



Open. Together.



OCP
REGIONAL
SUMMIT

Data Center Network Next Generation

Hans-Juergen Schmidtke

Facebook, Director of Engineering



Open. Together.



FACEBOOK



INSTAGRAM



MESSANGER



WHATSAPP

WHATSAPP

65

BILLION

Messages sent
per day

>2

BILLION

Minutes of voice and video
calls per day

MESSENGER

81

BILLION

Messages sent between
businesses and
customers each month

VIDEO

3.5

BILLION

FB Live Broadcasts

>10

MILLION

FB Live Broadcasts on
New Year's Eve

PAPILLION, NE



LOS LUNAS, NM



PRINEVILLE, OR



FOREST CITY, NC



FORT WORTH, TX



NEWTON COUNTY, GA



NEW ALBANY, OH



ODENSE, DENMARK



LULEÅ, SWEDEN



ALTOONA, IA



CLONEE, IRELAND



HENRICO, VA



EAGLE MOUNTAIN, UT

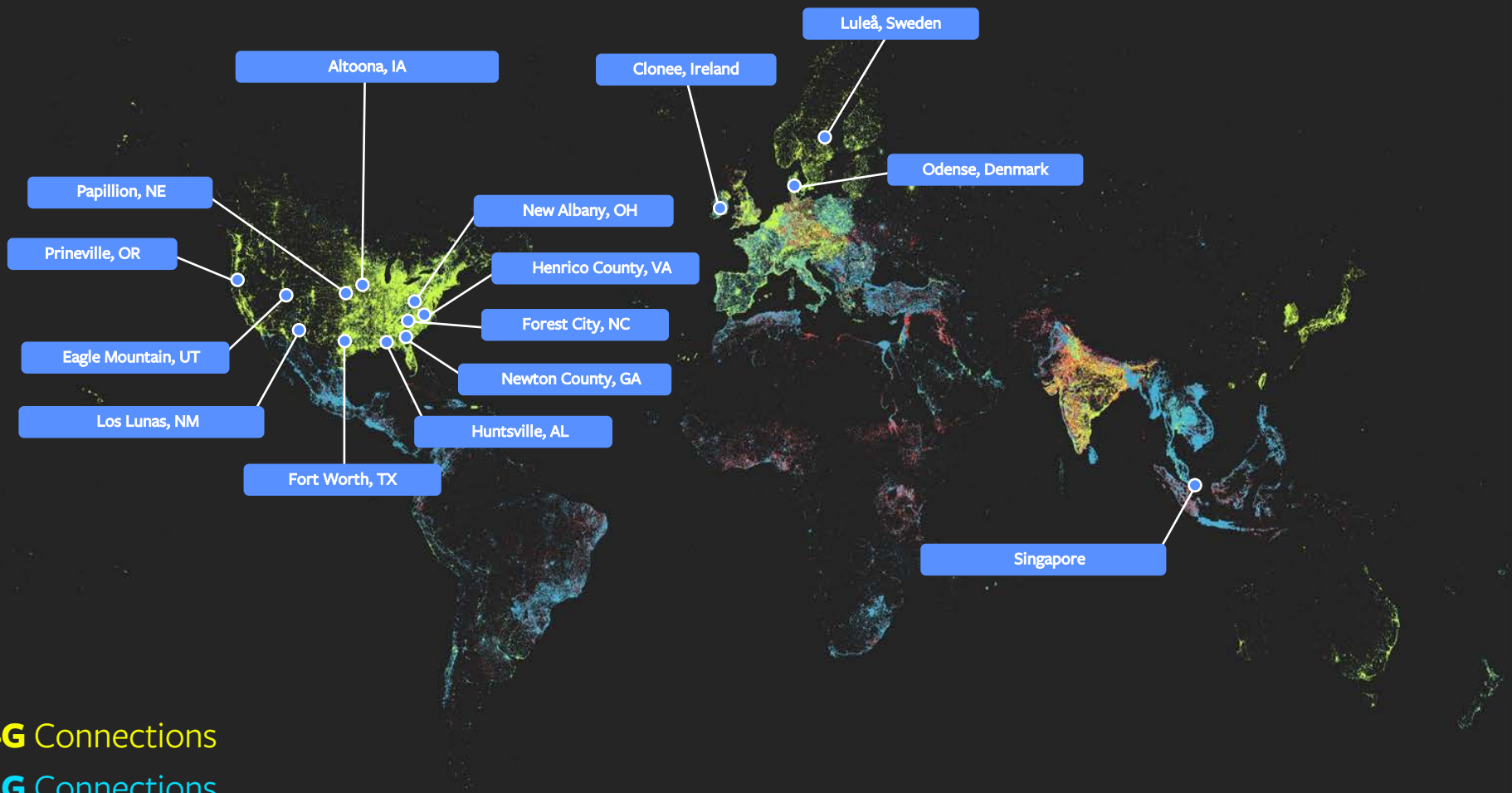


HUNTSVILLE, AL



SINGAPORE





4G Connections

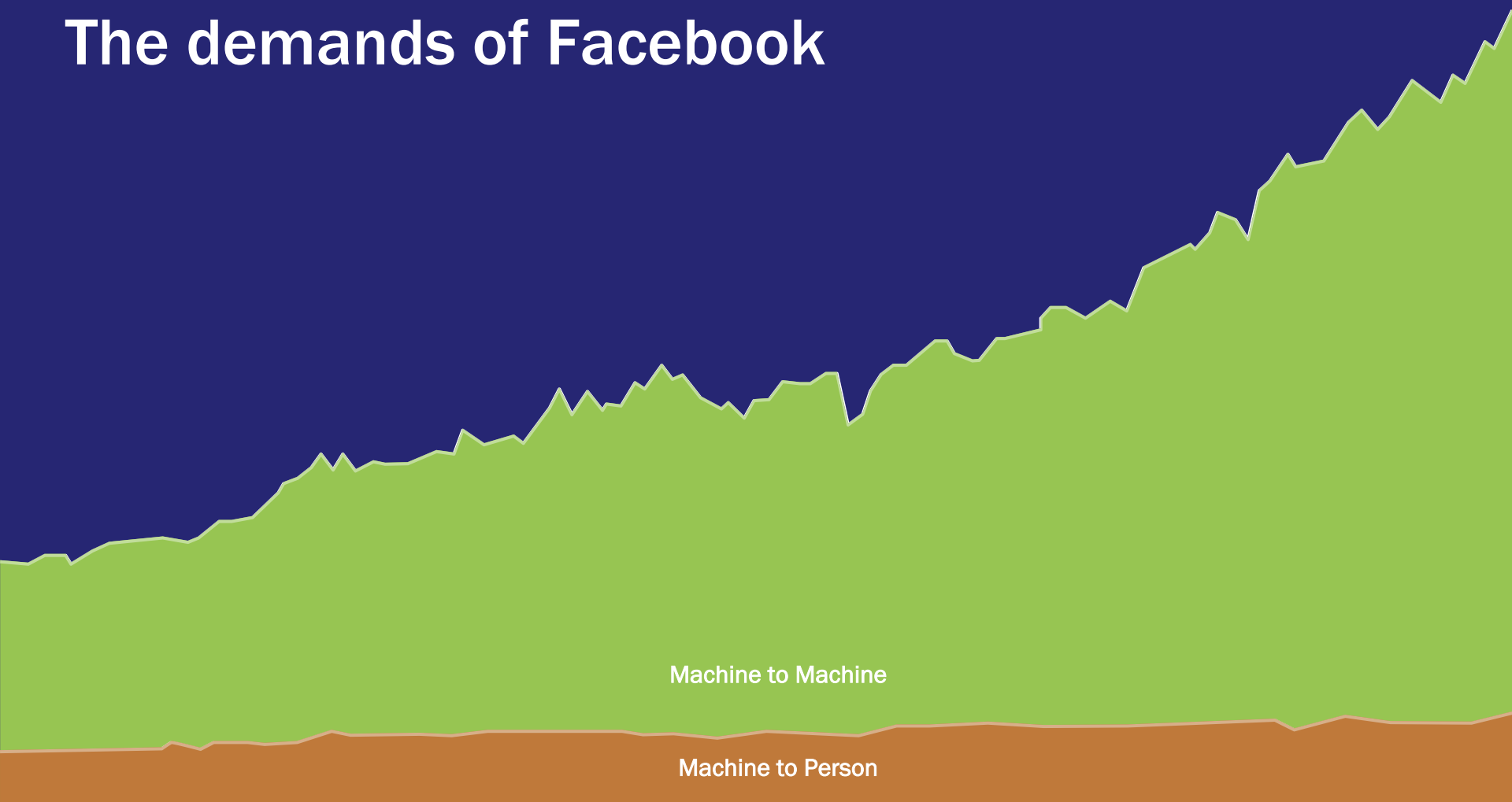
3G Connections

2G Connections

The background is a solid dark blue. On the left side, there is an abstract graphic consisting of numerous thin, flowing yellow lines that curve and sweep across the frame, creating a sense of movement and depth. These lines are more concentrated on the left and fade out towards the right.

