Independent market research and competitive analysis of next-generation business and technology solutions for service providers and vendors

HEAVY READING

WHITE PAPER

800G Client Optics in the Data Center

A Heavy Reading white paper produced for Cisco

AUTHOR: SIMON STANLEY, ANALYST AT LARGE, HEAVY READING

EXECUTIVE SUMMARY

Demand for the latest high speed network solutions has grown rapidly, driven by the massive shift to cloud services by businesses and individuals. Leading cloud service providers, including AWS, Google, Meta, Microsoft, Baidu, Alibaba, and Tencent, are continually building and upgrading hyperscale data centers with the latest server and networking solutions. The market for client optics is now dominated by these data center operators, which are demanding ever higher bandwidths and quicker time to market for new networking technologies to meet the needs of their customers. The applications demanding the highest bandwidths include artificial intelligence (AI) and machine learning (ML), video processing, and networks for high performance quantitative/scientific research.

The deployment of 400GE client optics was accelerated by the demand from hyperscale web players and service providers, along with other data center operators, coinciding with the availability of a wide portfolio of optical module and cable solutions. Technologies that had been spearheaded for 100GE, such as PAM4 modulation, forward error correction (FEC), and breakout solutions, together with double-density form factors, were key to delivering these 400GE solutions. The backward compatibility of the double-density QSFP-DD form factor has given end users the flexibility to manage the migration from 100GE to 400GE as demands on their networks have grown. These elements, along with the ability to bring coherent pluggable solutions directly to a client port, present a tremendous opportunity for next-generation networking.

The next key development is 800G, and the industry is already gearing up to deploy this next generation of client optics in hyperscale data centers. Developments in three distinct areas are needed for 800G deployment: optical modules and direct attach copper (DAC) cables, switch ASICs, and 800GE standardization. Not all these need to be fully delivered for data center operators to benefit from 800G upgrades.

By understanding the key developments for 400G and 800G, as well as the standards planned for 800G and 1.6T, data center operators can ensure that they benefit from 800G upgrades as solutions evolve. Further, by reviewing the 800G use cases in this white paper, data center operators will gain a better appreciation for which 800G upgrades should be considered initially and which should wait for later deployments.

THE GROWTH OF 400G

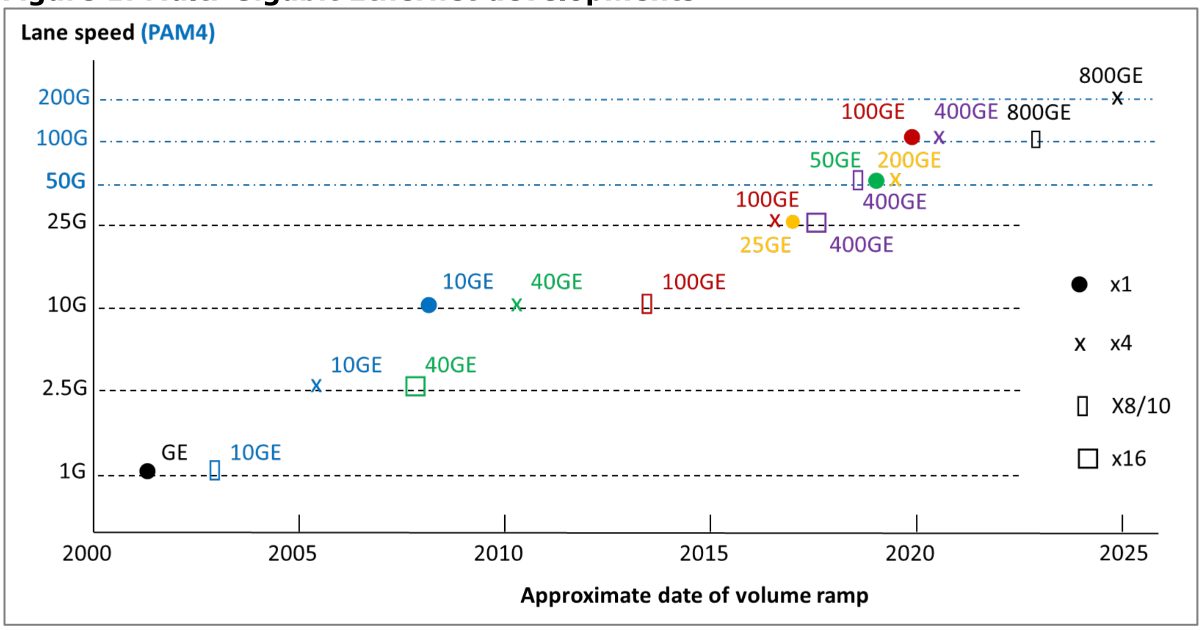
The speed with which hyperscale data center operators have moved to the high volume deployment of 400G demonstrates the huge transition that has occurred in the market for client optics. A market that was dominated by enterprise networks in the early 2010s is now dominated by public and private data center networks. This transition reflects the impact of the shift by individuals and businesses to using cloud-based services.

