Google

Low Latency for Modern Datacenter Applications

Xiang Zhou

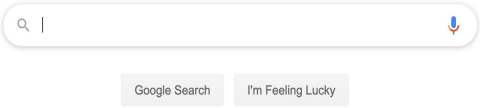
Platforms, Google Inc.

OFC Workshop: Low Latency Communications — Where Do We Need It?

Datacenter-Enabled Latency-Sensitive Apps

- Web-page based applications: O(100ms) RTT desirable
- Cloud streaming interactive video/game: O(10ms) RTT desirable

Google



Google Cloud

YouTube

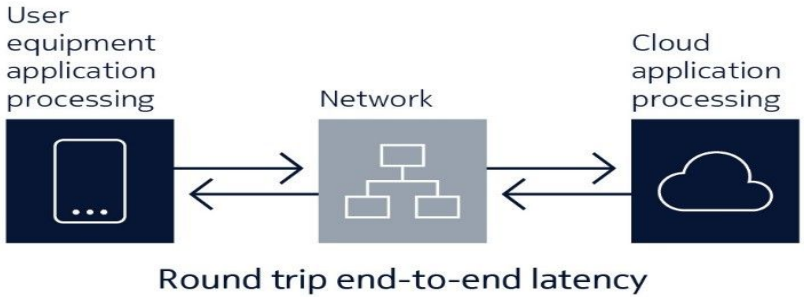
Google Meet

STADIA

Most latency sensitive applications

Latencies that Impact End Users Experience

- **Network latency**
  - Transmission medium (fiber, air): fiber=5µs/km; Air=3.34 µs/km;
  - Transport Optics: O(ns) to O(10µs)
  - Switch/Router: O(100ns) to O(10µs)
- **Computation latency** (App dependent)
  - Datacenter as a computer
    - Massive parallel computation enabled by high-speed interconnect technologies
    - Machines and interconnects both critical for lower cloud processing latency
  - End device computation



Round trip end-to-end latency

Google's DC Interconnection Network

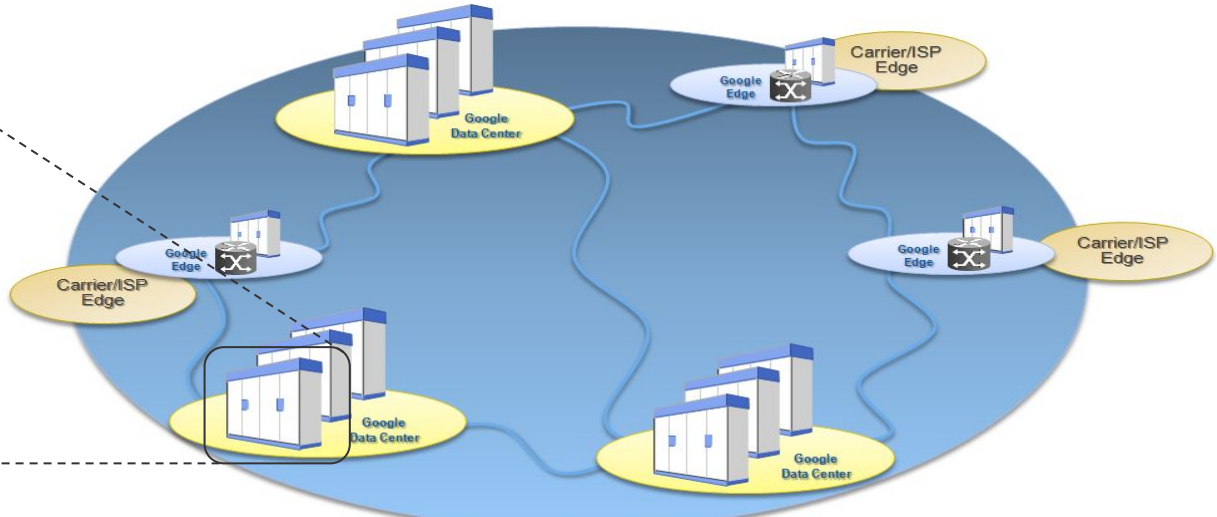
Intra-DC optics latency much more critical than inter-DC optics

Intra-DC (Clusters)	Intra-Campus (Campus)	Intra Metro POP - POP (Metro)	Inter-metro DC - DC; DC-POP (Backbone, LH / Subsea)
<5µs	**<50µs**	**Up to 400µs**	**>5ms**
< 1km	< 10km	~40-80 km	1000s km