The Challenges

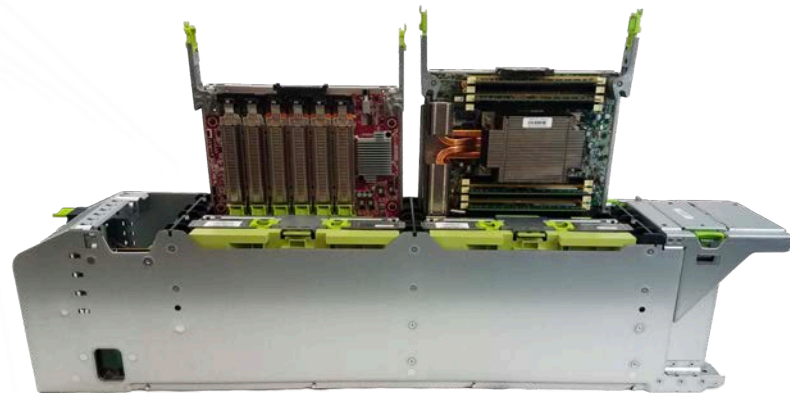
The demands of Facebook



Increasing Workloads



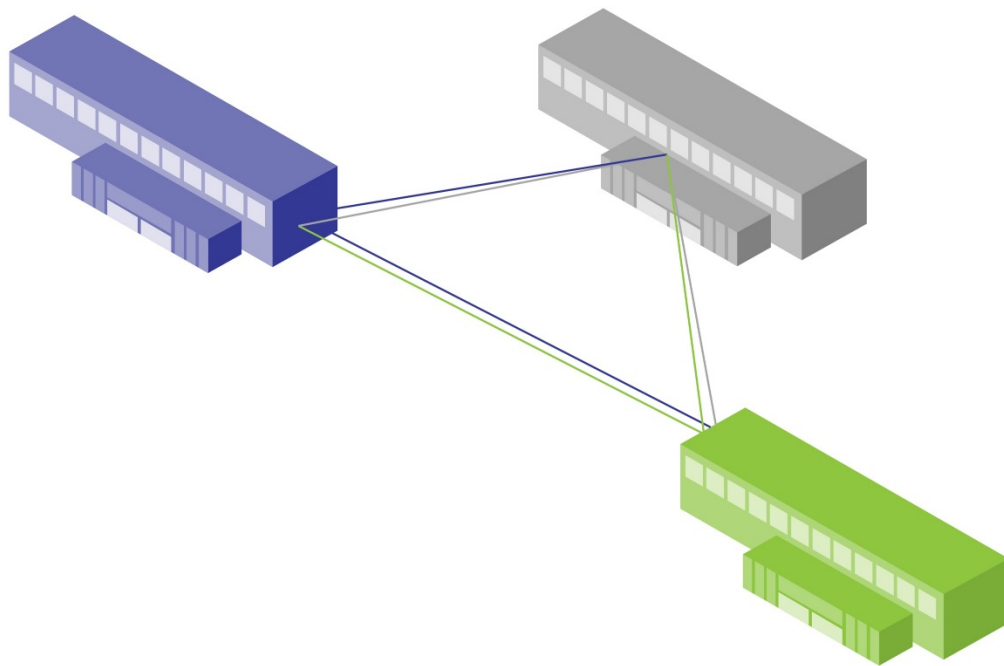
Zion



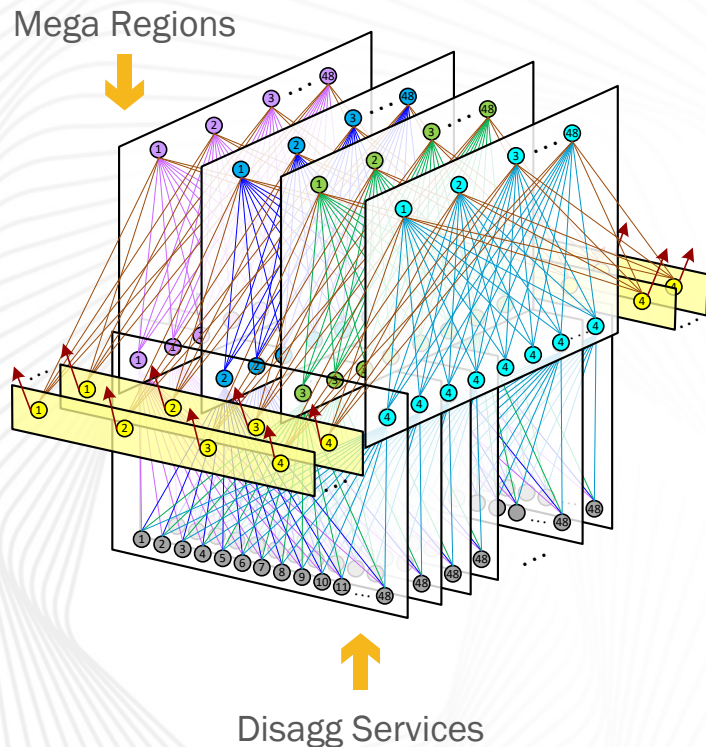
Yosemite + Glacier Point + Accelerator
Modules

Regions Hitting Limits...

- **Fabric Aggregator (FA):**
inter-building fabric of fabrics
- Up to **3 large buildings** (fabrics)
- **100Ts level** of regional uplink capacity per fabric (max)
- **Power is limited**



Growing Pressures on the Fabric...



Expanding **Mega Regions**
(5-6 buildings) = accelerated
fabric-to-fabric East-West demand

Compute-Storage and AI disaggregation
requires **Terabit capacity per Rack**

Both require larger fabric Spine
capacity (by **2-4x**) ...

Dilemma of scaling up and wide

Prediction ↓

Year	Throughput	SERDES	Ideal optical speed	Implemented @scale
2016	3.2T	25G	100G	100G
2018	6.4T	25G	200G	100G
Estimates				
2019	12.8T	50G	400G	100G
2021	25T	50G	800G	200G
2023	50T	100G	1.6T	400G
2025	100T	100G/200G	3.2T	400G?

Optics Running 1-2
Generations Behind

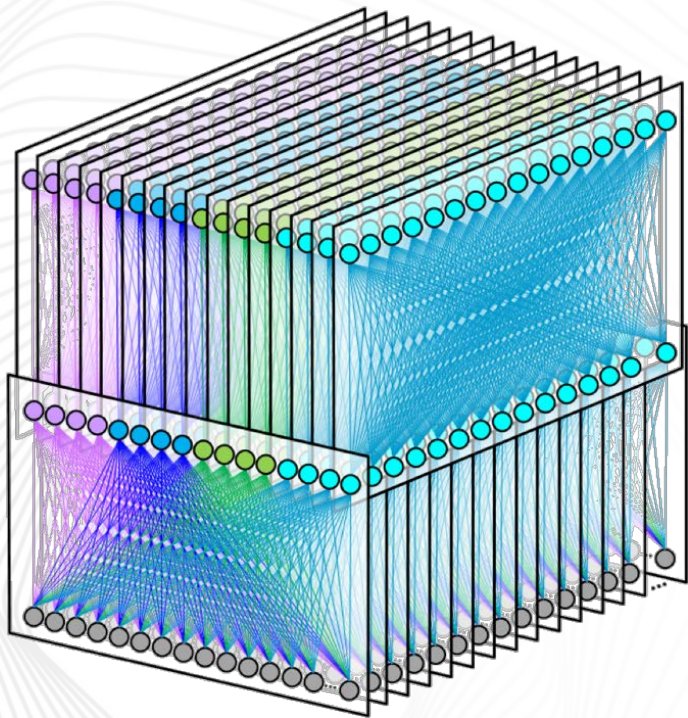
Power
Limit

- Challenge of optical manufacturing
- Challenge of SERDES scale
- Challenge of Chip edge I/O
- Challenge of Power

The background is a solid dark blue. On the left side, there is an abstract graphic consisting of numerous thin, wavy, yellow lines that flow from the top left towards the bottom right, creating a sense of movement and depth. The lines are more densely packed in some areas and more spread out in others, giving it a three-dimensional, tunnel-like appearance.

