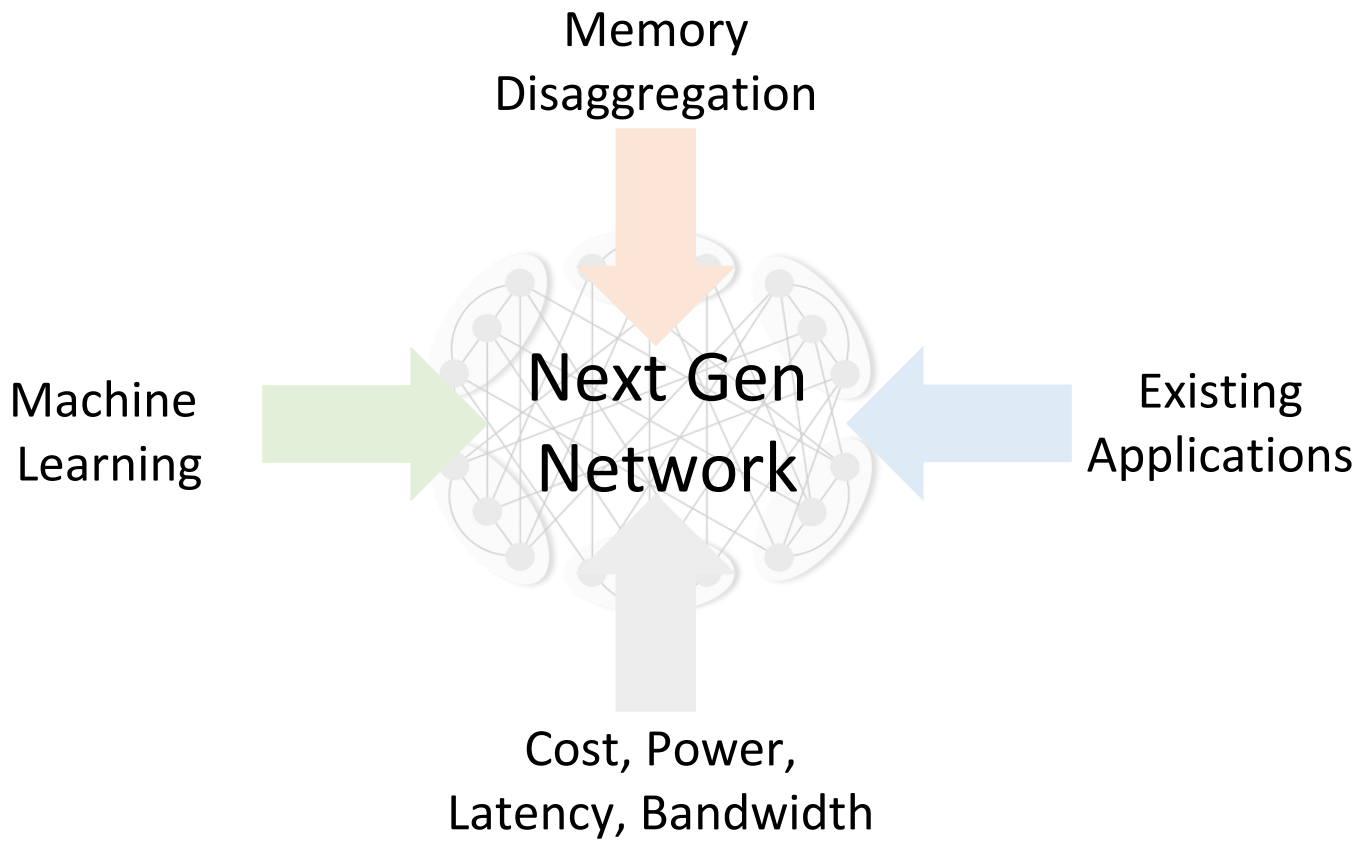
Brad Booth

Ethernet Alliance TEF January 2021

Key Drivers for Next Generation Networking

Observations
- Growing data intensive usage models
- System repartitioning trends (CapEx)
- Increasing sensitivity energy consumption (network tax)

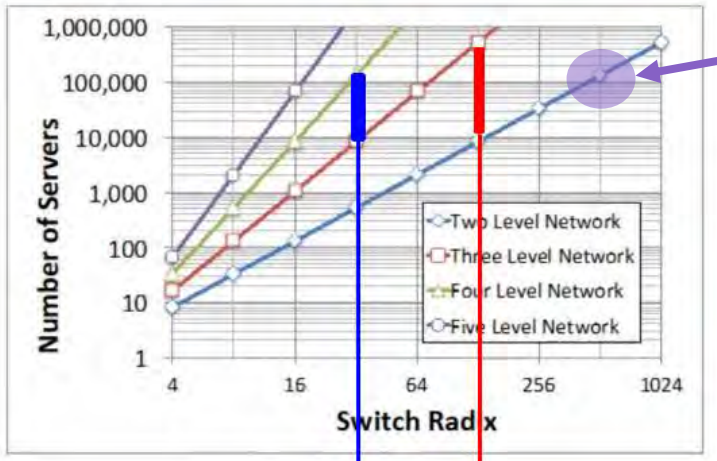
Implications
- Energy efficiency (low pJ / bit)
- Maximum use of available bandwidth
- Low latency
- Silicon bandwidth density (Gbps/mm^2)
- System and component cost
- Error handling/correction
- Link reliability/redundancy

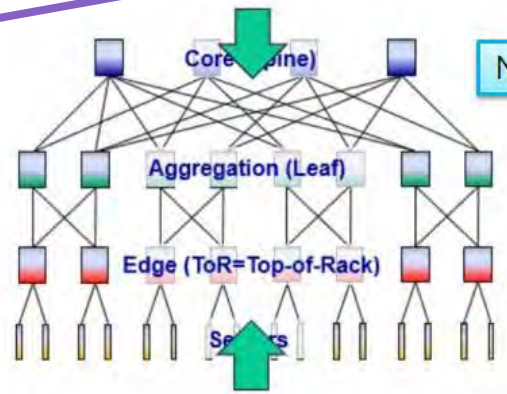


Networking Topologies Shifting to Meet Next Generation Demands

Focus on Flows

Radix = 512, 100K servers, 32MW DC, Two Tiers





Network flattening

-Source: Facebook

Switch Generation	Radix = 32	Radix = 64	Radix = 128
12.8T	400G	200G	100G
25.6T	800G	400G	200G
51.2T	1.6T	800G	400G

