Programmability is Key from Edge to Cloud in the Data-Centric, 5G-Enabled World

October 1, 2020

Executive Summary

How will networks for communication service providers (CoSPs), cloud service providers (CSPs), and enterprises evolve to handle the dramatically increasing data volumes expected in the coming years? Increasing data volumes are being driven today by smartphones, laptops, IoT, and, in the near future, by emerging 5G-enabled services. 650 Group's internal projections indicate that data entering/exiting the data center (north/south) is driven mostly by consumer content (e.g., video). In contrast, a wide range of use cases ranging from enterprise applications, consumer data, and cloud applications drive data between machines.

IoT sensors don't transmit a lot of data individually. Still, there are more IoT devices today than the number of smartphones globally[1], which is driving significant aggregate data transmission. 5G is accelerating new network use cases, such as self-driving or autonomous driving. As a result, according to 650 Group research, data continues to multiply, doubling every two to three years[2].

To support this expansion of data, network bandwidth is going through a renaissance that includes upgrades at the network edge to 5G to data center access networks, the latter moving to 25Gbps to 100Gbps, and soon 400Gbps and beyond (Figure 1).

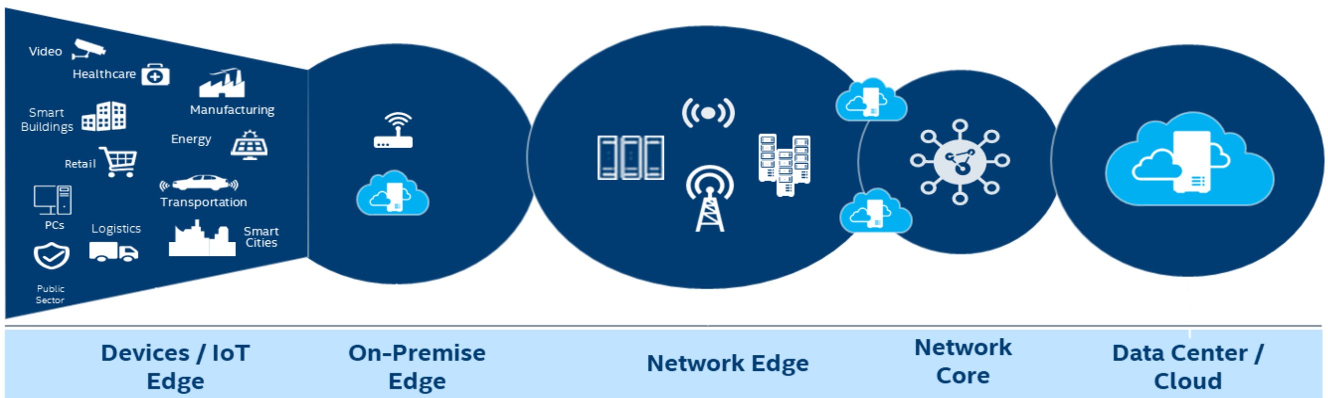


Figure 1: Innovation in Network Connectivity Key to Address Growing Demands From Edge to Cloud

All this data needs to be processed, analyzed, and transported to users worldwide with large data centers and more intelligent Artificial Intelligence (AI) and Machine Learning (ML) edge systems ready to crunch data workloads, the size of which was unimaginable just five years ago.

Insight into this data requires high performance compute, storage, and networking that is not only workload-optimized, but is secure and can scale for bleeding-edge, hyper-scale cloud data centers, CoSPs, and enterprise multi-cloud deployments. As packet throughput increases, network congestion can have a much more significant impact on network performance, thus driving the need for real-time telemetry data that can identify and react to network hot spots rapidly.

Research from 650 Group has identified seven macro technology trends emerging to enable CoSPs, CSPs, and enterprises to stay ahead of this data deluge. This paper is written for Intel to show how the company's connectivity technologies can benefit CoSP, CSP, and enterprise networks.

---

[1] 650 Group IoT Report, July 2020; IOT Device Installed Base of 21B devices at the end of 2020.