The Network: F16 and HGRID

F16 — Facebook's New DC Topology



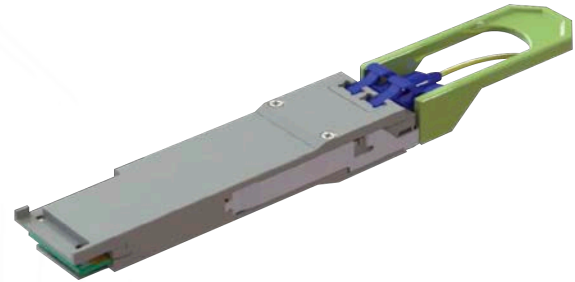
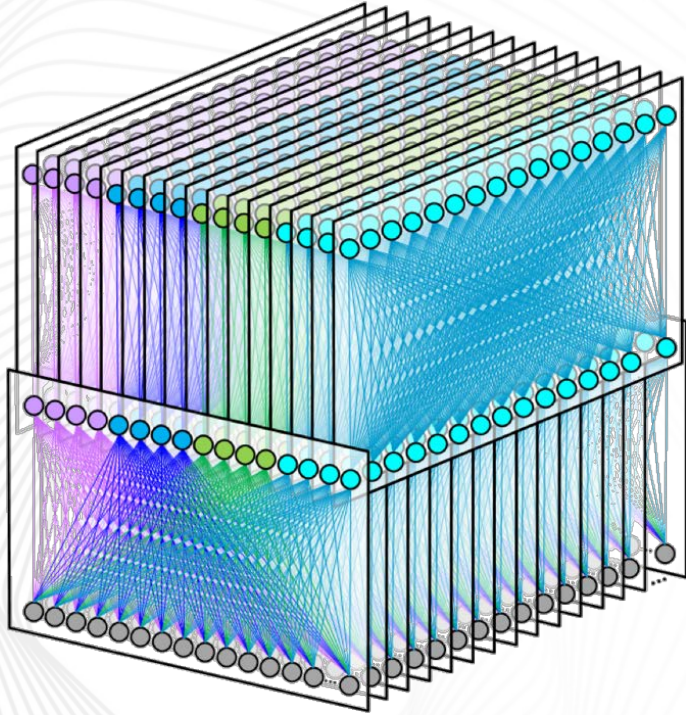
16-plane architecture

6-16x spine capacity on day 1

1.6T raw capacity per rack

Fewer chips = better power & space

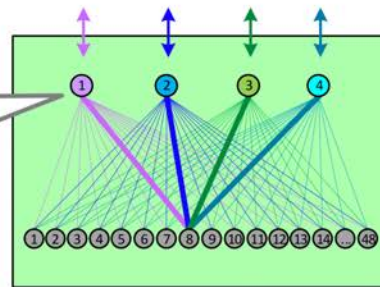
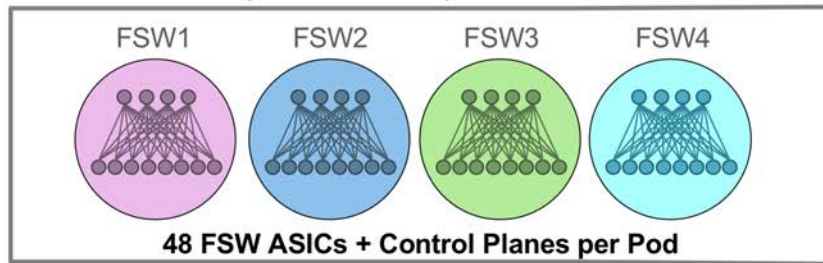
400G Speeds with 100G Optics



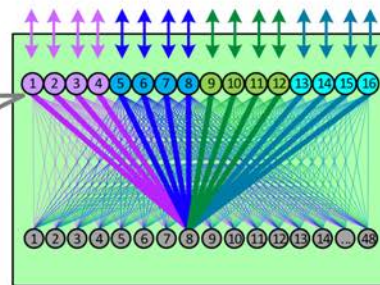
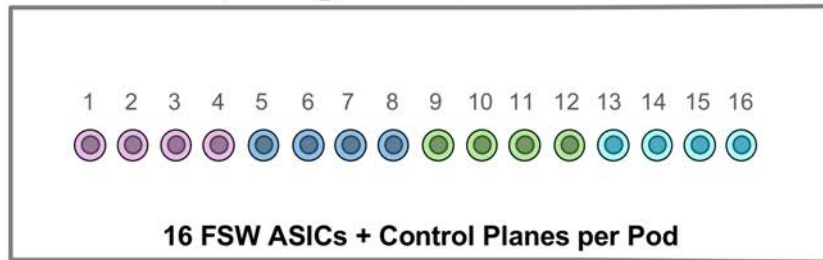
100G
CWDM4-OCF

400G Speeds with 100G Optics

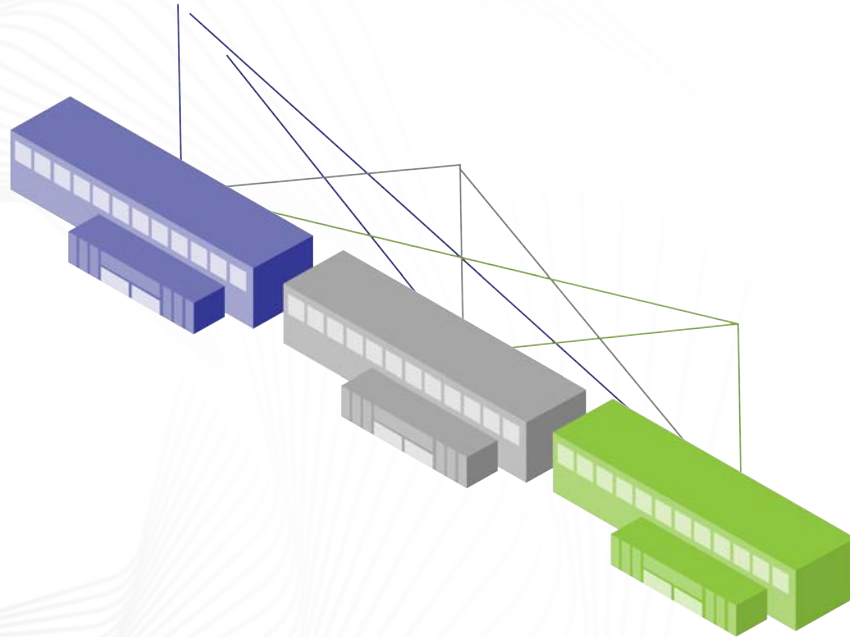
→ from 4 x 128p multi-chip 400G fabric switches