The vast data centers used by cloud service providers have thousands of identical racks of servers and networking equipment. When hyperscale data center operators start deploying a new generation of client optics, they immediately require massive volumes of optical modules to build out switching fabric and router networks. This puts huge pressure on vendors to ramp up production very quickly and introduce the latest technologies as soon as practically possible. In contrast, enterprise networks have mostly been upgraded over time, requiring a relatively slow production ramp-up of new networking technologies.

Figure 1 shows the approximate dates of volume ramp-up for multiple generations of Multi-Gigabit Ethernet technologies. The most important Multi-Gigabit Ethernet speeds have been 10GE, 100GE, and now 400GE. The next key development is 800GE. For each Ethernet speed, the figure shows configurations of lane speed (1–200Gbps) and the number of lanes (1, 4, 8/10, or 16), which may be wavelengths, fibers, or copper connections. For each of these generations, the time between successive implementations is reduced.

The figure also shows several Ethernet speeds that are, or have been, important for specific applications or as a stop-gap solution. 40GE was widely used until 100GE became cost-effective. 25GE is widely used for access networks and server connections, and some connections are now being upgraded to 50GE. 200GE offers a simple upgrade from 100GE, replacing the four 25G lanes with 50G PAM4 lanes.

Figure 1: Multi-Gigabit Ethernet developments



Source: Earlswood Marketing

To meet the demands from hyperscale data center operators for higher speed networks, the industry has accelerated the time to market for each successive generation by using multiple approaches:

- Reducing the multipliers between key generations:10x (10GE➔100GE), 4x (100GE➔400GE), and 2x (400GE➔800GE)
- Introducing PAM4 modulation and FEC to double the data rate for each lane without increasing the baud rate
- Using more lanes and double-density modules

The result has been faster time to market for the first high volume form factor for successive generations: 10GE using SFP+ modules with one lane, 100GE using QSFP28 modules with four lanes, and 400GE using QSFP-DD or OSFP with eight lanes connecting to the switch ASIC. This has been particularly important for 400GE, where the first high volume modules have used the same baud rates as 100GE modules. Manufacturers have been able to meet the relatively fast time to market demanded by data center operators by capitalizing on the high volume production technologies developed for 100GE.

PAM4 is now well established and supported by a wide range of switch/router ASICs and optical modules. The first high volume generation of 400G client optical modules being deployed in hyperscale data centers are connected to the switch/router ASICs by eight 50G PAM4 lanes. For this implementation, most optical modules integrate a gearbox between the eight-lane switch ASIC connection and the four optical lanes. A new generation of double-density optical module form factors, QSFP-DD and OSFP, were developed to support the eight-lane switch interface. These form factors will also support DAC or active electrical cables (AECs) for the shortest (<7m) 400G links in the data center.

The QSFP-DD form factor is widely deployed in 400G connected data centers and has been a critical factor in delivering cost-effective 400G optical modules with 50G PAM4 lanes. This form factor was developed from the broadly adopted QSFP+/QSFP28 form factors and is backward compatible. Backward compatibility has become an important feature for 400G modules. It allows end users to integrate the newest platforms with the highest bandwidth and latest features while continuing to use lower speed modules as they manage the migration of their network to higher speeds. The QSFP-DD form factor is very flexible, with options for heat sinks on the cages and on the modules. A 400G port using QSDP-DD may support a single 400GE link, two 200GE links, or four 100GE links using a breakout cable. This flexibility is important in data center applications where switches can be upgraded to 400G ports, dramatically increasing capacity; however, the connections to servers can continue to use 100GE. The QSFP-DD specification has been further developed to support 800G applications using 100G PAM4 lanes, and support of 1.6T applications using 200G PAM4 lanes is in progress. These are backward compatible with the QSFP family of modules, including 200G QSFP56 and 400G QSFP112.

The OSFP form factor was designed specifically for 400G applications and has been further developed to support 800G and 1.6T applications. OSFP modules are used by a small number of data center operators and switch manufacturers. The basic OSFP module has an integral heat sink and fits into a cage that supports one or four modules side by side.

The introduction of 400ZR and ZR+ means that QSFP-DD and OSFP pluggable modules can now also be used for longer distance connections between data centers. At 400G, they can (for the first time) use the same form factors as the shorter reach optical and copper interfaces. 400ZR is an Optical Internetworking Forum (OIF) specification, initially developed for data center interconnects, and it supports 400G coherent optical links up to 120km. ZR+ modules cover a much wider range of applications by using enhanced FEC and link-specific optimizations to drive longer distances, and they can support Ethernet rates from 100G to 400G.