Fiber latency

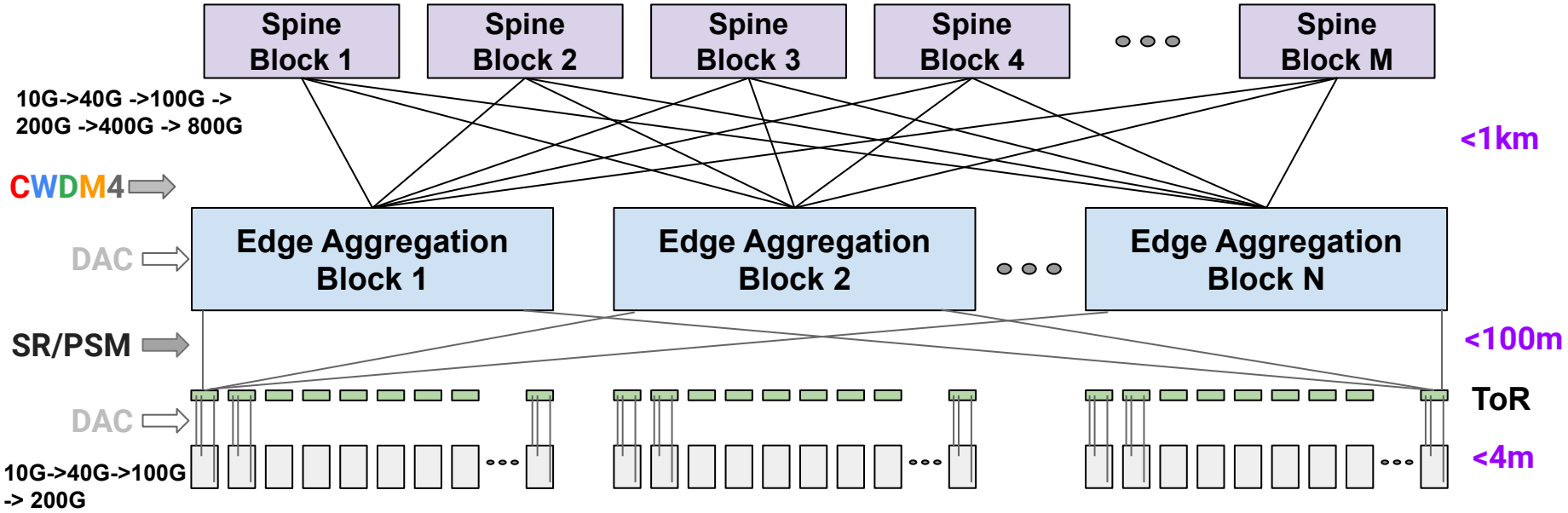
Distance

High Efficiency Servers
(www.google.com/about/datacenters)



Google

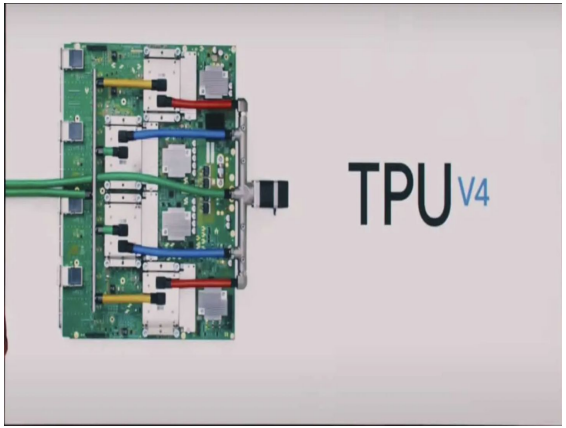
Datacenter Warehouse-Scale Computer (Cluster)



- Low-latency interconnect technology needed for warehouse-scale computer
- TOR to machine and TOR to Edge most latency sensitive
  - Interconnect latency O(1µs) or less desirable