[2] 650 Group Merchant Silicon Report, June 2020; Networking bandwidth inside the data center doubles on average every two years

Executive Summary

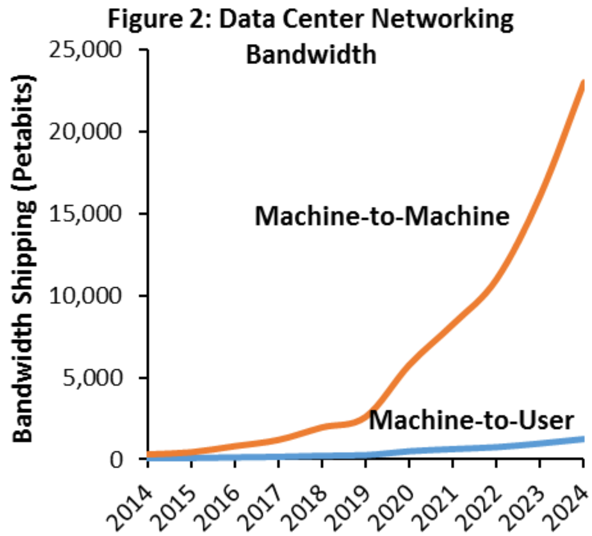
Intel has a broad portfolio of connectivity solutions from Ethernet NICs and switch ASICs to leading silicon photonics products, along with Intel® Xeon® Scalable and Intel Atom® processors, Intel FPGAs, and accelerators to process network workloads. Intel also has been a significant contributor to open source initiatives for networks, including Data Plane Development Kit (DPDK), Open Network Edge Services Software (OpenNESS), Hyperscan, Linux®, Open vSwitch© (OvS), SONiC, and P4, among others, to enable industry innovation. With the combination of an end-to-end portfolio, a broad ecosystem of partners optimizing solutions using Intel's technologies, and a decades-long history of serving small businesses up to the largest hyper-scale cloud customers, Intel's product portfolio helps solve real-world demands placed on networks from the edge to cloud.

In conjunction with this white paper, 650 Group did a webinar with interviews across Intel's DCG executive team to talk about additional market trends. The webinar is complimentrary to this white paper and can be viewed with the link below:

https://www.brighttalk.com/webcast/10773/434496?utm_source=650-Group&utm_medium=brighttalk&utm_campaign=434496

Understanding Today's Connectivity Technology Macro Trends

Humankind is in the era of data and device explosion. Compute, networking, and storage needs to scale to accommodate data processing and analysis to serve customers better and glean valuable insights from data. Hybrid and multi-cloud deployments are the new normal as enterprises increasingly embrace multi-cloud infrastructure and services to keep up with growing demands, increase levels of efficiency, and better utilize growing volumes of data. The amount of bandwidth is growing rapidly, especially machine-to-machine (Figure 2)

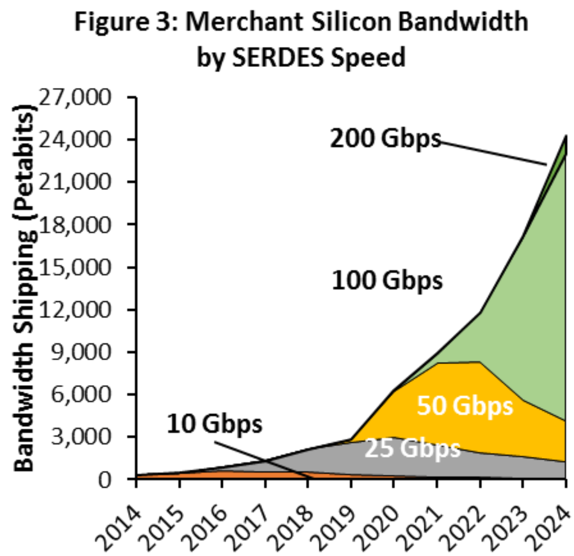


The amount of data processed at each device, and subsequently being sent back to CoSP, CSP, and enterprise networks for further processing, is increasing rapidly.

In 2020, the ecosystem no longer resembles the start of the previous decade. Hyper-scale cloud providers are typically at the bleeding edge of adopting new generations of compute, networking, and storage at an unprecedented scale. To put this in perspective, just one of the largest hyper-scale cloud providers has more compute, network, and storage capacity today than the entire CoSP industry[3][4][5]. While these hyper-scalers are very large-scale consumers of IT infrastructure, the amount of enterprises, big and small, still drive a significant volume for the market as well.