Data center interconnect links that had previously relied on separate optical transport systems with integrated coherent transceivers that support DWDM connections can now use QSFP-DD or OSFP, 400ZR, and ZR+ coherent optical modules plugged directly into the switch/router ports, removing the need for separate optical transport systems. This approach is already making a significant impact on the design of data center networks, and it enables routers and switches to support 400G links from 1m to 600km+ using copper cables, client optics, or coherent optics in a single form factor.

KEY TECHNOLOGY ENABLERS AND DRIVERS FOR 800G

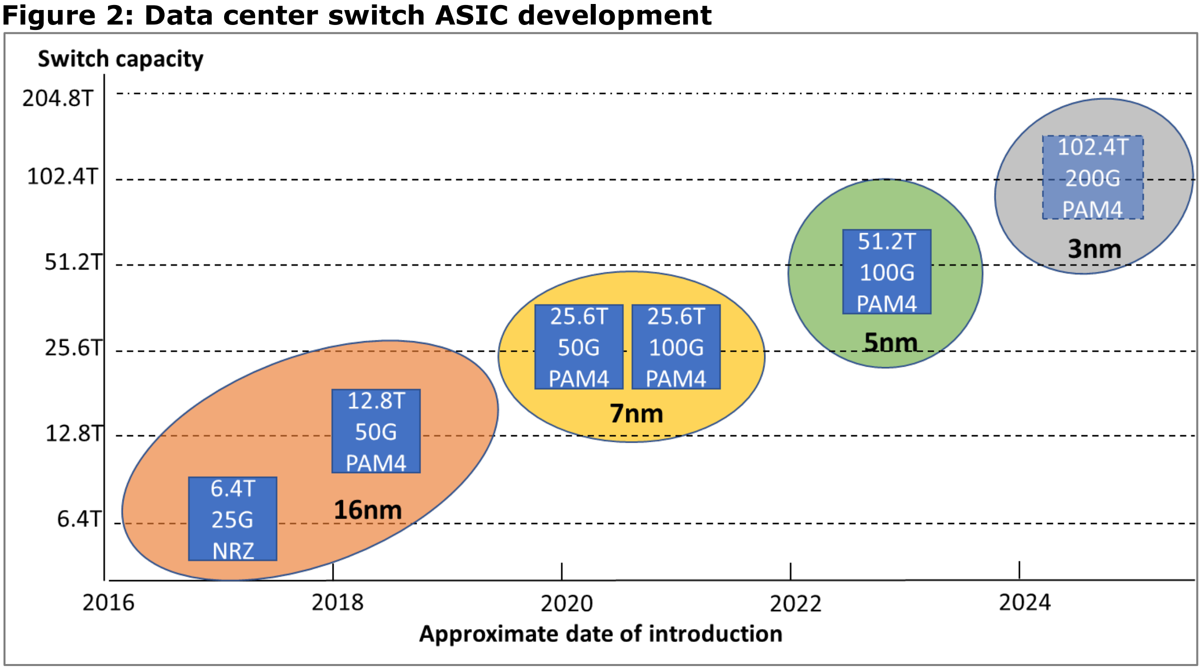
PAM4 modulation has proved to be a very cost-effective way of doubling the effective data rate of a single lane without increasing the baud rate (number of symbols per second). Previous generations used non-return-to-zero (NRZ) modulation, which has two signal levels, and each symbol represents one bit of data (0 or 1). PAM4 has four signal levels, and each symbol represents two bits of data (00, 01, 11, or 10). The drawback to PAM4 signaling is the increased sensitivity to noise, which means that other techniques are necessary to increase the signal robustness by using FEC and equalization that had not been necessary at lower speeds.

The first PAM4 implementations were for 50Gbps lanes, and they rapidly displaced the 50Gbps NRZ solutions in development at the time. 50G PAM4 (56Gbps maximum bit rate) is now well established and supported by a wide range of switch/router ASICs and optical modules. 50G PAM4 enabled the first high volume generation of 400G client optical modules using QSFP-DD/OSFP and the deployment of 200GE connections in data centers using QSFP56 optical modules.

100G PAM4 is also well established now, enabling more cost-effective 100GE using one wavelength (single lambda) and 400GE using four wavelengths for single mode fiber (SMF) links. The availability of 25.6T switches and routers with 100G PAM4 interfaces will enable a quick ramp-up for 800G ports.

The next generation of solutions currently being developed will support 200G PAM4 wavelengths, reducing the optical complexity and cost of the modules, and enabling greater switch and router system capacity and 3.2Tbps ports. This will make 800G more cost-effective as the industry standards converge on the coding, equalization, and error correction techniques necessary for the variety of applications to be supported.