Optical interconnects

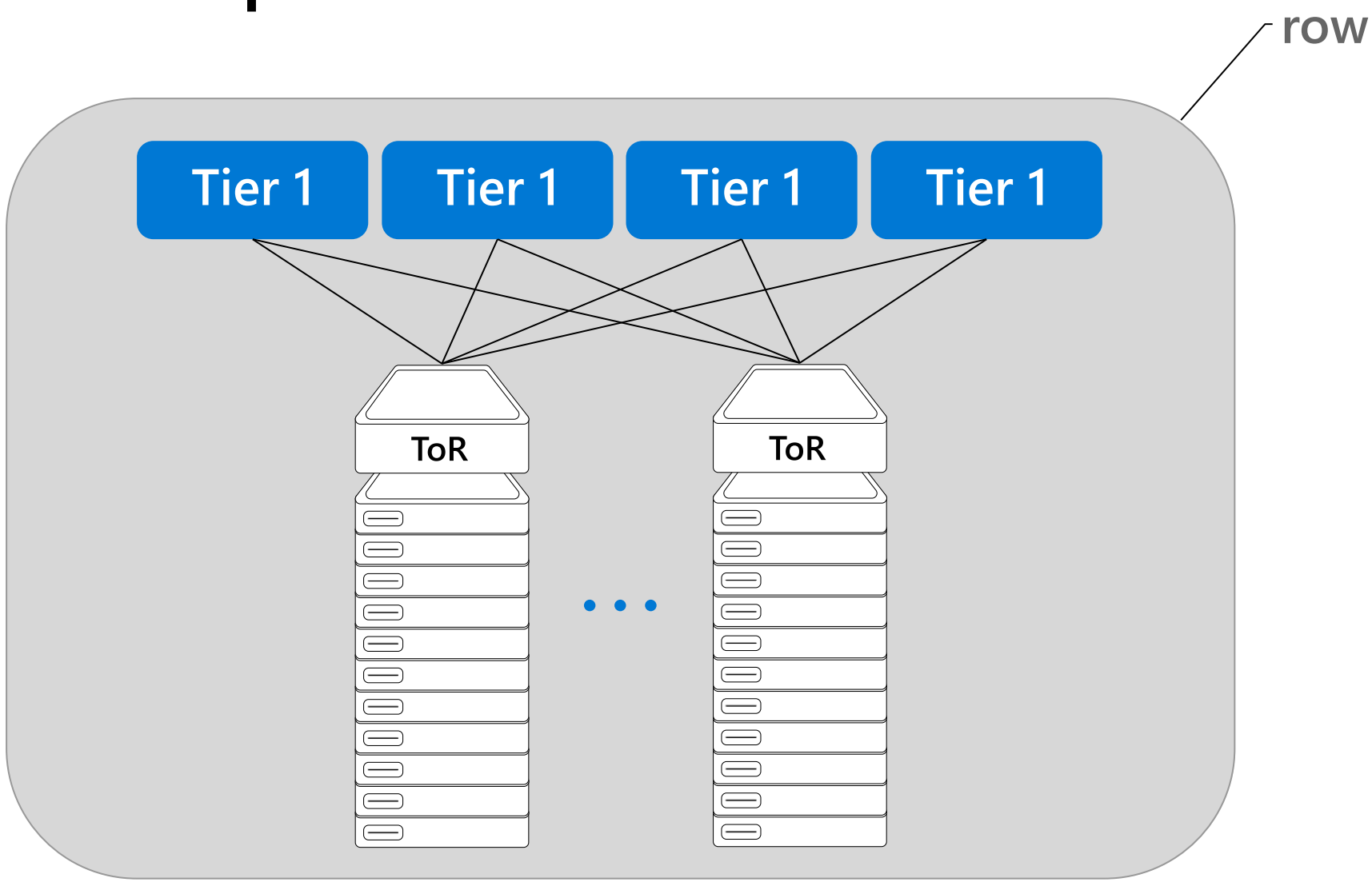
Inphi Copyright R. Nagarajan, Ilya Lyubomirsky, "Next-Gen Data Center Interconnects: The Race to 800G" 3 Inphi

- Fewer tiers = decreased latency, lower power
- CPU bandwidth ~ 1G/core
- Volume of servers vs. power grid

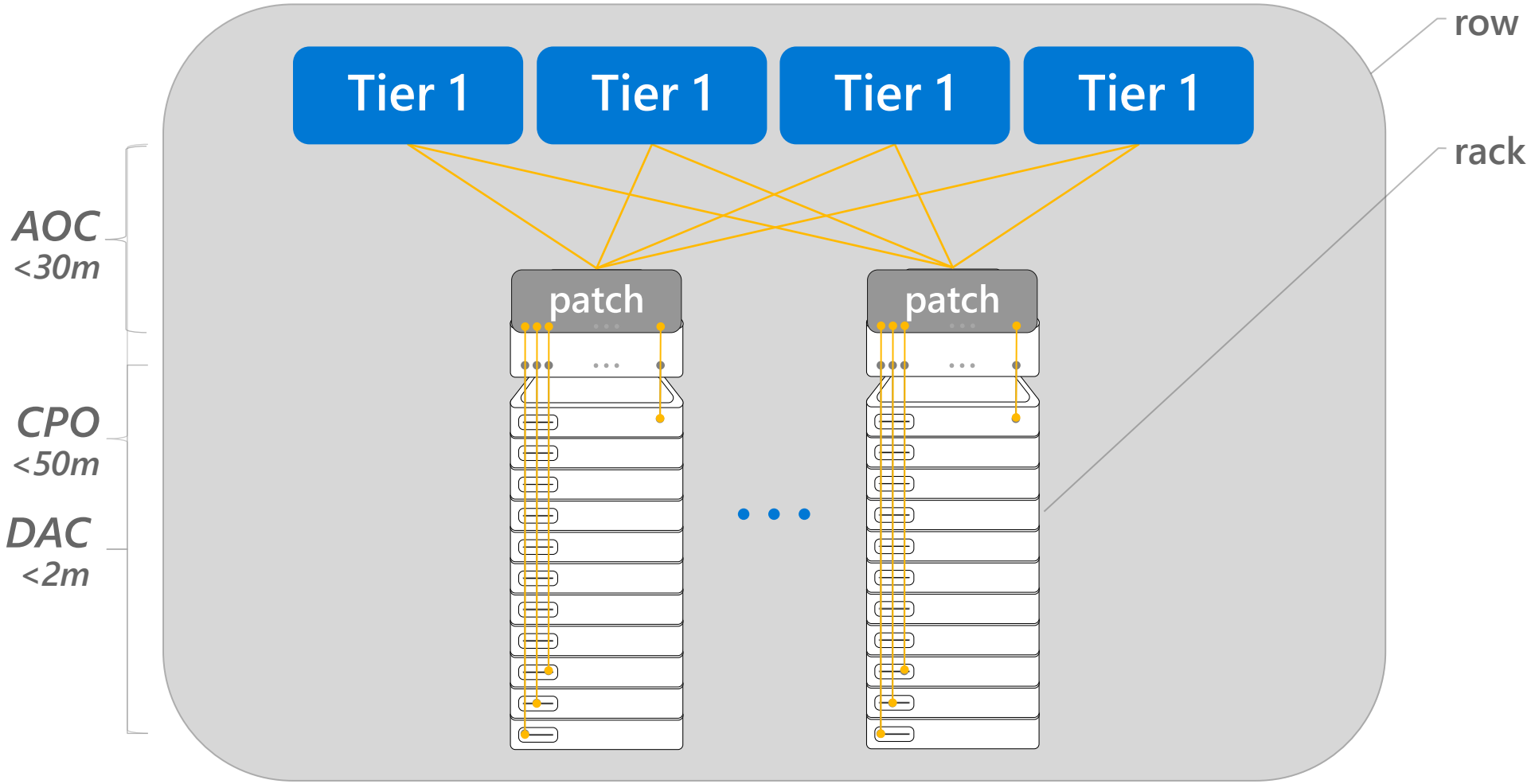
Topology Change Example

Is a ToR valuable?

- Stranded ports in high radix switches
- Additional latency & power
- Single point of failure
- Enables the use of DAC… for now



ToR Elimination



Summary

- ✓ Power is a growing challenge with increased data rates
- ✓ High radix switches plus optical interconnects enable new networking topologies
- ✓ Topology changes eliminate port stranding, reduce latency and improve energy efficiency
- ✓ New networking paradigm: Focus on the Flows