At the same time, the evolution and promise of 5G is a major initiative among both CoSPs and CSPs. 5G offers a technology platform for services such as enhanced mobile broadband (eMBB), ultra-reliable low latency (URLLC) applications, and massive machine-type connectivity for IoT. Edge computing is becoming increasingly more pervasive, driving the need for more intelligent and higher performance compute capabilities at the edge. Optimizing the infrastructure from the edge to cloud is increasingly more important to deliver a great customer and consumer experiences.

New, emerging workloads and usage models, such as artificial intelligence and machine learning (AI/ML), visual computing, autonomous driving, and more, are putting pressure on the network infrastructure. At the heart of this data and device era is the actual data itself. Data is only useful if it can be processed efficiently to uncover essential insights. For consumers, data can mean personal photos and memories shared with friends and family. For enterprises, information is the lifeblood of the company, directing the strategy and future of the company. To drive insight, new AI and ML tools process immense quantities of data for intelligence. This vast volume and variety of data to/from the consumer drives the need for higher bandwidth connectivity.



---

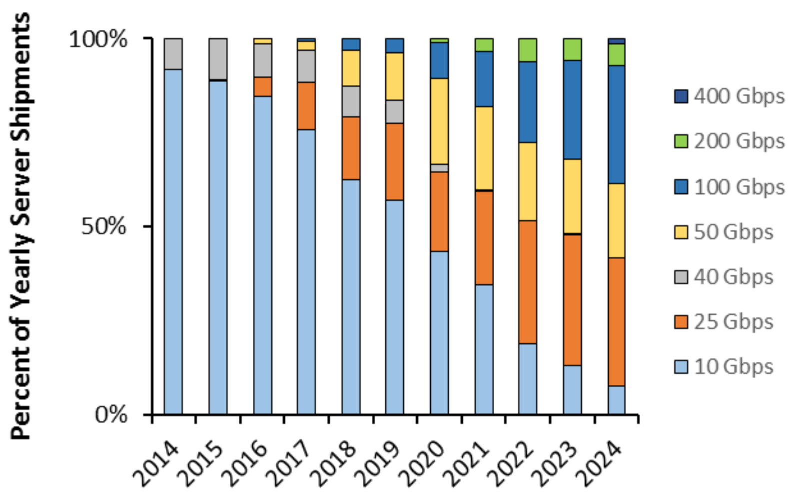
[3] 650 Group Server Report, 1Q20, June 2020
[4] 650 Group Data Center Switching Report, 1Q20, June 2020
[5] 650 Group Storage Report, 1Q20, June 2020

Based on 650 Research, there are eight key innovation areas for networking technology: bandwidth, virtualization/containers, 5G/Edge, NICs and switching, security, programmability, storage, and fiber optics. The next part of the paper will explore each of these trends individually and highlight key Intel innovations that impact each corresponding trend.

Bandwidth – Network bandwidth continues to increase at every layer of the network. Bandwidth for data center switching continues to grow over 30+% per year with evolution to higher speed SerDes (Figure 3, previous page). Consumer broadband connections are moving towards multiple gigabits per second (Gbps) up to 10 Gbps. Workstations and WLAN APs in the enterprise are moving away from 1 Gbps to multigig and up to 10 Gbps links for the first time in nearly two decades. Server access in the enterprise is moving from 1 Gbps to 10G Base-T or 25 Gbps Direct Attach Cable (DAC), and hyper-scale cloud providers are adopting millions of 25, 50, and 100 Gbps connections each quarter to fulfill their needs. 1-100 Gbps meets most market needs through 2022 for server access bandwidth. (Figure 4). In switches, 3.2 Tbps or smaller ASICs is the sweet spot for the market until the wave of 400/800 Gbps product usher in a new wave of innovation in the middle third of the decade.

Intel Innovations in Network Bandwidth – Intel has been driving innovation across its connectivity portfolio to support growing bandwidth demands with Intel Ethernet NICs, Intel Tofino series of P4-programmable switches, and higher bandwidth Intel Silicon Photonics. It's more than just "bigger pipes" because this portfolio delivers several advanced capabilities to address the need for more flexible, programmable, high-performance networks. More details are shared in the sections below.