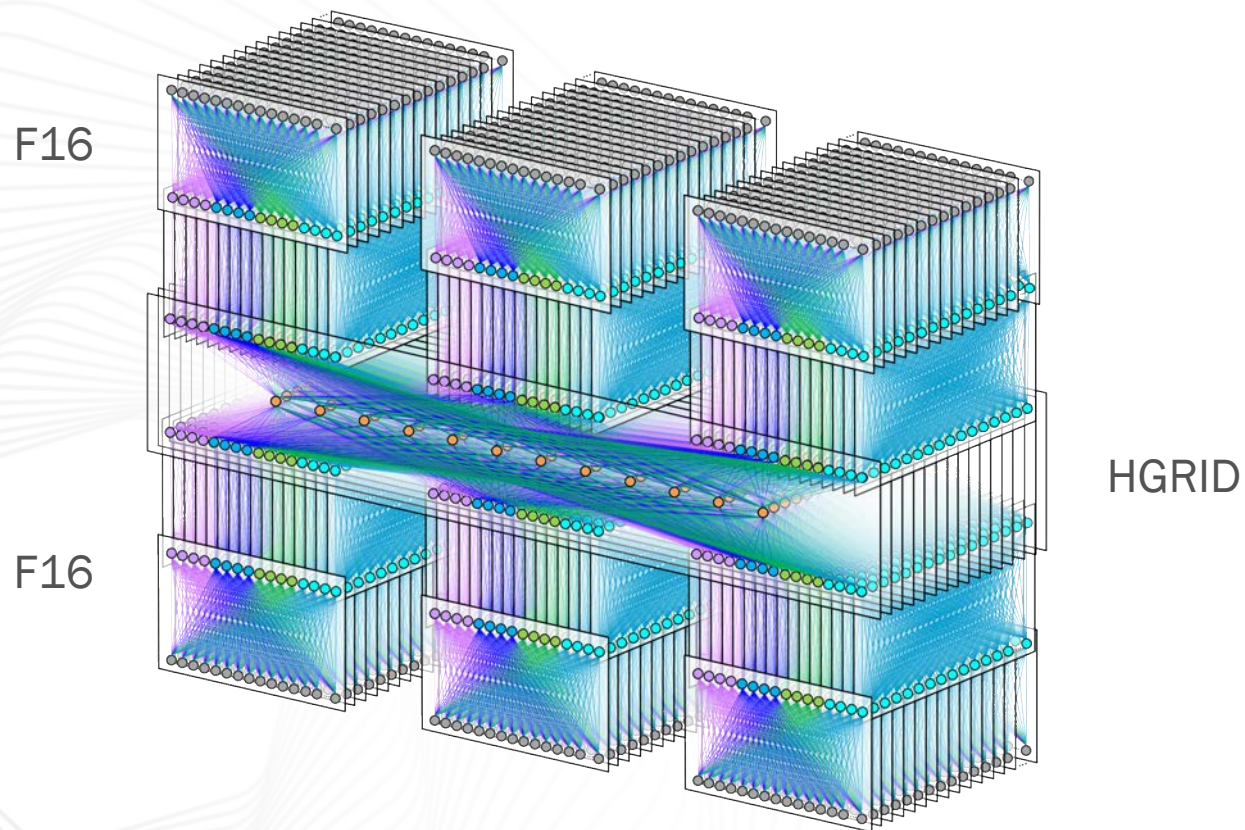
→ to 16 x 128p **single-chip 100G** fabric switches



HGRID — Scaled-up Fabric Aggregator



HGRID



The background is a solid dark blue. On the left side, there is an abstract graphic consisting of numerous thin, wavy, yellow lines that flow from the left edge towards the center, creating a sense of movement and depth. The lines are more densely packed in some areas, creating a gradient of yellow from light to dark.