Co-Package Optics for Datacenters

Rob Stone, *Facebook*

ethernet alliance

Co-packaged Optics in the Datacenter

Ethernet Alliance Technology Exploration Forum
January 2021

Rob Stone
robstone@fb.com

FACEBOOK Infrastructure

Agenda

Problem Statement

New Workloads and Endpoints

CPO for Network Switching

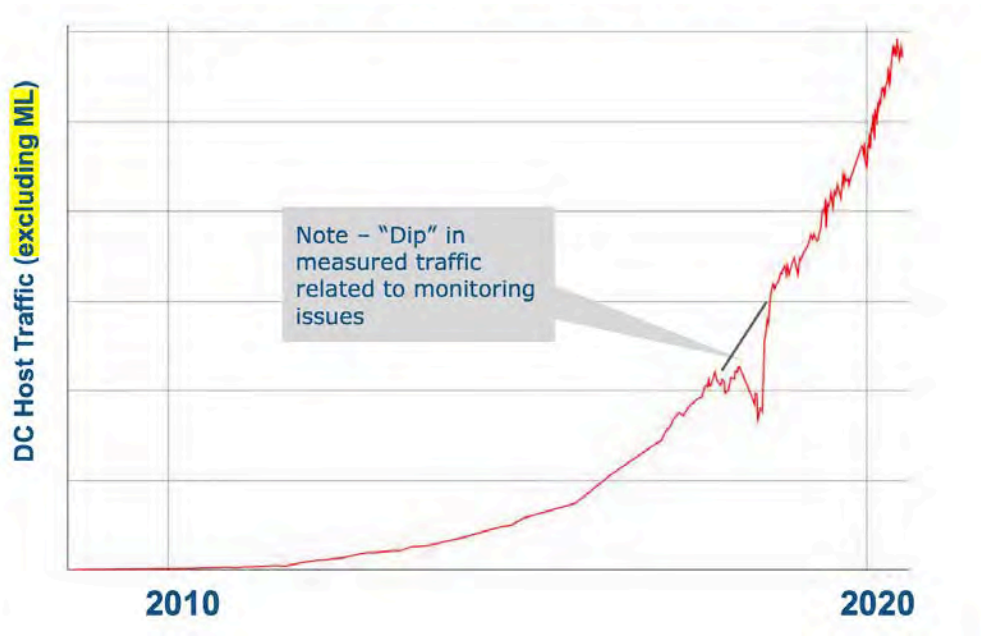
Network Timeline / Roadmap

Industry Initiatives

Conclusions

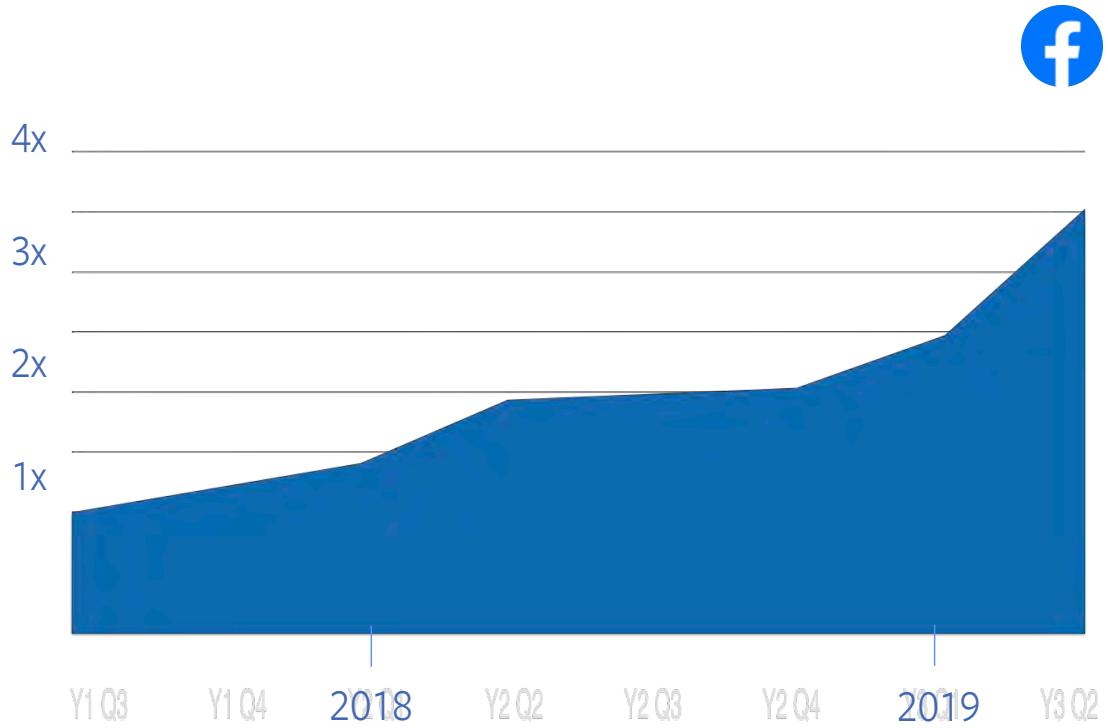
Continued Datacenter Bandwidth Growth

Data Center Traffic (ex ML)



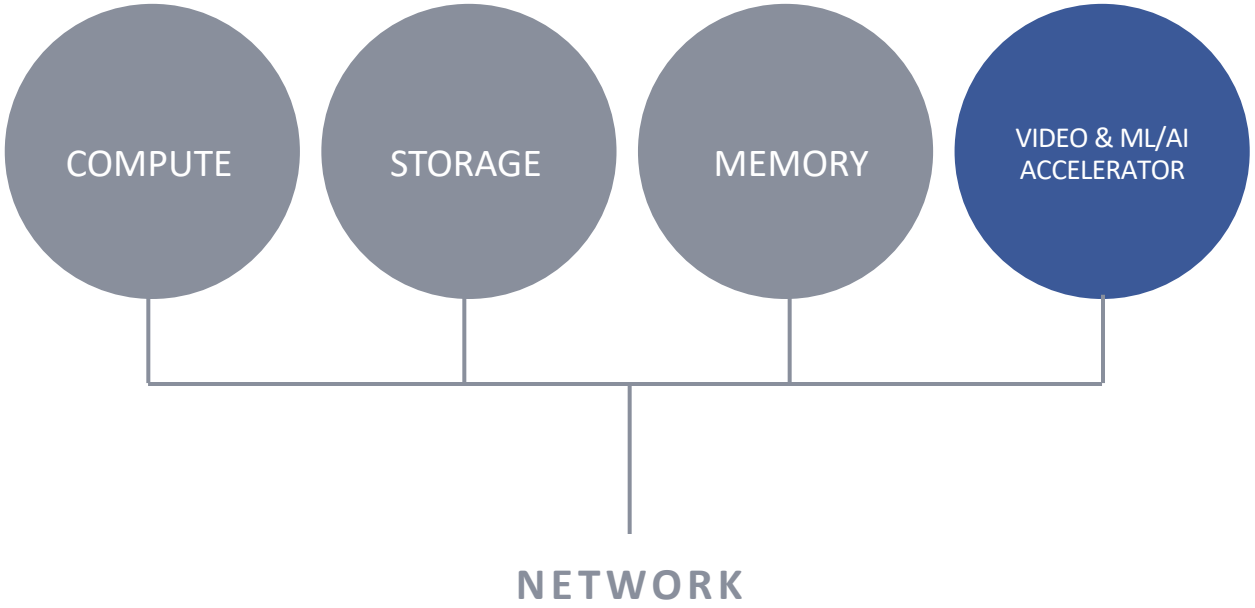
IEEE Beyond 400 Gb/s Ethernet CFI consensus presentation (Cedric Lam, Google)