Figure 2: Data center switch ASIC development



Source: Earlswood Marketing

Another key enabling technology is silicon photonics, which is the integration of photonics components and high speed transceiver functionality onto a silicon substrate. This technology has been widely used for 100G and 400G optical modules. Silicon photonics enables the use of standard wafer fabrication plants for high volume fabrication of optical subsystems to simplify and reduce the cost of optical module assembly. This integration approach is particularly important for dense optical modules with eight or more lanes and coherent optical modules with complex optical functionality.

As switch capacity grows and serial interfaces become faster, further optical integration becomes even more important. Early work is happening on switch implementations using co-packaged optics. In this case, silicon photonics chiplets are co-packaged with the switch ASIC, potentially removing the need for optical modules plugged into the front panel. At 102.4T and above, it is expected that co-packaged optics will become increasingly important.

The Ethernet standards for 800G and 1.6T are in development by the IEEE802.3df Task Force, which started work in January 2022. The primary objectives defined are to cover the following:

- 800 Gigabit Ethernet using eight 100G lanes or four 200G lanes
- 1.6 Terabit Ethernet using eight 200G lanes
- 200 Gigabit Ethernet using a single 200G lane
- 400 Gigabit Ethernet using two 200G lanes

The specifications will cover single mode fiber (SMF), multimode fiber (MMF), copper twin-axial cables, and chip-to-module electrical interfaces.

At the September 2022 meeting, the IEEE802.3df Task Force voted to split the work into two projects:

- IEEE802.3df working on specifications based on 100Gbps lanes
- IEEE802.3dj working on specifications based on 200Gbps lanes

This should allow the completion date for 800GE using eight lanes (IEEE802.3df) to be brought forward by more than a year to June 2024. The proposed completion date for implementations based on 200G lanes (IEEE802.3dj) would be pushed out, allowing more time for this work; however, baseline proposals are already maturing. These changes to the objectives and schedules are subject to full approval by the IEEE.

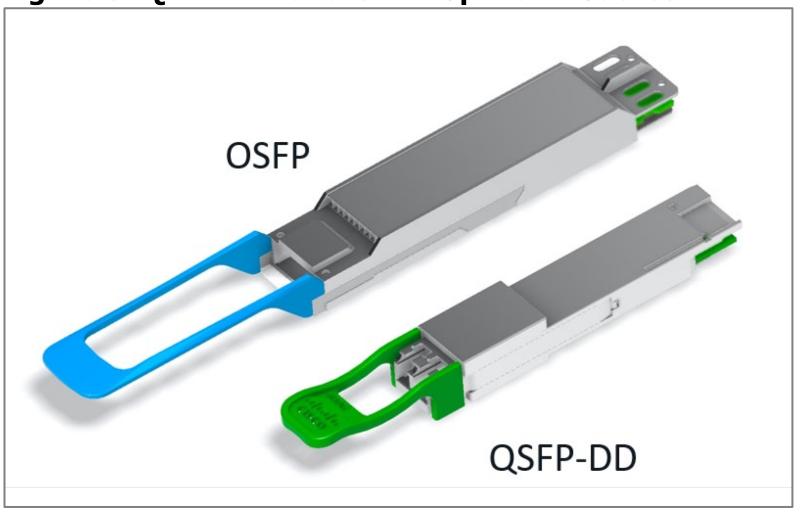
The OIF is running a new project to define 800G coherent line interfaces for two applications. 800ZR is for amplified, single-span WDM links up to 80−120km and will be a direct upgrade for 400ZR. The project is also working on a solution for unamplified, fixed wavelength links up to 10km that could be used in a variety of applications that currently use direct-detect optical modules. The IEEE 802.3dj also has 800G objectives for 10km and 40km reach. Coherent interfaces will be adopted for the 40km solution and are being proposed as a candidate for the 10km solution.

OPTICAL MODULE DEVELOPMENTS

Optical module developments have been driven by several key considerations, including physical size, number of lanes, data rate per lane, backward compatibility, and thermal management, including heat sinks and airflow.

Two types of double-density optical module form factors were developed for 400G client optics applications with eight 50G PAM4 lanes: QSFP-DD and OSFP (see **Figure 3**). Both optical module form factors, with subsequent enhancements including 100G PAM4 lanes, will support 32 400G or 800G ports in a 1RU system and thermal cooling capabilities up to 30W.

Figure 3: QSFP-DD and OSFP optical modules



Source: Cisco

The QSFP-DD form factor was developed by the QSFP-DD MSA and is the latest in a long line of modules developed from the small form-factor pluggable (SFP) modules used for 100 megabit Ethernet and Gigabit Ethernet. The SFP specification was enhanced to support 10GE (SFP+) and expanded to support 40GE with four 10Gbps lanes (QSFP+). Further developments supported 100GE with four 25Gbps lanes (QSFP28) and 200GE with four 50Gbps lanes (QSFP56).