The Hardware: Minipack

Minipack – next-generation 128 x 100G Switch

Single 12.8T ASIC

Modular design

Mature optics

Lower power/smaller size

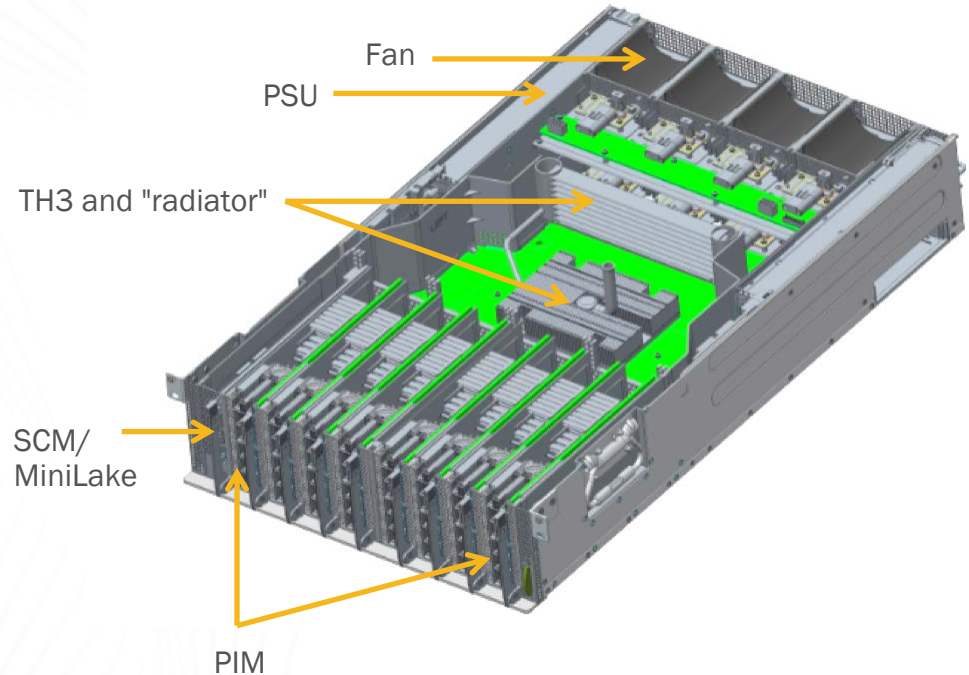


Minipack Modular Architecture

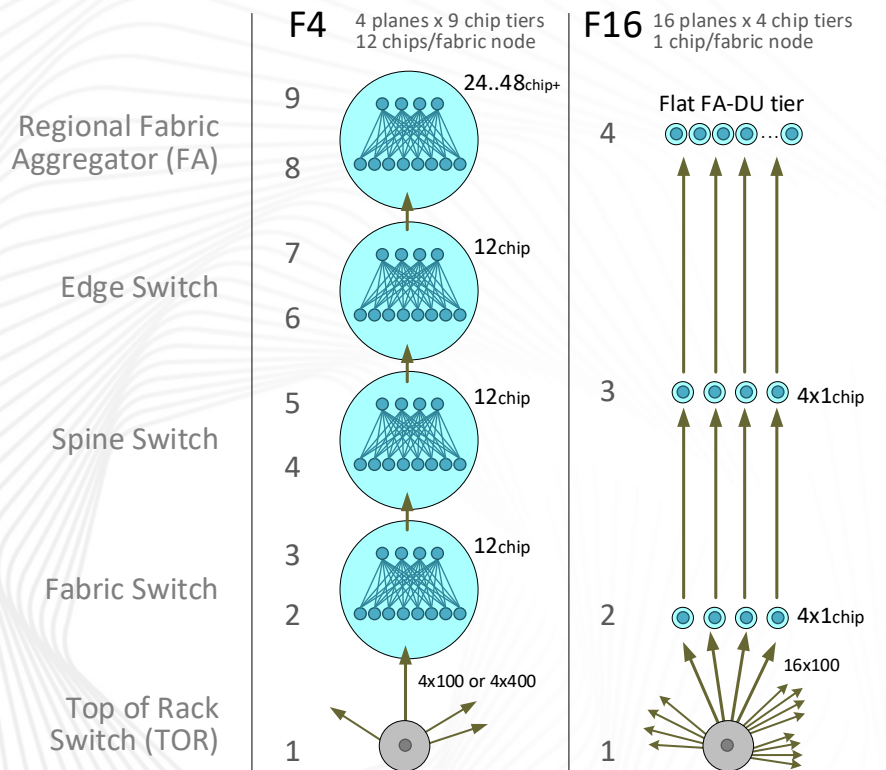
Orthogonal-direct architecture

FRU-able, modular PIM
(line card)

FRU-able SCM (carrying **MiniLake**
microserver) for easy DIMM and
SSD replacement



Philosophy: Simpler and Flatter



Optimize the number of I/O's and chips for the end-to-end application.

A concept that can be applied widely.

Moving Fast

- Develop more hardware abstractions