AI Inference



Vijay Rao OCP Keynote 2019

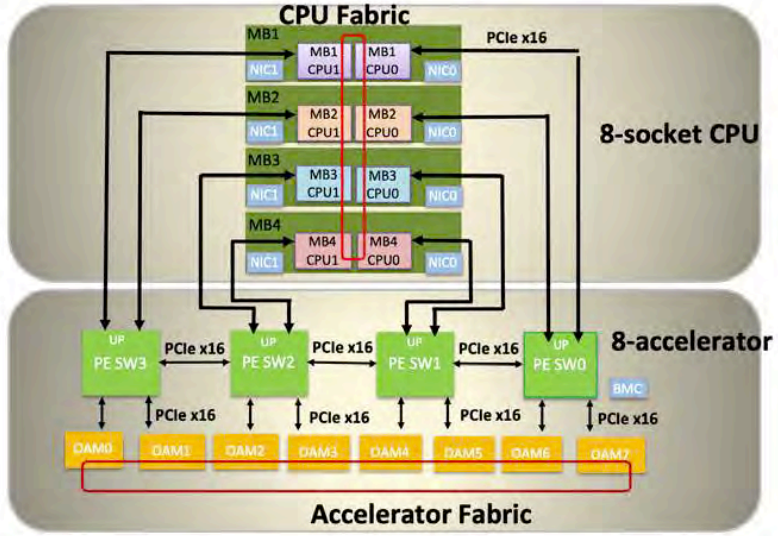
Changing Workloads



New Accelerators Driving Higher Network Demand

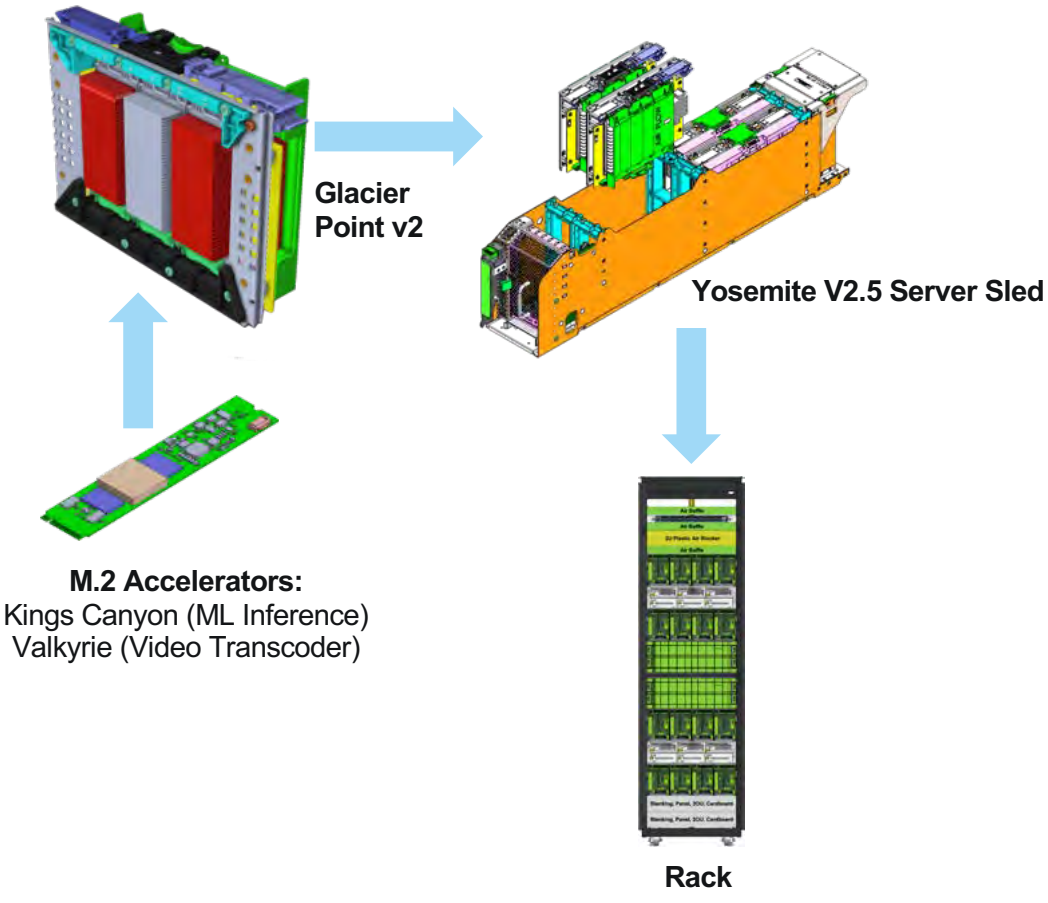
ML Training: Zion

8 x 100GbE scale-out network per node



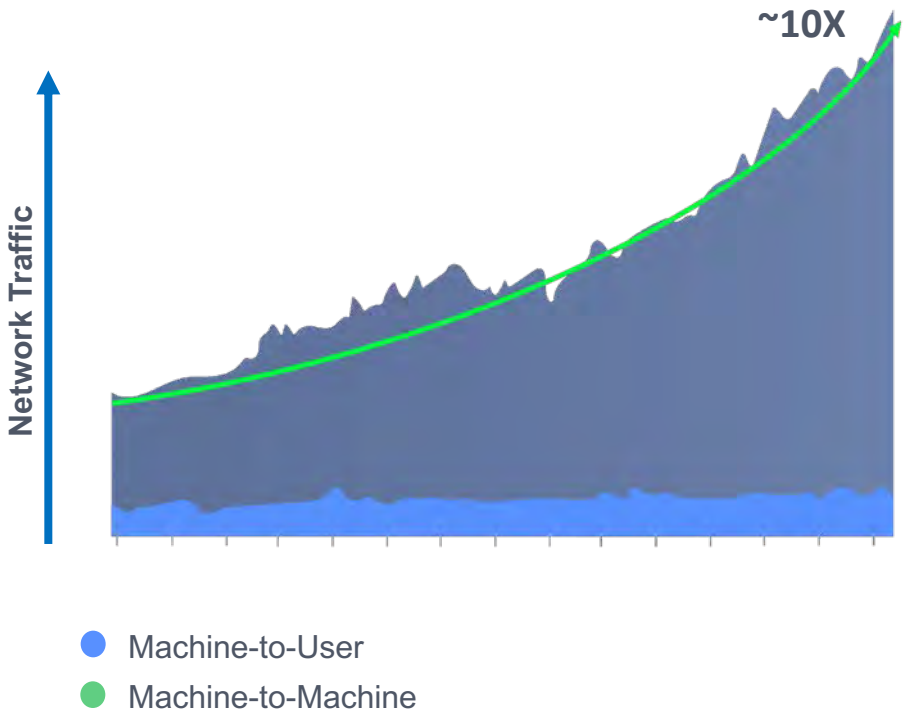
Compute Accelerator: Yosemite + Glacier Point

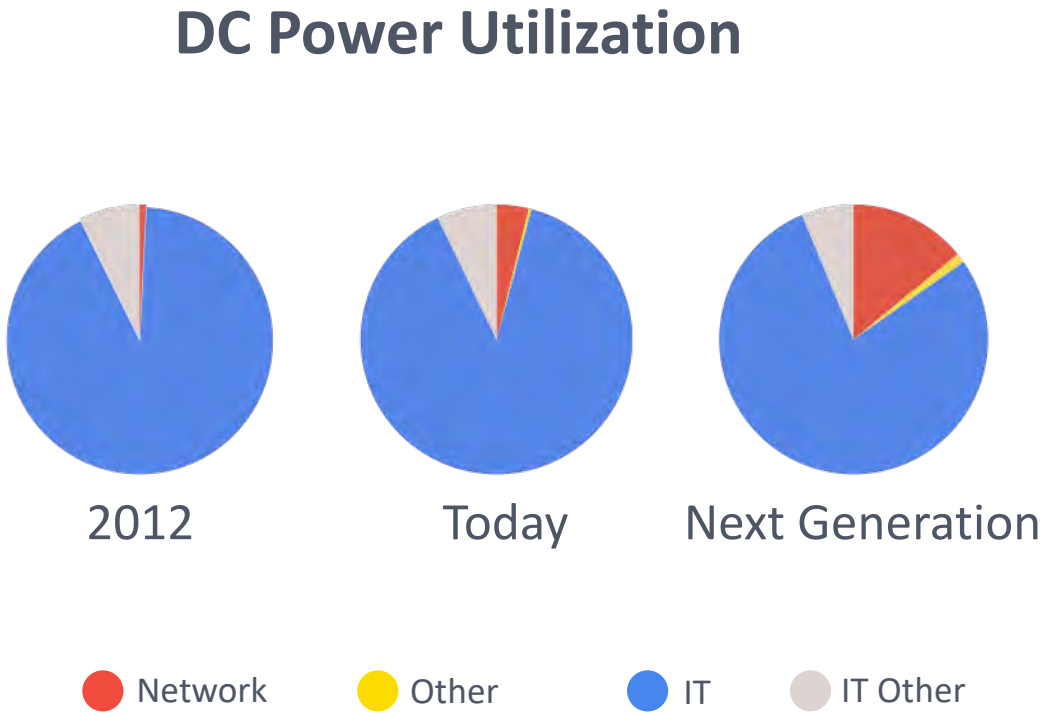
100GbE per Server



https://engineering.fb.com/2019/03/14/data-center-engineering/accelerating-infrastructure/