https://www.ieee802.org/3/B400G/public/21_03/lam_b400g_01a_210329.pdf

Datacenter Machine Learning (ML) Cluster

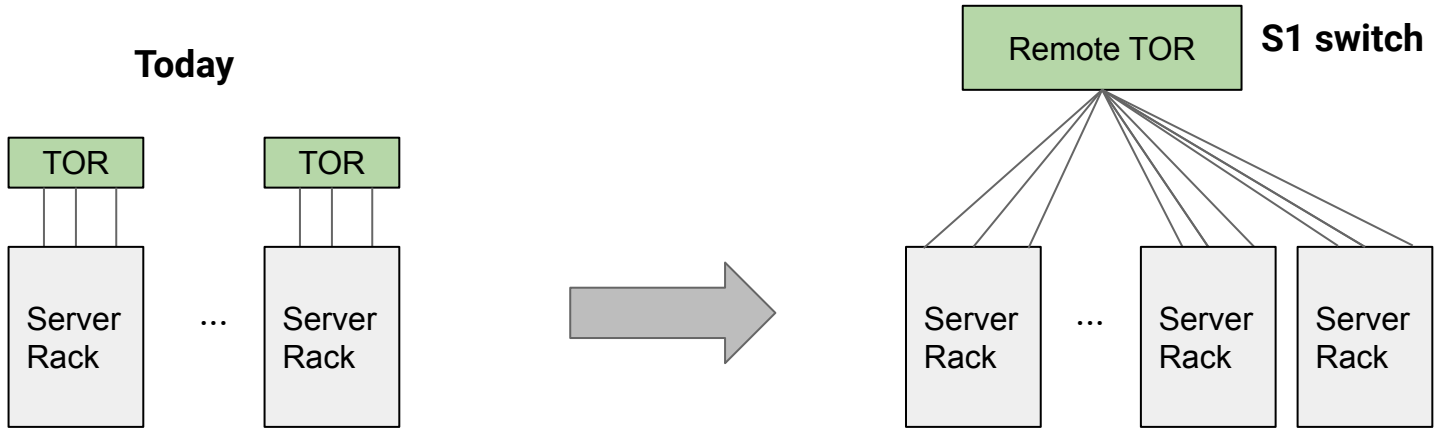


- Gen4 ML cluster/Pod (4,096 TPU chips)
  - Provide more than an exaflop of compute, equivalent to about 10 million average laptop processors at peak performance
  - Enabled by very high bandwidth and low-latency optical and electrical interconnect technologies

Datacenter Network Disaggregation

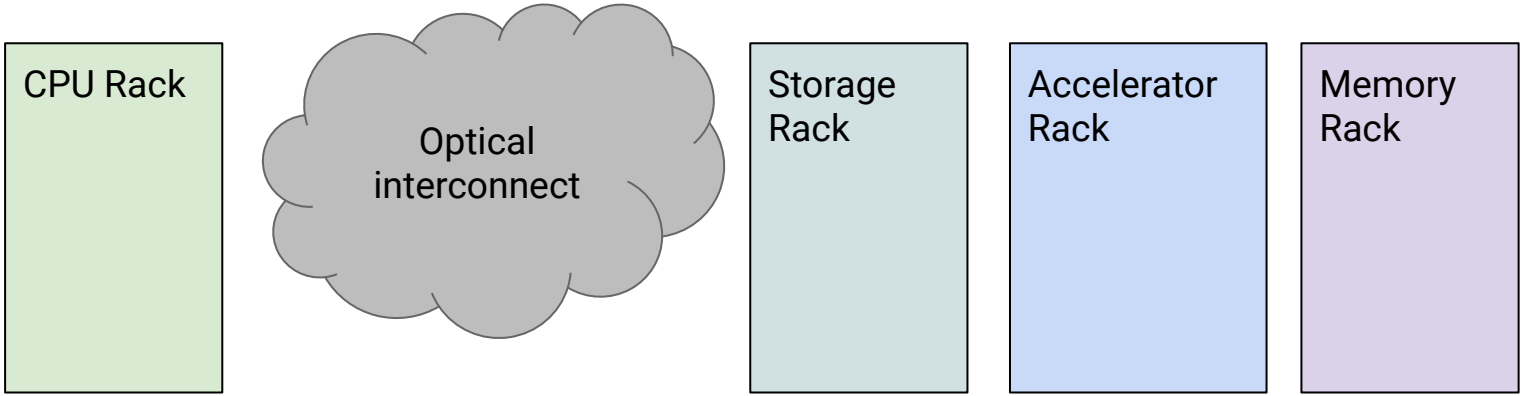
An Exploratory Concept

- Separate network from machine racks with optical interconnect. Provide flexible BW to servers and solve/mitigate network stranding
- Increased transmission distance between server and S1 switch (Remote TOR) imposes tighter latency requirements on the interconnect technology (applications dependent)



Datacenter Server Disaggregation

- Separate servers into core components: compute, memory, accelerators, storage and connect them with optics. Resource can be flexibly allocated per application.
- Much tighter interconnect latency requirements, especially for Memory disaggregation !!



Ref: Requirements of Optical Interconnects for Next Generation Data Centers, Z. Shen, OI'19

Low-Latency Optical Interconnect Technology

- Low-latency optical transceiver
  - Low-latency and high-gain FEC
    - Needed for higher SE modulations (PAM4/6, coherent QAM etc) to address SNR scaling and error floor challenges
  - Low-latency DSP
    - More DSP needed to scale the lane speed (components BW constraint and link impairments). Low latency implementation is critical !
  - Low-latency and low-cost optical amplification technology
    - To enable future optically-switched low-latency datacenter
- Low-latency transmission medium
  - Low-cost and large-volume manufacturable hollow-core fiber
  - Free-space optical links for optical backplane (within the rack) ?

Conclusions

- Low-latency optical interconnect technology enabled warehouse-scale computer and Supercomputer, which has resulted in
  - Dramatic reduction of computation latency for many computation-intensive applications
- The potential need for datacenter network and server disaggregation imposes more stringent interconnect latency requirements
- Innovations are needed in low latency FEC, DSP, optical amplification and transmission medium technologies to continue scale our cloud datacenter capability