The QSFP-DD form factor adds a second, deeper row of contact pads along the module's edge connector. This doubles the number of signal and power connections to support the eight lanes and increased power requirements, but it also allows the insertion of the existing four lane modules into the same connector. With the connector able to support backward compatibility, the cages also need to physically have the same dimensions so that older QSFP modules can be inserted. There are two basic designs for a cage: the surface mount (SMT) cage supports a single module and the 2x1 stacked cage supports two modules, one above the other. With the push to higher speeds and complexities, these 400G and above modules are much higher power than the ~3.5–5W QSFP28 modules. To support these increases, the cages are designed to include heat sinks that the module slides under to aid with heat dissipation.

With this riding heat sink approach to addressing cooling, it was found that QSFP-DD was able to meet all system and module thermal requirements for any module at 400G or 800G speeds. Key to this ability is the fact that the heat sink is not part of the module and therefore can be optimized for the system design and requirements. Heat sinks can be added on top of the module (or between the modules in the case of a stacked cage). These can be independently sized and can use better quality materials. Heat sinks can also be added to the nose of the module outside the faceplate to take advantage of the airflow there. (This is known as a Type 2 variant QSFP-DD.) This flexibility is key to designing switch and routing systems that optimize airflow and power efficiency.

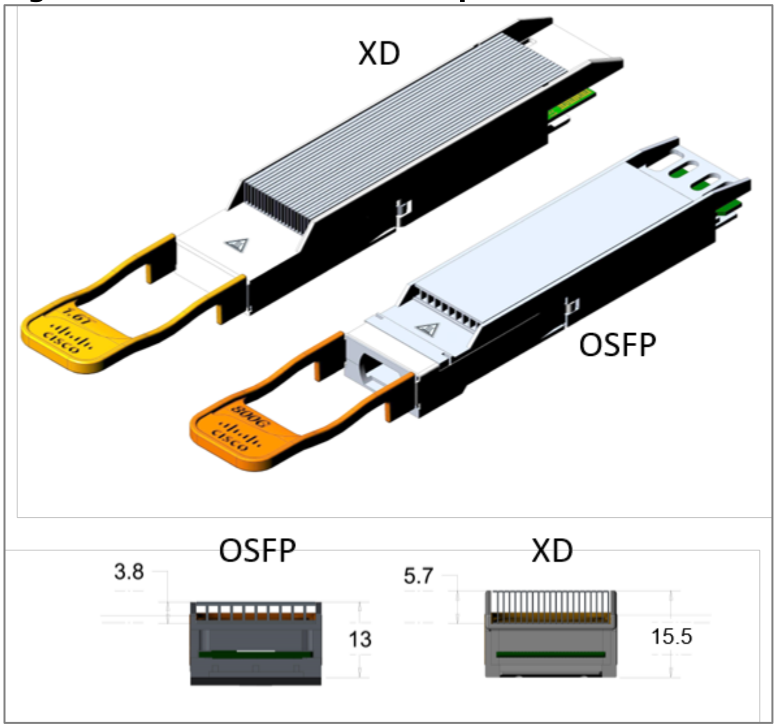
The QSFP-DD specification was expanded in 2021 to support 800G applications, using 100G PAM4 lanes (QSFP-DD800), and increased maximum module power dissipation, with products supporting 30W modules already being demonstrated. The QSFP-DD MSA is currently working on the next-generation QSFP-DD to include support for 1.6T with 200G lanes (QSFP-DD 1600).

The OSFP module specification was developed by the OSFP MSA. The basic OSFP module supports eight 50G PAM4 lanes and has an integral heat sink and plugs into a cage that supports one or four modules side by side. The OSFP specifications do include an alternative module option without the integrated heat sink. The OSFP Riding Heat Sink (OSFP-RHS) has the heat sink mounted on the cage and the modules are lower, making them incompatible with other OSFP solutions.

The OSFP specification was expanded in 2021 to include support for 800G modules with 100G PAM4 lanes (OSFP800) and increased module power support to support a maximum of approximately 30W per module. The OSFP specification was further expanded in 2022 to include support for 1.6T modules with 200G PAM4 lanes (OSFP1600).

The OSFP MSA group has now started working on a new module form factor, OSFP-XD, that is designed to support future applications requiring 16 electrical lanes or higher power consumption of up to 40W, including coherent pluggable modules. OSFP-XD modules (see **Figure 4**) integrate an enhanced heat sink that is 20% taller than OSFP modules and is not expected to be backward compatible due to optimizations needed to support the higher power and higher speed lanes. 1.6T OSFP-XD modules with 100G PAM4 lanes could be compatible with future 3.2T modules using 200G PAM4 lanes.