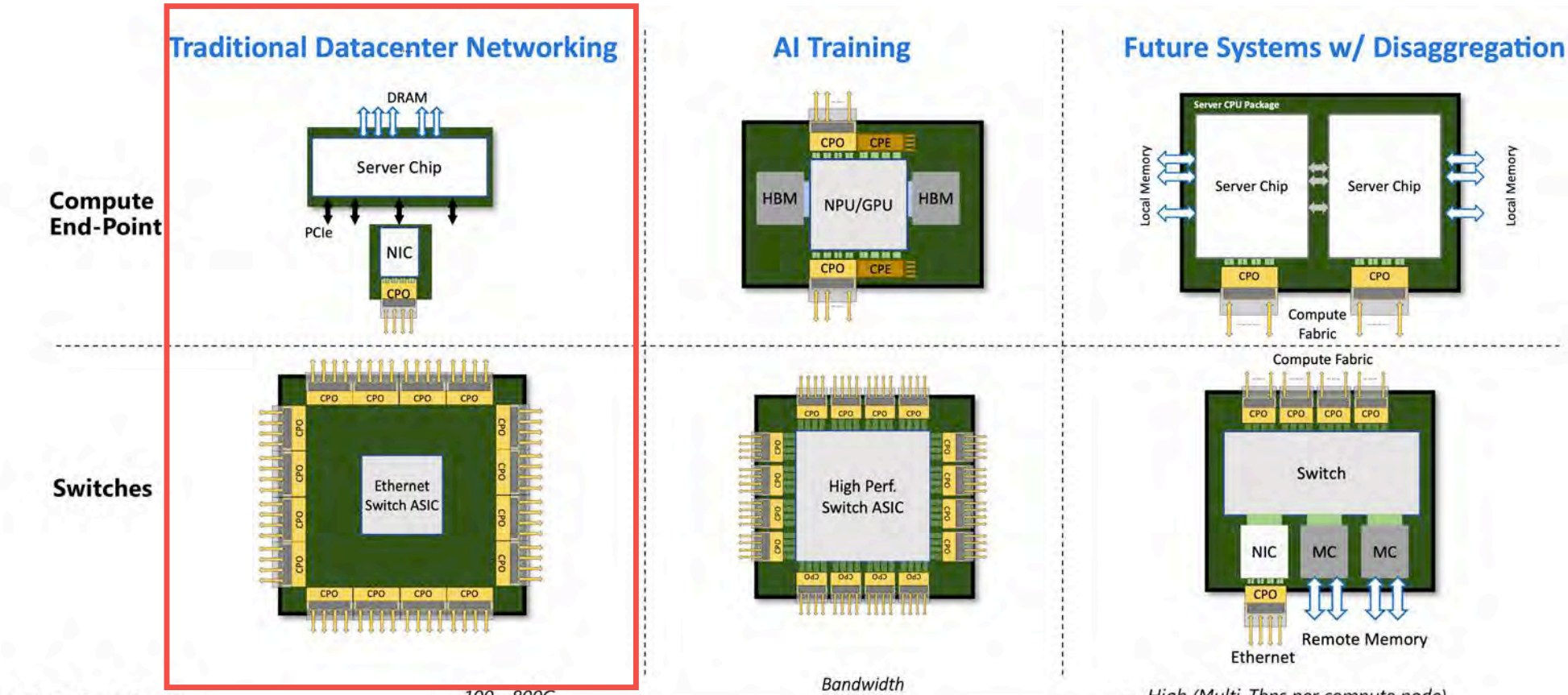
Growing Network Power Allocation





- Networking is consuming a higher proportion of the data center power budget

Application Areas for CPO

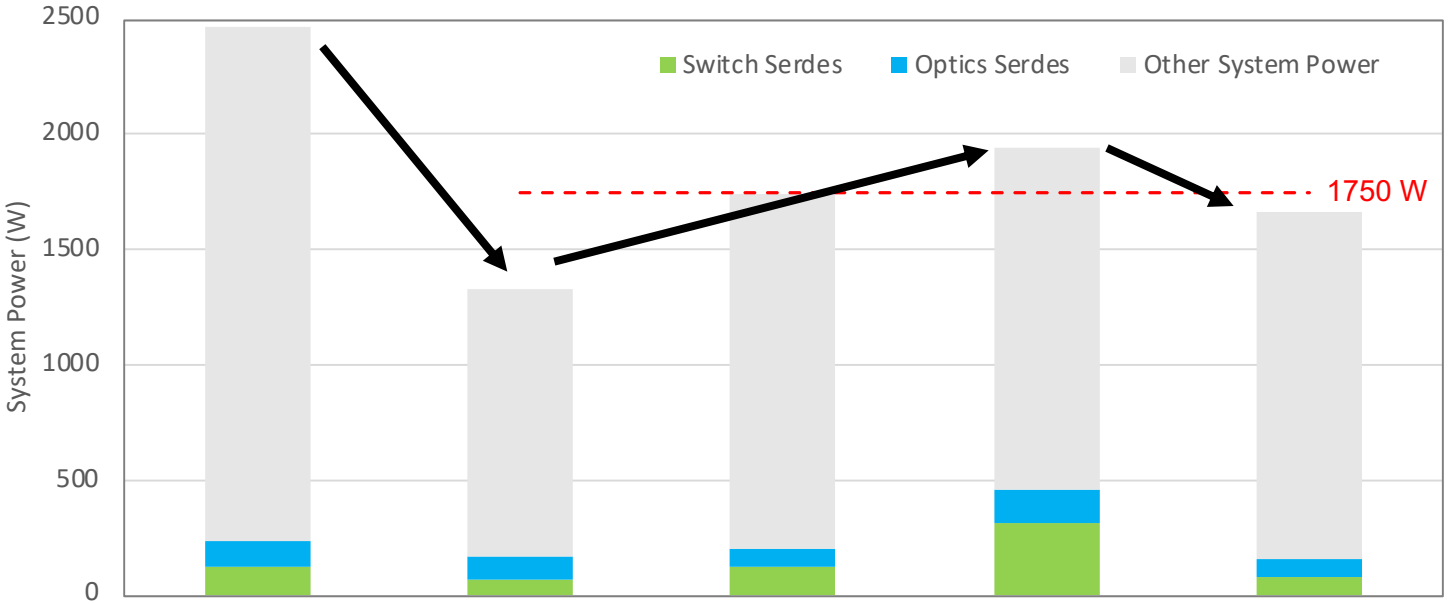


Anticipate a wide range of applications for CPO with alternative CPO designs

Microsoft

Ram Huggahalli, Microsoft (used with permission) OIF CPO Webinar November 2020

Power Reduction Motivation for CPO



Bandwidth	**12.8T**	**12.8T**	**25.6T**	**51.2T**	**51.2T**
Codename	Backpack	Minipack	Minipack 2	Next-Gen	CPO Switch
System Architecture	12 Chip, LR	Single Chip, LR	Single Chip, LR	Single Chip LR	Single Chip, XSR
Optics Form Factor	128 x 100G QSFP-28	128 x 100G QSFP-28	128 x 200G QSFP-56	128 x 400G FPP	CPO