Figure 4: Server Speed Migration



Network Function Virtualization – CoSPs are catching up with CSPs in embracing the network agility, and simplified service deployment that comes with network functions virtualization (NFV) and containerization. Network virtualization and containers play a pivotal role in network build-outs. Hyper-scale cloud providers look towards placing containers directly in each network element to virtualize the network to allow up to thousands of simultaneous workloads to run securely on a server. Without innovation here, the humans managing the networking simply could not scale with the number of workloads being added to the network each day (Figure 3).

Intel Innovations for Network Virtualization: Intel has a long history of delivering technologies, enabling server virtualization to bring a wide range of capabilities to support virtualized networks, including technologies for SR-IOV, enhanced network virtualization overlays, Flexible Port Partitioning (FPP), and several other capabilities. Intel's offerings and efforts go beyond hardware, with many of the company's software engineers working on network virtualization for the company's connectivity offerings. To enable rapid ecosystem innovation, Intel is very active in the open-source community and is a top contributor to several open-source frameworks and projects, including DPDK, SPDK, P4, SONiC, Open vSwitch, OpenNESS, and several others.

Security – While all these bits of data transfer through wired links and wirelessly all around us, it must be secure. Each step of the bits' journey must be secure. The NIC needs to encrypt and accelerate security features instead of burdening the CPU. The network needs to act as a secure pipe with the ability to recognize and detect rouge packets and separation between applications. Without security build into the design of each component as well as adaptive to future threats, the bits of data would not be safe. Intrusions need to be prevented if possible, detected when they do happen, and steps taken to recover back to a known and trusted state.

Intel Innovations in Security:
With Intel's Ethernet 800 Series, Intel implements a design philosophy of platform resiliency with three vectors of security: Protect, Detect, and Recover. Hardware Root of Trust protects the firmware and critical device settings with built-in detection of corruption and automated device recovery to ensure returning the device to its originally programmed state. Intel also delivers accelerators, such as Intel QuickAssist Technology (Intel® QAT), that speed up encryption/decryption, as well as data compression for improvements in security, authentication, and file storage. At the same time, Tofino's design incorporates security principles, and the programmability of P4 can react and adapt to threats.

Programmable Networks – The bulk of the networks in the world have been built using fixed-function Ethernet switches. And although the speeds and feeds have gotten a lot faster over the years, the techniques and methodologies used in Ethernet switch ASICs have largely remained fixed and proprietary.

Programmable switch ASICs using the P4 language have changed this network paradigm by providing a switch fabric that can support new network protocols, run networking applications in the switch fabric, and provide in-band network telemetry (INT). P4 programmable switches were initially designed for the data center, where flexibility and performance advantages are in high demand. But as programmable switching becomes more mainstream, it is finding new applications at the network edge and CSP and CoSP networks.

Intel Innovation in Programmable Switching – To expand its Ethernet product portfolio, Intel in 2019 acquired Barefoot Networks, a pioneer in programmable switching. Their P4-programmable Tofino switch ASIC successfully debunked the myth that programmability came with lower performance, higher power, and higher cost. In fact, networking research showed that a P4-programmable switch like Tofino could deliver low power, high performance, and programmability[6].

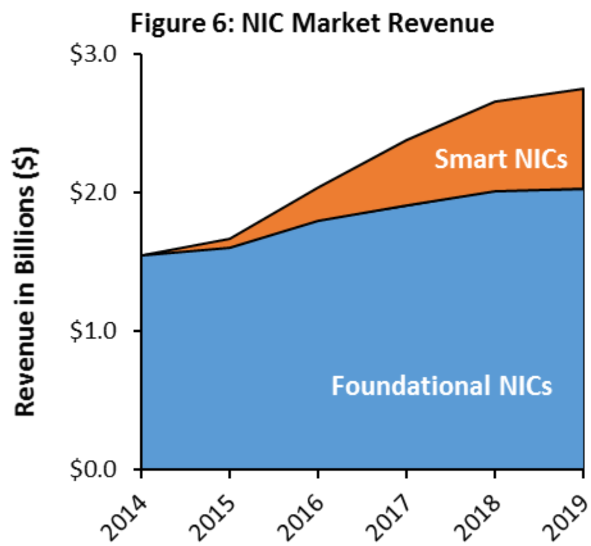
Figure 5: Intel's Tofino (6.4 Tbps) and Tofino 2 (12.8 Tbps)

Intel's Tofino and Tofino 2 are the first commercial, P4-programmable family of Ethernet switch ASICs on the market, and offer switching speeds of 6.4 Tbps and 12.8 Tbps, respectively (Figure 5). The company has also innovated with open-source In-band Network Telemetry (INT) and created Deep Insight. This management platform enables fine-grain measurement of all traffic flows with an ability to interpret, analyze, and identify the root cause of issues in real-time with nanosecond accuracy.