Figure 4: OSFP-XD and OSFP optical modules



Source: Cisco

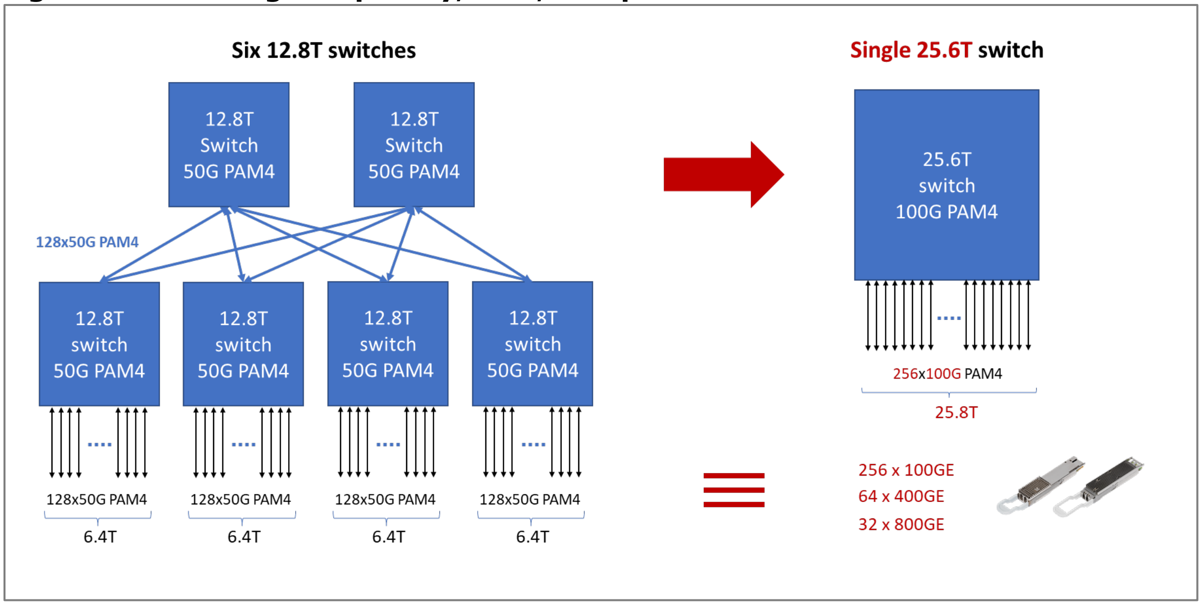
The OSFP-XD specification is expected to be released before the end of 2022. The OSFP-XD module electrical interface uses the double-density approach adopted by QSFP-DD to provide two contact rows, allowing support for 16-lane modules. OSFP-XD is being designed to be flexible to support different applications and system cooling challenges. The extra height allows a tradeoff between a module that supports 16x passive/active copper cable (DAC/AEC) solutions, which need more room but consume less power, and a higher power optical module that can benefit from the maximum heat sink fin height. Balancing these applications so the system airflow and thermal management can support any intermix of module types was a key target for OSFP-XD. To achieve this extra bypass, air channels are being defined to support the cooling of the hot ASIC behind the optics.

800G USE CASES

The introduction of 25.6T switches with 800G ports creates new opportunities for data center operators to take advantage of use cases that can deliver significantly higher performance networks with reduced complexity, cost, and power.

As shown in **Figure 5**, a single 25.6T switch can be used to replace six 12.8T switches. Each of the 800G ports on the 25.6T switch has eight 100G PAM4 lanes and will support eight 100GE connections, two 400GE connections, or (when the specifications are complete) a single 800GE interface. By using 100G PAM4 serial interfaces, the number of signal lanes on the 25.6T switch is the same as each of the six 12.8T switches, reducing the overall complexity of the solution. The replacement of six switches with a single switch will deliver a significant reduction in power and cost for the same 25.6T network capacity.

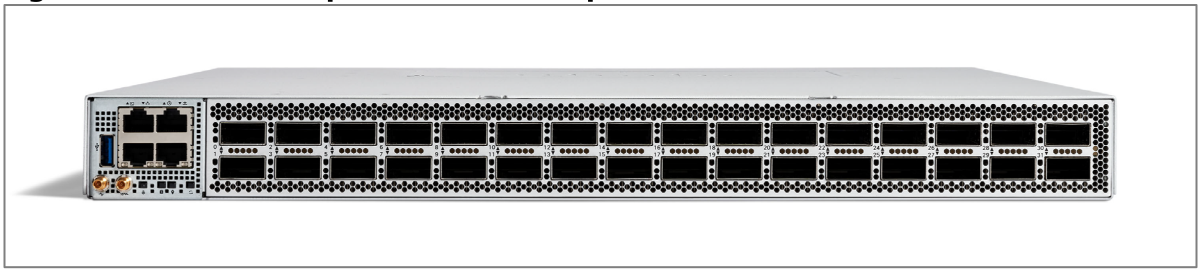
Figure 5: Reducing complexity, cost, and power with 25.6T switch



Source: Earlswood Marketing

Figure 6 shows the Cisco 25.6T platform with 800G ports. Both the Cisco 8111-32EH and the Nexus 9232E have 32 800G ports in a compact 1RU form factor and will support 64 400GE links or 256 100GE links using breakout cables. Heavy Reading explores several use cases that utilize this flexibility below.