Front-Panel Pluggable Optics

Co-Packaged Optics

Backpack

Minipack

- Co packaged optics (CPO) provides the next big step in power reduction

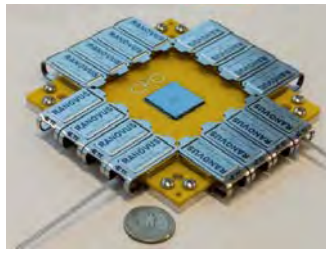
Industry Initiatives

- CPO Demonstrations / Concepts

Cisco / Luxtera Mechanical Concept
OCP Summit 2018

Intel CPO Live Demo
OFC 2020

Rockley Photonics Mechanical Concept
OFC 2020



Ranovus Mechanical Concept
OFC 2020

- Standards Development:
  - Microsoft / Facebook JDF

    www.copackagedoptics.com
  - OIF CPO Framework Project
  - COBO

Data center locations

Luleå, Sweden

Odense, Denmark

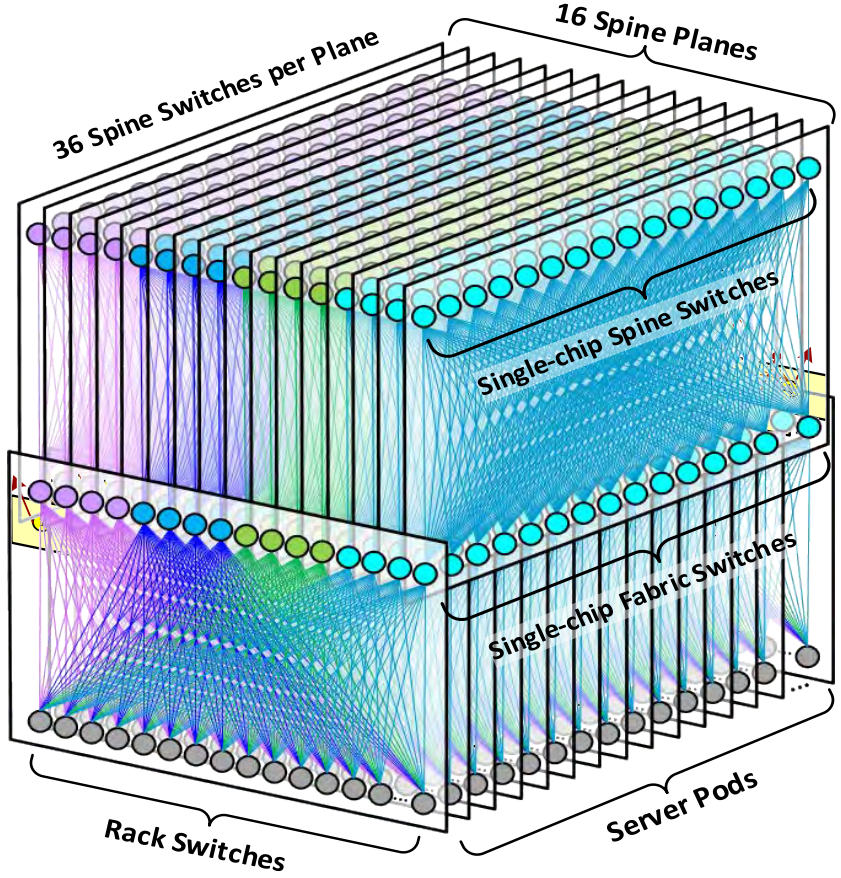
Clonee, Ireland

Singapore

New construction

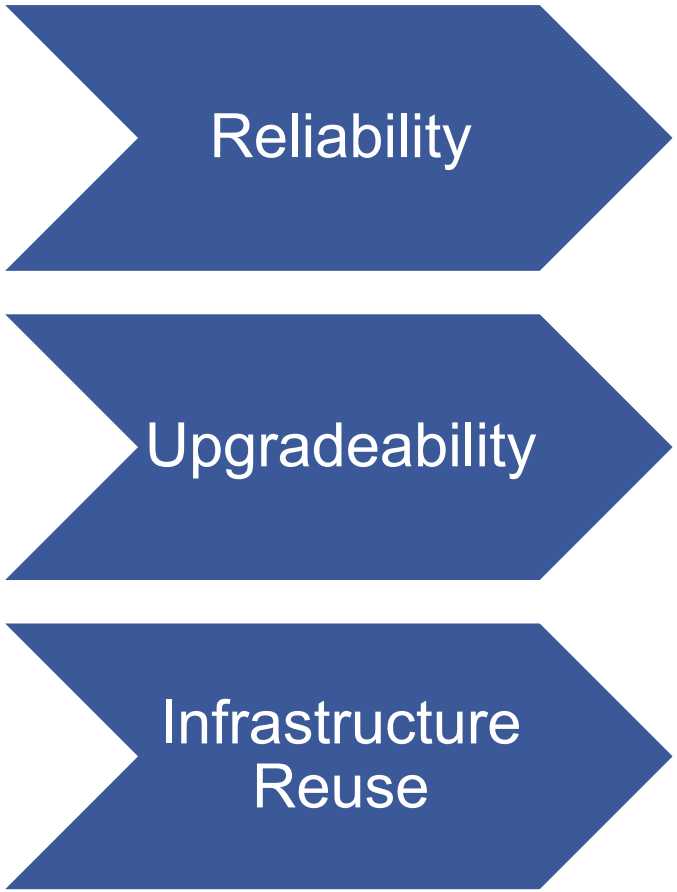
Serving traffic

Datacenter Scale



- Up to 16 plane architecture:
  - 16 x duplex fiber pairs / rack
- Connects compute, storage and switches
- Server pod: 48 racks

Networking Operational Implications at Scale



- Challenge for CPO vs Front Panel Pluggables (FPPs)
  - Serviceable External Laser Source, spare channels possible options
- Security of Supply – broad supplier ecosystem (→ standardization)

- Minimize disruption to network during upgrades
  (backwards compatible modes 200GE 4 x 50G → 100GE 4 x 25G)
  Forwards compatibility desirable, but harder!

- Compatibility with existing network topology
  (scale switch bandwidth while preserving radix)
- Maintain networking power footprint

Intra Data Center Network Evolution

FB Initial Deployment Year	Port Speed (Gb/s)	Electrical Lane Speed (Gb/s)	Switch Silicon Bandwidth (Tb/s)	Switch System Bandwidth (Tb/s)	Switch Hardware	Optical PMD	Optical Lane Speed (Gb/s/λ)	Optical Module Type		
								Front Panel Pluggable Optics	On Board Optics	Co-packaged Optics
2015	40	10	1.28 (x 12)	5.12	Six Pack	40GBASE-LR	10	QSFP+	-	-
2016	100	25	3.2 (x 12)	12.8	Backpack	100G-CWDM4 (OCP)	25	QSFP-28	-	-
2018	100	50	12.8	12.8	Minipack	100G-CWDM4 (OCP)	25	QSFP-28	Mini-Photon	-
2021	200	50	25.6	25.6	Next Gen	200G-FR4 (OCP)	50	QSFP-56	Next Gen OBO	-
2023	400	100	51.2	51.2	Next Gen	400G-FR4	100	TBD	Next Gen OBO	CPO Gen 1
…	800	200	102.4	102.4		800G-FR4	200	-	-	CPO Gen 2

Standards Development Organization

OIF OCP IEEE 802.3 SNIA / SFF

Conclusions

- Higher network demand from new applications
- Co-packaged optics is important to enable offer power and density savings to support next generation system scaling
- Standards supported CPO ecosystem is required for wide adoption

Ongoing Priorities: scale network bandwidth power and cost efficiently

Team Acknowledgements!




















Thank you!