Meanwhile, Intel's FPGAs, such as the Arria 10 and Stratix 10, provides programmable compute functions within switch systems in use cases that require additional scale and features serving cloud, enterprise, and communications service providers.

[6] P. Bosshart, G. Gibb, H.-S. Kim, G. Varghese, N. McKeown, M. Izzard, F. Mujica, and M. Horowitz. Forwarding metamorphosis: Fast programmable match-action processing in hardware for SDN. In ACM SIGCOMM, pages 99–110. ACM, Aug. 2013.

Ethernet NIC Evolution – Ethernet NICs have evolved into two categories - Foundational NICs, which address the majority of market needs for Ethernet connectivity, and SmartNICs, which offer compute capabilities to address a segment of the hyper-scale CSPs and CoSPs. Both markets thrived through 2019[7] (Figure 6).



Foundational NICs can be used as the workhorse for the majority of networking needs. Foundational NIC capabilities have continued to evolve with higher bandwidth, greater levels of intelligence, and capabilities to address data-demanding requirements for most segments of the market.

A SmartNIC is a Network Interface Card that accelerates host networking functions and is adaptable to existing and emerging use cases, including security, virtualization, storage, load balancing, and data path optimization. SmartNICs leverage existing Ethernet networking capabilities in Foundational NICs where acceleration is achieved by processing tasks, which the system CPU would normally handle, through a tightly-coupled compute engine such as CPU and/or FPGA. Adaptability is achieved by standard and easy-to-use programmability of hardware functions and is significantly software-defined.

By moving some of the infrastructure workloads to the SmartNIC, the host CPU cores are freed up for greater levels of scaling with an ability to add even more applications to the host. We find that there is a wide diversity in customer requirements for SmartNICs. Today, some sizeable hyper-scale cloud customers that have deployed custom SmartNICs optimized for their particular needs and other CSPs and CoSPs are also exploring how SmartNICs customized to their needs can improve performance in their infrastructure.

Intel Innovation in Ethernet NICs

Intel Ethernet NIC offerings have driven technology and market leadership for multiple decades with a consistently strong lineup. Supported by a broad ecosystem of systems providers worldwide, Intel Ethernet 700 Series and Intel Ethernet 800 Series continue to add greater bandwidth, as well as more features and functionality with each product release (Figure 7). NIC feature lists often span pages, but two important highlights from our end-user interviews include Application Device Queues (ADQ), which improve application response time predictability, lower latency, and improve throughput for a wide range of critical applications, including web, caching, and database

Figure 7: Intel's Ethernet 800 Series Adapter and Controller

[7] 650 Group Server and Smart NIC Report, 1Q20, June 2020

tiers within the datacenter. The second important feature is Dynamic Device Personalization (DDP) that delivers more efficient, higher performance packet processing of new and advanced network protocols by taking advantage of Intel Ethernet's programmable pipeline.

Given the diversity of customer requirements for SmartNIC functionality, Intel enables focused SmartNIC solutions built with Intel FPGAs and SoCs, primarily for large CSPs and CoSPs, and has long- term roadmap. One example for CoSPs is the Intel FPGA Programmable Acceleration Card N3000 that uses Intel FPGAs and Ethernet NICs, which provides programmability via RTL to tailor the functionality of the SmartNIC.

Storage – Many early storage deployments were based on dedicated Fibre Channel networks for storage systems within a data center. This changed with mainstream cloud architectures that have relied on Ethernet for storage networks, given its flexibility, low cost, and ease of implementation. The advent of flash-based storage for servers led the network to play a critical role in high-performance storage in the cloud and within the enterprise. Today, connecting NVMe storage using NVMe over Fabric is common, with Ethernet being the most popular transport. Depending on the use-case, both RDMA and TCP are important for the implementation of NVMe over Fabric.