- Expand testing capabilities
- Continuous deployment

The background is a solid dark blue. On the left side, there is an abstract graphic consisting of numerous thin, golden-yellow lines. These lines originate from the left edge and flow towards the right, curving and overlapping to create a sense of motion and depth, resembling a stylized wave or a series of flowing ribbons.

Packet Optical Strategy

Co-packaged optical solutions



- Conventional pluggable optics – not scalable



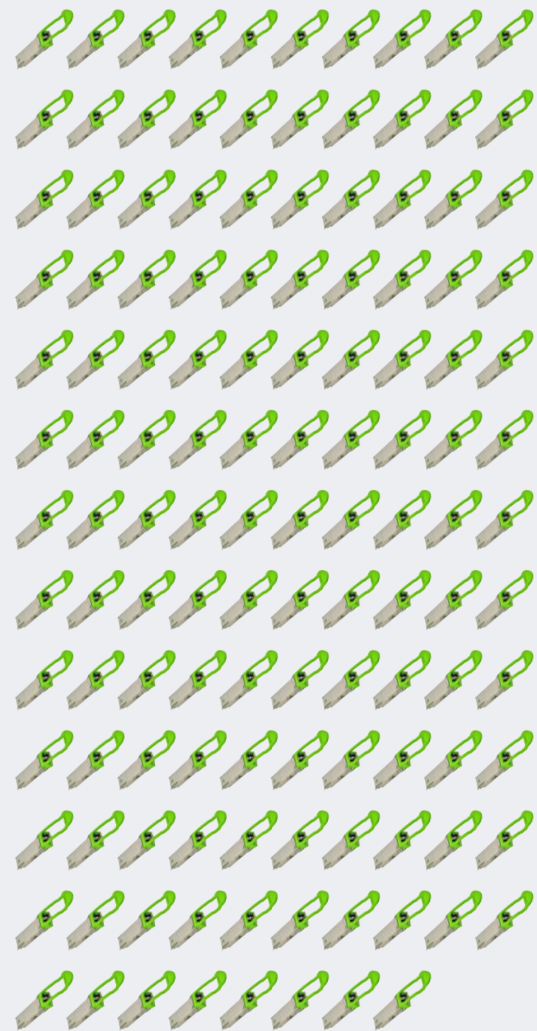
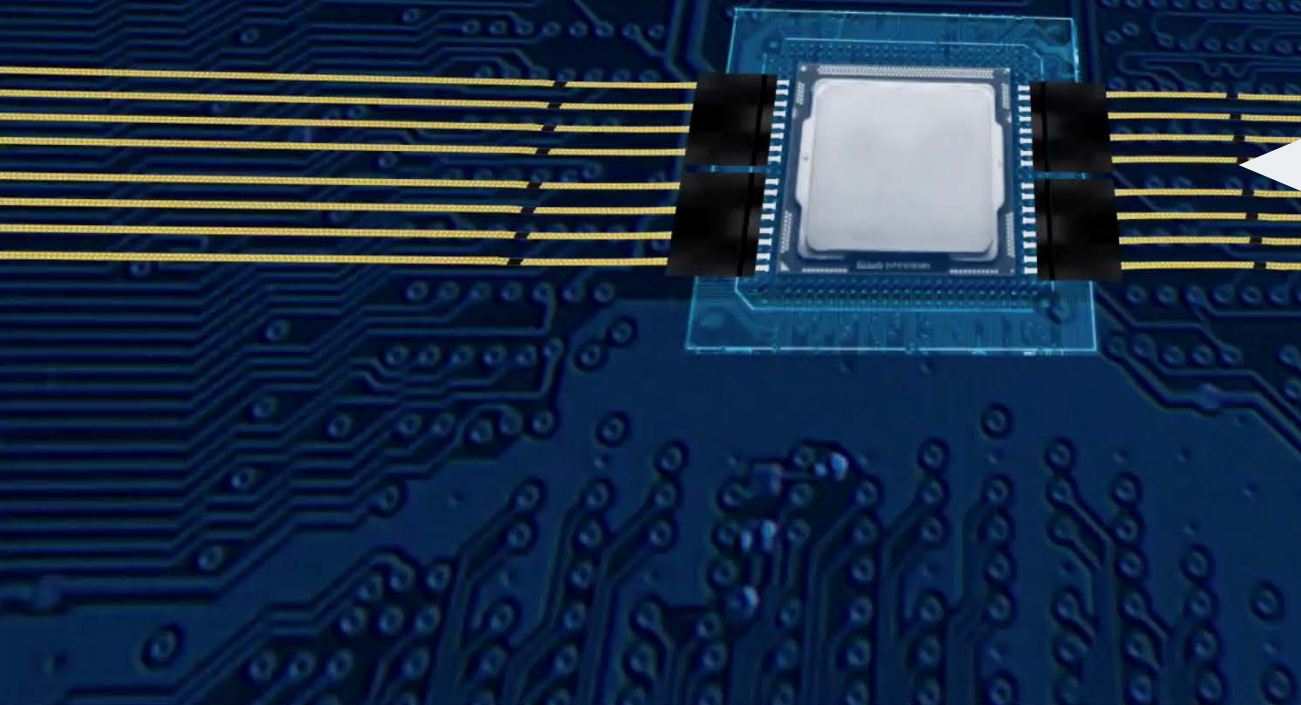
- On-board optics