Figure 6: Cisco 25.6T platform with 32 ports of 800G

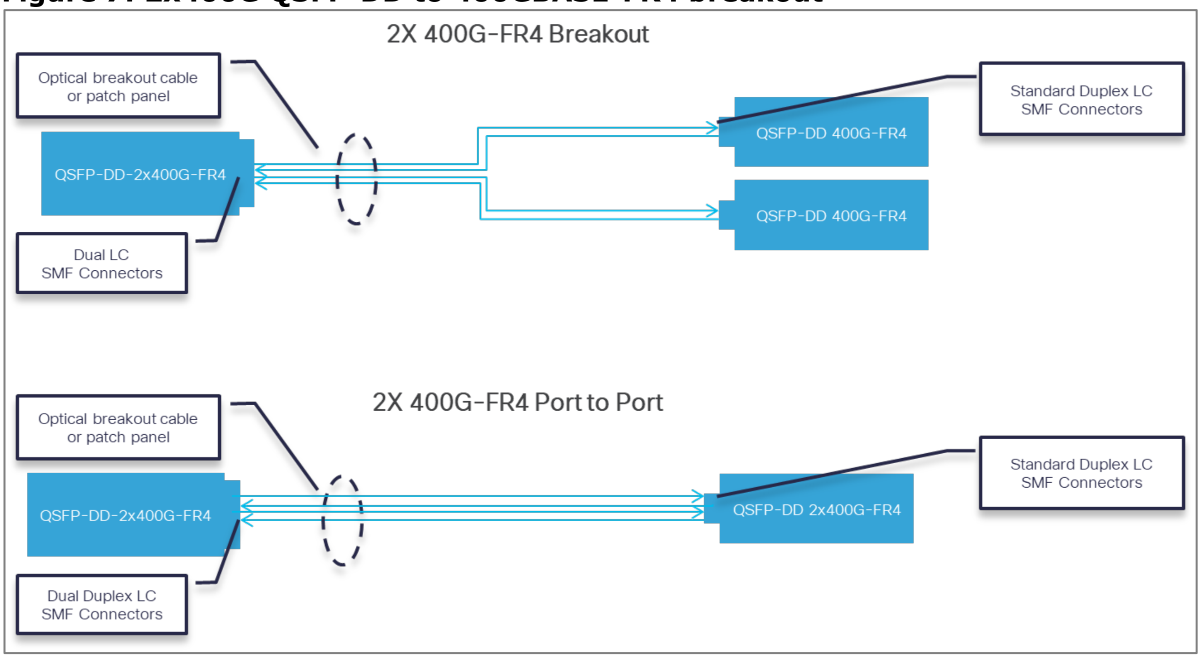


Source: Cisco

Figure 7 shows an 800G port used to support two IEEE400GBASE-FR4 connections up to 2km, offering high density 400G interfaces. The upper configuration using a breakout cable directly doubles the number of 400GE links supported by a 1RU 25.6T switch system compared with 12.8T systems. Deployments using this general configuration are moving forward in multiple applications, including high density AI/ML clusters in data centers and ultra-high definition video processing.

The lower configuration implements an 800G connection up to 2km using dual IEEE400GBASE-FR4 links. Network deployments using this configuration are being rolled out for high performance computing (HPC) systems supporting scientific research and other very high processing applications.