Intel Innovation for Storage Networks – Intel Ethernet 800 Series NICs provide support for high-performance storage networks and NVMe over Fabrics through the support of multiple Ethernet storage protocols, including iWARP, RoCE v2 RDMA, and NVMe over TCP supported with ADQ. Intel's participation in the standards bodies and storage working groups ensures a standards-based approach with a more significant vendor ecosystem compared to Fibre Channel.

5G / Edge – Driven by 5G networks and the need for computing closer to the edge, CoSP networks are quickly pushing server functionality to points of presence (PoP)s, base stations, and other edge locations. CSPs are also responding by offering its infrastructure to support 5G networks. Over the past five years, Ethernet switch-based networks have replaced routing platforms in many CoSPs and CSPs, and 2020 marks the first year of significant movement further down the stack as Ethernet switch platforms, with long-reach optics, begin to replace purpose-built optical transport systems in the metro. Ethernet now plays a role from the base stations to the CoSP central data centers, as well as to the mega CSP data centers. The key to Ethernet's pervasiveness is it's built on open standards, widely available through a broad ecosystem, and easy to deploy with support for a growing number of use cases including VNFs, virtual gateways, virtual firewalls, vRAN, OVS DPDK, vRouter, UPF, and others. One of the primary reasons why the Ethernet switch data center market segment is growing from $10 billion -$25 billion in 10 years[8] is due to the increase in use cases outside the data center and into the transport and long-distance transmission portion of the market.

5G also ushers in new technologies and topologies from the cloud to the edge. Edge computing will range from an individual's own device/s (smartphone, tablet, AR glasses, car) to smart sensors to CoSP PoPs such as base stations, local edge data centers, and others. With that context in mind, edge computing will drive an increase in computing capability and network connectivity needed at all layers and devices.

Depending on the end user's application, compute can occur directly in the edge server at a base station, in a central office, or at a distributed cloud data center. As application developers become savvier on latency, and new applications are created, the cloud will likely turn latency into a clickable or programmable feature for application developers. Large, real-time data sets are a challenge, and edge computing and filtering is the advantage to help process new and continuously evolving data sets. Data sets will need to be right-sized at all layers. For example, all telemetry data from the network will not be sent back to the cloud for processing. Preferably, telemetry data will be processed and analyzed at the edge, and only a subset will transmit to the cloud. Many applications will also work in this paradigm with the intelligence at the edge, processing appropriate amounts of data in the proper location.

[8] 650 Group Data Center Switching Report, 1Q20, June 2020

Constraints such as bandwidth, responsiveness, power, and data set size will determine where the computation occurs.

Intel Innovation for 5G and Edge Networks – Intel brings its heritage as a leader in cloud computing technologies to transform 5G networks from the cloud to the edge—becoming part of the fabric of the network in the same way it is the backbone of the data center. Intel is embedded throughout the 5G value chain, offering an unmatched portfolio of products and solutions that provide agility, flexibility, and performance with Intel® Xeon® Scalable processors, support for all RAN and core network configurations, accelerators, software, a common toolchain, reference architectures and a broad ecosystem of solution providers.

Optical Evolution – For 20 years, network builders and OEMs cared a lot about how large the optical components were and how many ports could fit into a line card or 1RU "pizza box" system. The next generation of optics technology – offering 100 Gbps per lambda – ushers in a significant and profound change.

Optical connectivity historically has focused on the interconnects within a CoSP network such as cell-site backhaul, metro and long-haul optical links, and residential backhaul from PON, Cable, and DSL. A focus on the needs of the CoSP was driven by the fact that they were the most advanced and highest volume customers driving next-generation technologies and speeds. However, with the advent of cloud computing and further advancement in the largest cloud providers, innovation and technological leadership has shifted toward the hyperscale data centers.

Innovation in optical connectivity is on track to evolve through several rapid technological changes over the next five years. First, innovation in silicon photonics is rapidly scaling to support high-volume demand. Second, as higher bandwidth connectivity becomes a necessity in data centers, the market is also driving the need for co-packaged optics.

The industry needs silicon photonics for networks to keep pace with the innovation occurring at the compute portion of hyperscale data centers. Silicon photonics is often shrouded in mystery with an industry perception of optical cables plugging directly into ASICs. While this is a possibility, multiple components benefit from silicon photonics, and in fact, silicon photonics products have been shipping for several years already. For example, the tighter integration within the pluggable module comes from silicon photonics innovation. These modules look no different on the outside than optics from the past two decades. However, they hold a crucial technology innovation in the inclusion of silicon photonics, which, by virtue of the process, deliver high yield and reliability as well as lower cost.