OIF Considerations for Beyond 400ZR

Tad Hofmeister, *Google*

ethernet alliance

OIF Considerations for Beyond 400ZR

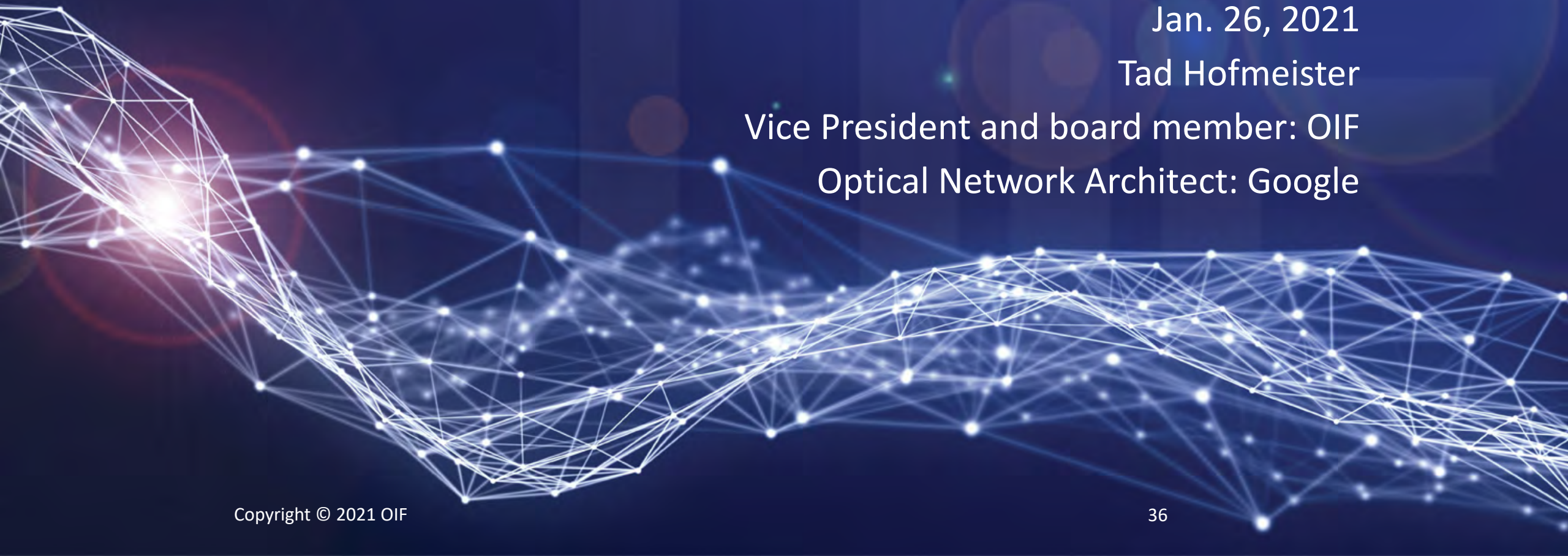
Ethernet Alliance Technology Exchange Forum

Jan. 26, 2021

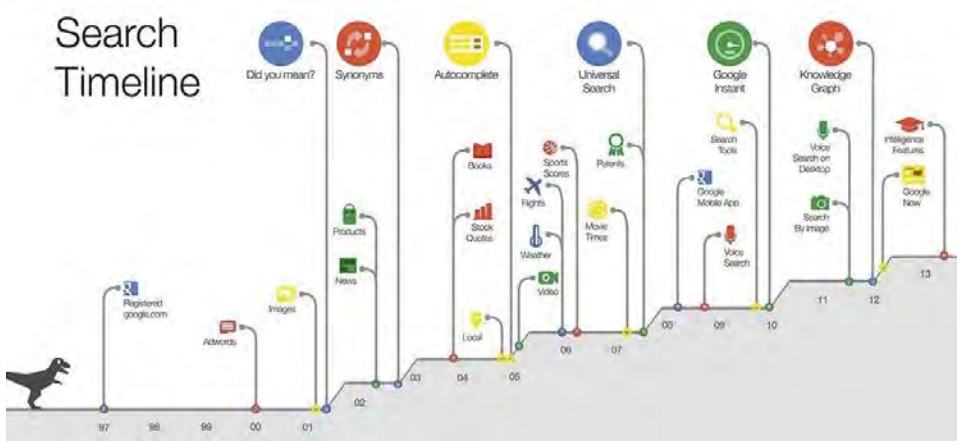
Tad Hofmeister

Vice President and board member: OIF

Optical Network Architect: Google

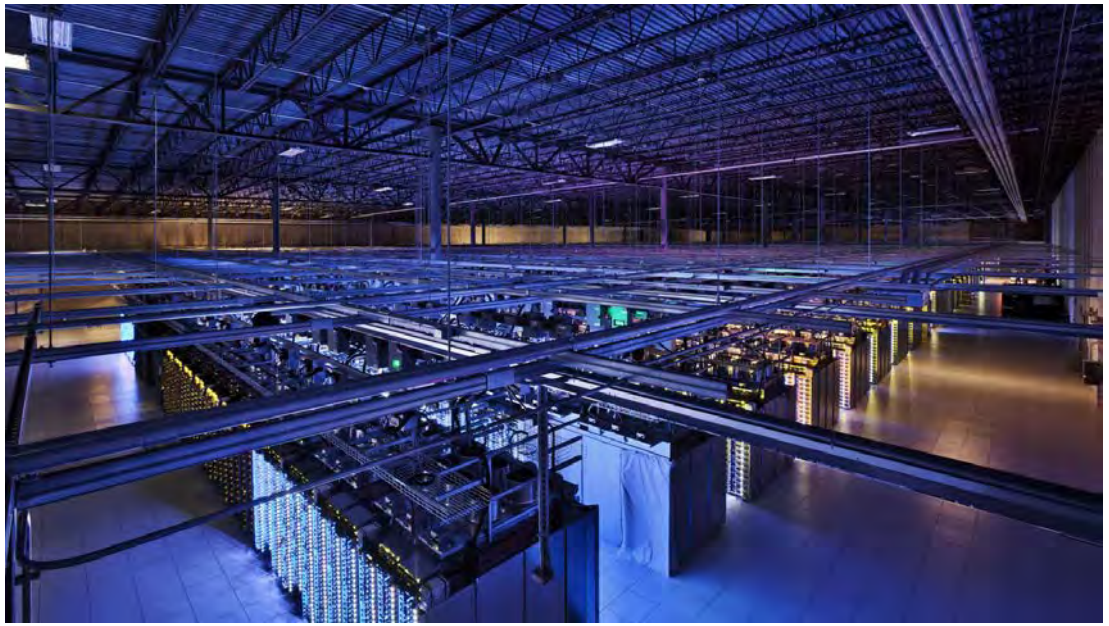


Google Services Growth

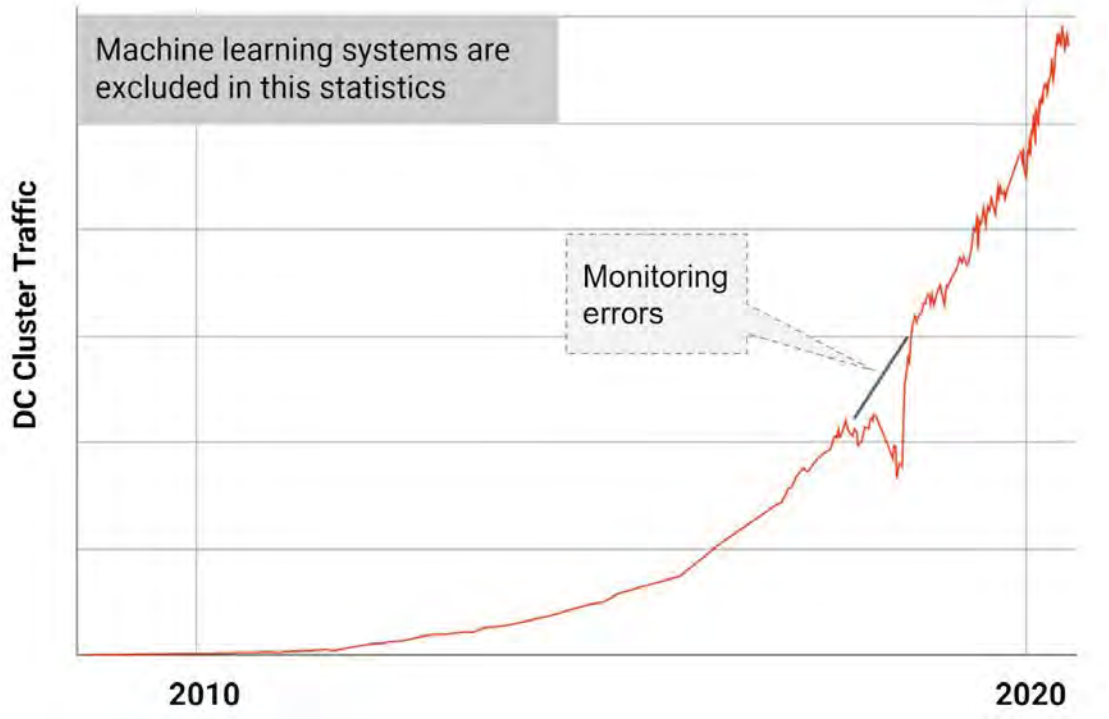


- 100 Billion searches per month
- 9+ apps with 1 Billion users
- 1 Billion individual IPs connected to by Google Cloud Platform (GCP)