- Co-packaged optics
 - Wafer level integration/test
 - Packaging challenge to wafer
 - Better process control

Benefits of co-packaged optics

Enables higher density, lower power, and lower cost

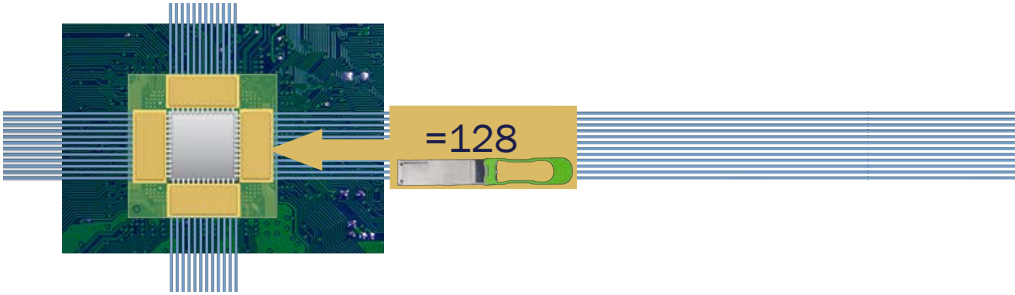


Evolution of Optical Interconnect

Optical
Transceiver

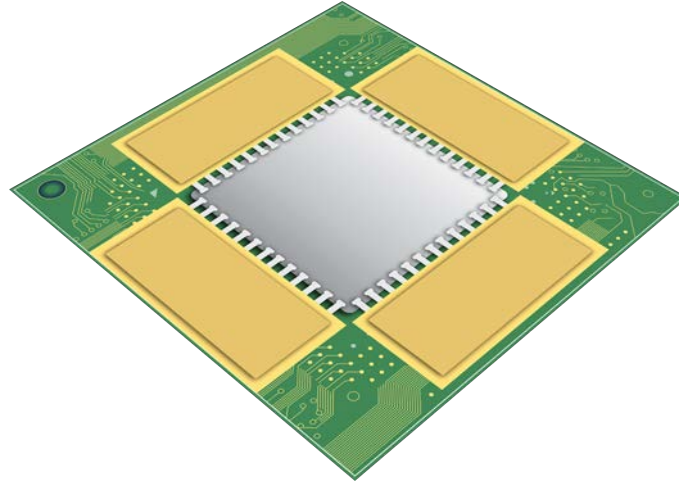


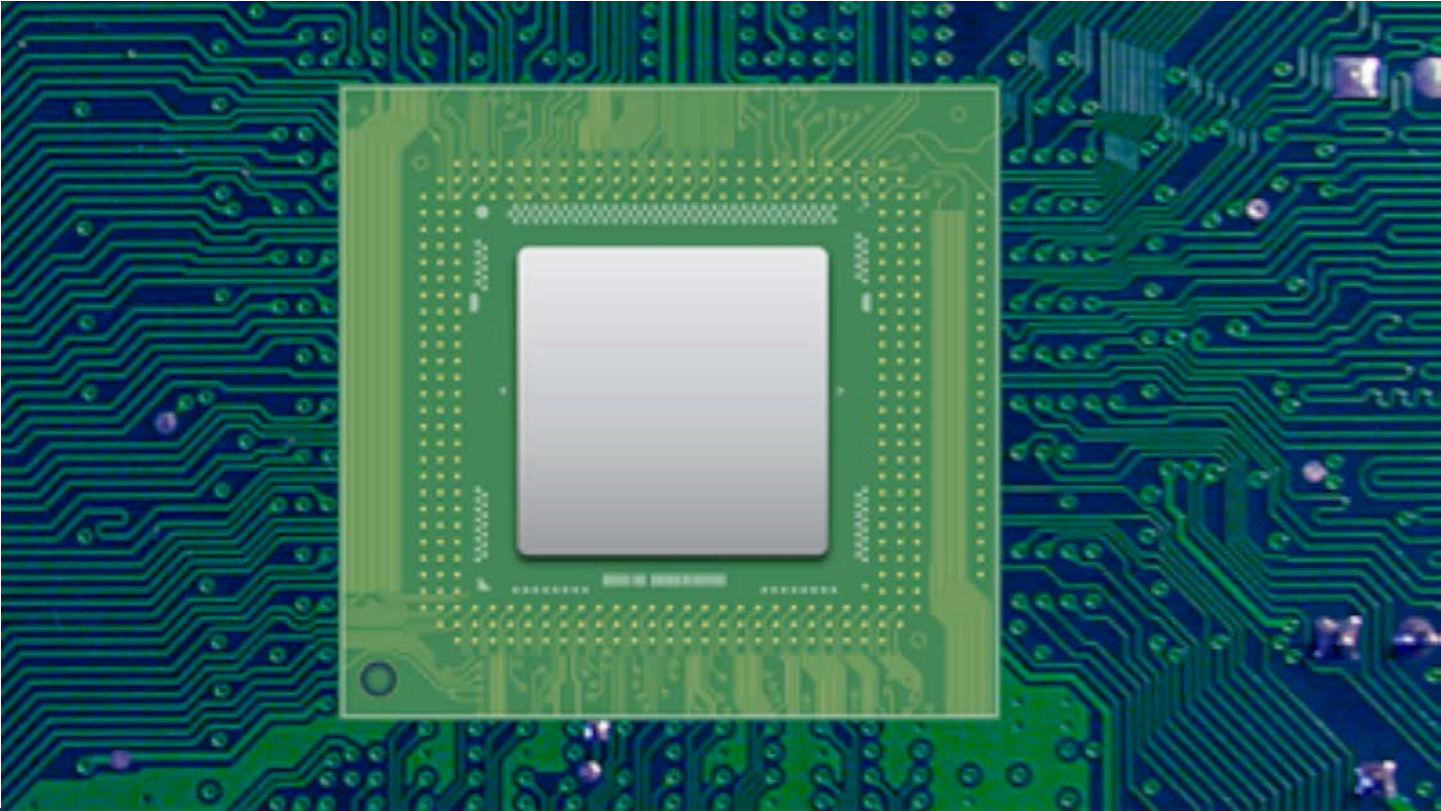
Co-packaged
Optics



Co-packaging Optics and Switch ASIC

- Improved reliability
- Scalable manufacturing
- Designed for cost





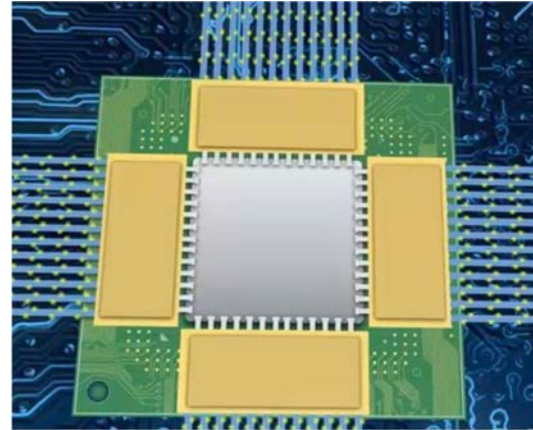
Co-Packaged Optics (CPO)

Collaboration with Microsoft
Build ecosystem to foster many business models

Three discussion documents

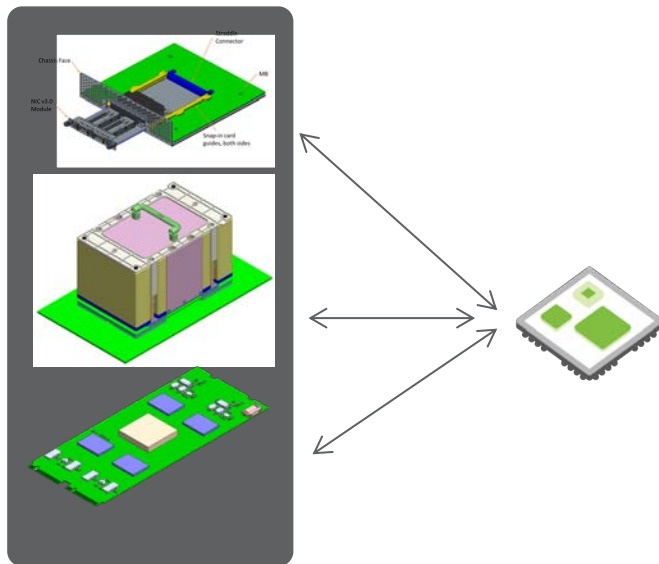
- Co-Packaged Optics Assembly
- Optical Module (Released)
- Optional External Light Source

Inquires to info@copackagedoptics.org
Group site

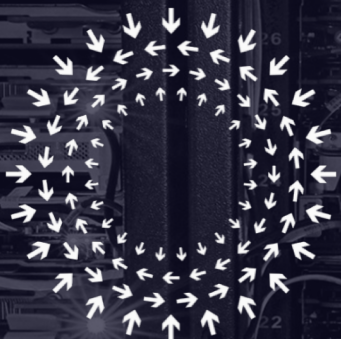


Chiplet Interconnects

ODSA working group within OCP



- Accelerators drive the requirements
- Chiplets are a means to meet requirements
- OCP Form Factors drive Power, I/O Footprint, Performance
- Reference architectures for: Networking, Storage, Inferencing, Training, Video and Image processing



OPEN
Compute Project



TELECOM INFRA
PROJECT

Summary

- Supply does not meet workload demand
- Solutions being worked on
 - Topology optimization, resource distribution including compute disaggregation, flash disaggregation
 - Component development
 - Co-packaged optics
 - Scale out of networks
 - FA & next generation fabrics
- Industry-wide implications
 - Any e2e cloud architecture can make use of tech
- Collaboration between TIP and OCP



Open. Together.



OCP
REGIONAL
SUMMIT