The next step for optical innovation is around co-packaged optics. Instead of terminating light in the pluggable module, the light ends directly next to the ASIC die. And with Ethernet switch ASICs moving towards a chiplet architecture, the logic and I/O portions of the ASIC are separated, enabling innovation to occur rapidly. This also allows for the silicon photonics engines to sit directly next to the switch ASIC die. Hence, co-packaging will become the natural evolutionary path of optics and ASICs over the next several years. The market will begin to see new classes of Ethernet switches in production in one to two switch generations. It is essential to remember that optics and co-packaging can and will exist simultaneously with many hybrid options in the networking layer.

Silicon photonics will permeate beyond just networking. As servers move towards more advanced AI and ML workloads, silicon photonics will be included directly into the NIC to provide high-speed optical links directly from the network to the server. Longer-term, this will redefine the relationship between top-of-rack switches and aggregation of the leaf/spine switch portions of the data center network. Simultaneously, silicon photonics will also become a critical technology for longer reach connections outside the data center, such as ZR/ZR+ type connectivity for Data Center Interconnect (DCI).

Intel Innovation in Optical Connectivity – In the hyperscale data center market segment today, Intel has become a leader in 100Gbps Silicon Photonics pluggable modules (Figure 8). Intel approached the module market for hyperscale cloud providers using new technologies. Intel Silicon Photonics modules, while on the outside looking no different than optics from the past two decades, hold a crucial technology innovation using a unique hybrid silicon laser made with wafer-scale manufacturing. The customer benefits from high-quality time-to-volume and reliability. An essential consideration for future speeds is the initial ramp of higher bandwidth optics that hyper-

Figure 8: Intel's 100 Gbps CWDM4 2km, 100 Gbps CWDM4 10km, and 400 Gbps DR4 Optical Pluggable Modules

scale cloud providers need. One hundred thousand or more optical modules each quarter is necessary for hyper-scale cloud providers to move to higher bandwidth like 400 Gbps. Old manufacturing techniques simply don't scale to these new demands.

In addition, Intel was the first in the industry to show co-packaged optics in early 2020, combining the company's Tofino 2 Ethernet switch ASIC with its Silicon Photonics engine (Figure 9). Only a handful of companies in the world will have the technology to manufacture next-generation switches combining these two technologies.

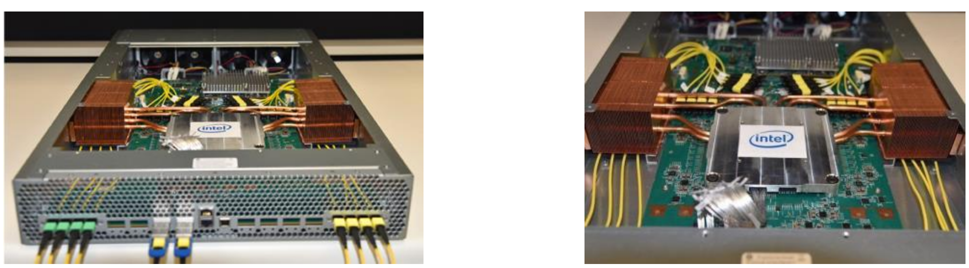


Figure 9: Intel's Tofino 2 Ethernet Switch ASIC with Silicon Photonics Engine

Summary

Processing, analyzing, and gaining insights from vast quantities of data can provide a competitive advantage for enterprises, CSPs, and CoSPs. Emerging workloads like AI/ML, 5G and, intelligent edge computing drive the continued build-out and modernization of data centers worldwide. To keep up with growing demands, Intel is delivering programmability and performance across compute, network, and storage infrastructure.

Network connectivity is the lifeblood of moving data at higher bandwidth and staying ahead of competitors. To that end, Intel has developed a robust portfolio of connectivity products, made significant contributions to open-source software, and created an industry ecosystem that continues to drive advancements.

The company's efforts have impacted foundational NICs, SmartNICs, Ethernet switching, Silicon Photonics, and Intel architecture processors. Intel is also optimizing their portfolio with additional levels of integration to address the demands of next-generation networks.