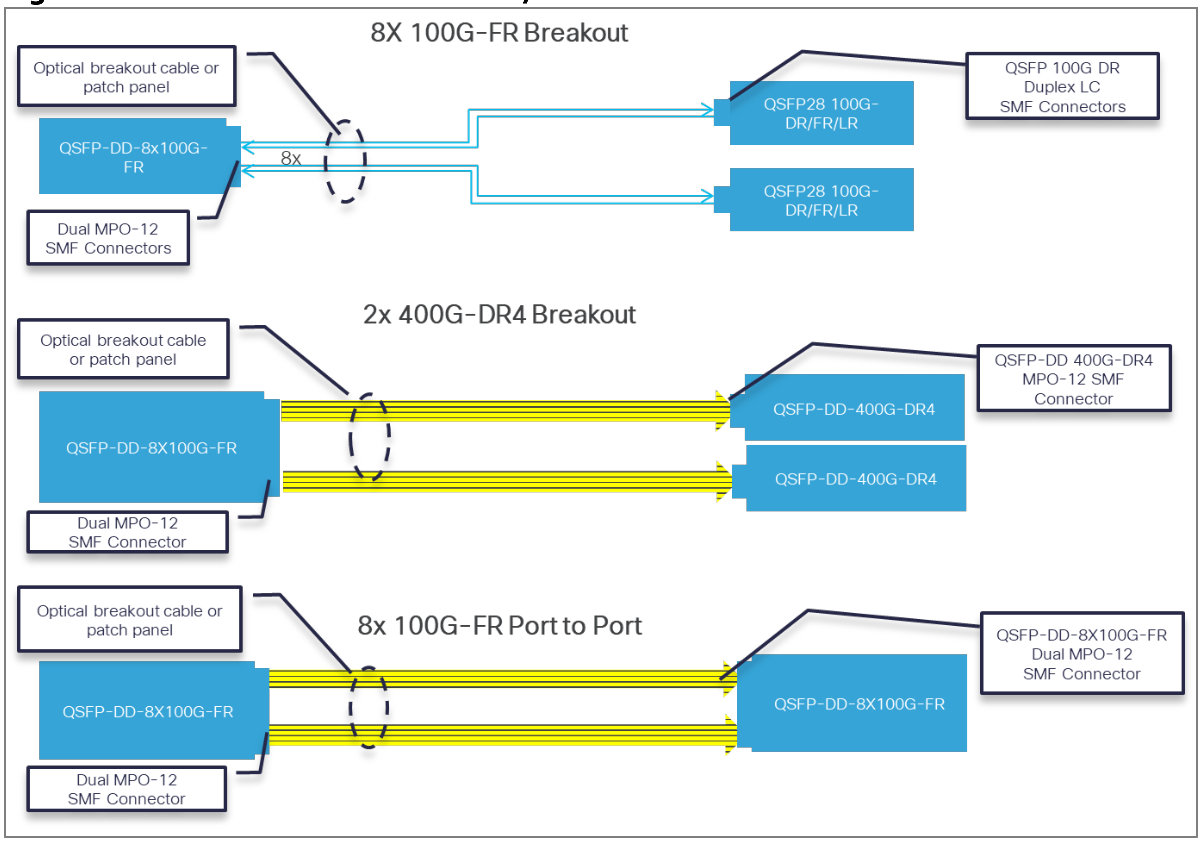
Figure 7: 2x400G QSFP-DD to 400GBASE-FR4 breakout



Source: Cisco

Figure 8 shows an 800G port used to support eight IEEE100GBASE-DR (up to 500m) or IEEE100GBASE-FR (up to 2km) connections, offering high density 100G interfaces. The figure shows three configurations: a breakout to eight separate 100GBASE-FR links, a breakout to two 400G connections, and an 800G connection over eight 100G links. These configurations are ideal for upgrades in peering/colocation networks and distributed data centers, which need many 100GE network connections.

Figure 8: 8x100G to 100GBASE-DR/FR breakout



Source: Cisco

CONCLUSIONS

The 400G upgrade for client optics was delivered faster than any previous generation and has enabled hyperscale data center operators to deliver enhanced cloud services to their customers. PAM4 modulation with FEC and double-density optical modules were game changers, enabling a quadrupling in bandwidth with only a doubling of signal bit rate for optics and links to switch ASICs. The shift to higher capacity, 400G, ports, and the use of breakouts enabled the industry to move to higher capacity switches without upgrading all the networking components in the data center. The rapid deployment and the lessons learned from 400G show that the industry has evolved and is now ready to deliver on 800G.

The first switches with 800G ports are available, so service providers can already move ahead with evaluating the opportunities and planning deployments. 800G client optical modules have been introduced supporting two 400GE connections or eight 100GE connections, and 800GE versions will be introduced as the standards develop. The technology developments made for 400G are easing the shift to 800G, enabling a staged approach using double-density optical modules and breakouts.

The introduction of 800G switch ports, optical modules, and DACs provides a significant opportunity for service providers to upgrade network performance without waiting for the 800GE standards. The work in the IEEE 802.3df Task Force for 800G and 1.6T, together with industry technology developments, will ensure a strong roadmap for future upgrades. The rapid deployment of 400G shows that the industry can deliver the latest solutions quickly, and it is ready to support an equally rapid shift to 800G as service providers build out and upgrade their data centers.

APPENDIX

About Cisco

Cisco is the worldwide leader in technology that powers the internet. Cisco inspires new possibilities by reimagining your applications, securing your enterprise, transforming your infrastructure, and empowering your teams for a global and inclusive future.

Cisco offers data center operators the flexibility to choose from a range of form factors, speeds, client optics, and network operating systems to achieve high performance, port density, and power efficiency for hyperscale architectures. To learn more about Cisco's new high-density QSFP-DD800 optical transceivers, please visit the Cisco Optics webpage.