Google Data Center Cluster Traffic Growth



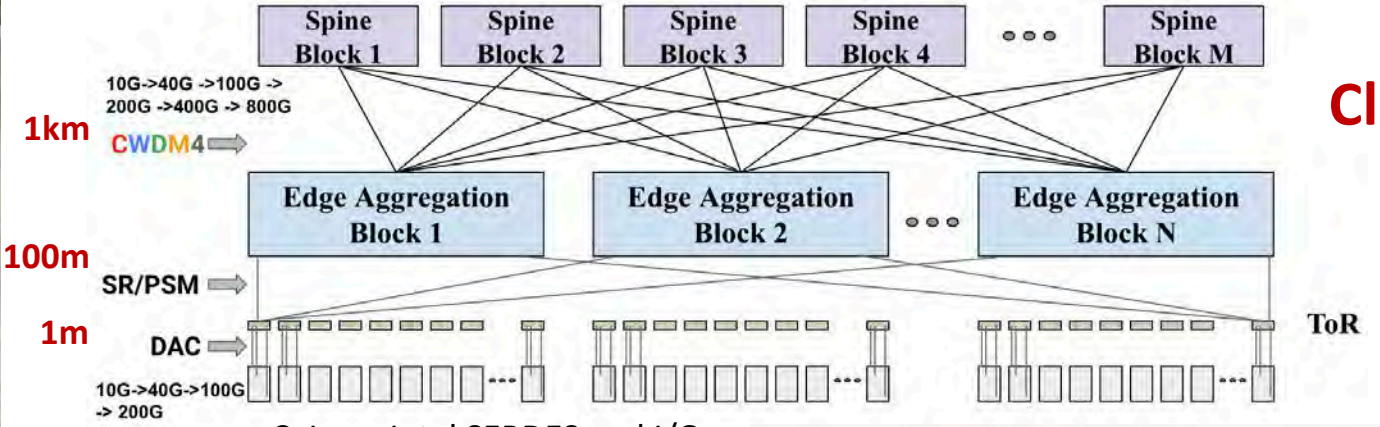
www.google.com/about/datacenters



C. Lam, ECOC Workshop Dec. 2020

Node Interconnect Distances

www.google.com/maps



Cluster

C. Lam, Intel SERDES and I/O Summit Dec. 2020

Campus
10km

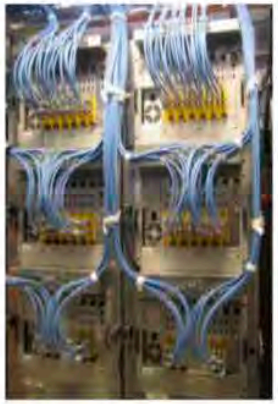
WAN
10000km

cloud.google.com/about/locations#regions

Google DC Network Evolution

5000x bandwidth increase!

Firehose (2006)



Watchtower (2008)

Saturn (2009)

10G SFP

Jupiter (2012)

1.3 Petabits/sec

40G QSFP

Jupiter 3.0 (2019)

6.5 Petabits/sec

400G OSFP

800G OSFP

Reach flexibility enabled by front panel pluggable transceivers & direct attached copper (DAC)

Motivation for 800G Pluggables

- Front panel pluggable transceivers:
  - Enables reach flexibility per port
    - Minimize cost and power for different applications
    - DAC (meters), SR (10s m), DR (<500m), FR (<2km), LR (<10km), ZR (<120km)
  - Pay as you grow, only populate ports as needed
- 800G per transceiver capacity:
  - Maximize front panel density for improved system density
  - Lower cost per bit of transceiver
    - Higher bit rate per laser/PD
    - Fewer transceivers per Tb/s

100Gb/s Electrical IO ➔ 800G Pluggables

OSFP

QSFP-DD800

OIF

800G Coherent

IEEE

802.3 Beyond 400Gb/s Ethernet Study Group

Ethernet Technology Consortium

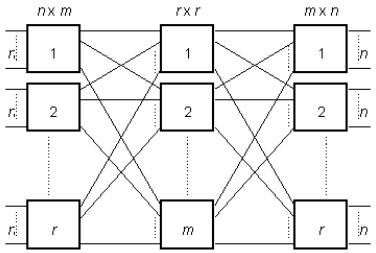
800GPluggable

MULTI-SOURCE AGREEMENT

800G Capacity is Not Limited to 800G-Ethernet

- 8x 100GE:
  - Enables backward compatibility with legacy devices
  - Increases capacity/port-count in the same rack-space
  - Enables larger fan-out to support larger capacity CLOS fabrics
  - DAC, 800G-SR8, 800G-DR8…
- Similarly, 2x 400GE:
  - DAC, 2x400G-FR4…

CLOS Fabric



en.wikipedia.org

OIF 800G Coherent Project

- Building on the success of the OIF's 400ZR coherent implementation agreement (IA)
- OIF Network Operator members desire similar functionality for 800G
- Prior to formal project start, the OIF NetOp WG:
  - Invited member and non-member network operators to 2 round table sessions at ECOC2019 and Microsoft OIF Day in Oct. 2019
  - Sent a beyond 400G survey to 24 network operators, receiving responses from 17 (7 hyperscalers and 10 traditional carriers) in early 2020
  - Confirmed strong demand for 800G coherent IA and enabled OIF to refine the scope of the project

800G Coherent Project Scope

- Single wavelength line interface with 2 applications:
  - Amplified, single span, DWDM links with target maximum of 80km – 120km
  - Un-amplified, fixed wavelength links with target maximum of 2km – 10km
- Support for multiplexing of independent clients:
  - 8x 100GE
  - or 2x 400GE
- IA will not be tied to a specific implementation form factor, but will include defined electrical power consumption classes, at least one of which will align with QSFP-DD800 and OSFP transceiver MSAs

Summary

- 800G Pluggable transceivers will be widely deployed
- 800G capacity for:
  - 8x 100GE
  - 2x 400GE
  - 800GE
- OIF 800G coherent project will enable interop for:
  - DWDM for campus, DCI, metro reaches
  - Single wavelength for campus reaches

Join Us Tomorrow, Wednesday, Jan 27th

Panel discussion – Beyond 100G Electrical

Moderator – Chris Lyon, *Ethernet Alliance President and Amphenol*

Panelists –

Cathy Liu, *Broadcom – "Electrical Interfaces beyond 100G"*

Nathan Tracy, *Ethernet Alliance Board Member and TE Connectivity – "Electrical Future Work"*

Ramin Farjad, *Marvell – "224G Dual Duplex Signaling over Single Lane Copper Links"*

Dave Stauffer, *OIF Representative – "OIF Common Electrical I/O 224G Project Overview"*

For our TEF 2021 on-demand content go to www.ethernetalliance.org

If you have any questions or comments, please email **admin@ethernetalliance.org**

@ethernetalliance

@EthernetAllianc

Ethernet Alliance