How to Accelerate 400-Gigabit Ethernet Rollout

Optical form factor convergence can support rapid bandwidth expansion

The impact of optics costs on purchases

Internet traffic has seen a compounded annual growth rate of 30 percent or higher over the last five years because more devices are connected and more content is being consumed. According to the Cisco Visual Networking Index, by 2022:

- IP traffic will increase to 396 exabytes per month.
- More than 28 billion devices and connections will be online.
- 48 percent of all connections will be video-capable.[1]

As a result of this growth, the networking and interconnect industries are pushing their technology development teams to meet bandwidth connectivity demands. The next major transition will be the general availability of 400-Gigabit Ethernet (GbE) technology. Hardware vendors are already releasing platforms to support this growth.

The 400 GbE will be mainstream soon enough, but the industry has yet to standardize on one pluggable module form factor, which is causing duplication in both development and manufacturing. This duplication creates risk that affects the market's ability to scale around a common solution and drive the necessary cost reductions that come from that scale.

The cost of optics may become a barrier to transition. The reduction of application-specific integrated circuit (ASIC) cost-per-bit is outpacing that of optics. Delays in choosing a common optical platform could affect the market adoption of 400-GbE platforms, which could lead to delays in growing the bandwidth market. Quad Small Form Factor Pluggable-Double Density (QSFP-DD) modules are an ideal connectivity option for all next-generation cloud infrastructures, hyperscale data centers, telecom operators, and Internet Service Providers (ISPs).

Contents

Cost as a barrier to adoption

In recent generations, the high-volume optical form factors for a new speed have always been the same as for the previous version. In other words, the optical form factors are backwardly compatible to the previous lower speed. Backward-compatibility ensures that service providers can spend their dollars wisely using the latest technology on new and installed infrastructure. They also can reuse modules if they want to protect their previous investments.

Significant costs are associated with developing any new technology, and new optical module factors are no exception. Each change in form factor delays the ability for manufacturers to reach economies of scale from mass production. Delaying a decision on the industry standard can hold costs artificially high because manufactures have to make limited quantities of multiple iterations and can't mass produce them to bring component costs down. However, getting to a common industry standard accelerates mass production and lowers costs by achieving economies of scale.

The transition to 400 GbE is unique. For the first time, the integration of components and technology is making it possible to integrate all the interconnect reach variants. Everything from copper to long-haul Dense Wavelength Division Multiplexing (DWDM) technology can be integrated into the same high-density pluggable form factor immediately when the new Ethernet speed is deployed. The industry can bypass the many years of product development related to working through various larger, transitional, form factors as the technology matures. Converging on a common industry-standard will significantly accelerate cost reductions and increase adoption.

The unfortunate duplication of module formats for 400 GbE prevents the market from taking full advantage of the other aspects of industry alignment with both the QSFP-DD and octal small form factor pluggable (OSFP) modules being available. Streamlining manufacturing is crucial to unlocking the economies of scale. Volume drives yield and cost. A common production line is key, and a vast ecosystem with dozens of companies benefits from standardization. The companies may include developers of manufacturing equipment, test equipment, software design tools, connectors and cages, thermal solutions, compliance and qualification equipment, and many more. Reaching this common production line quickly is crucial given the anticipated early and rapid growth of 400 GbE.

Network and system requirements, such as density, drive optical module requirements, and optical component integration capabilities drive optical module feasibility. In the past, these requirements have led to initial larger form factors that transition to smaller, denser, and less expensive form factors as the technology moves towards higher adoption. This cycle has happened before.

For example, 10-GbE optical modules moved from the bulky 300-pin module, to XENPAK, to X2, to XFP and finally to SFP+. Similarly, 100-GbE modules transitioned from the initial CFP form factor, to CPAK or CFP2, to CFP4, and finally to QSFP28. With Fibre Channel, there was no deviation from SFP throughout the speed increases from 1G to 2G to 4G to 8G to 16G to 32G because the use cases didn't change.

With Ethernet, these transitional form factors occurred as the dominant use case changed over time. Driven by service providers initially, the use cases required longer reach and lower density before they transitioned to include the data center requirements for shorter reach, higher density, and higher volume. The longer reach technologies miniaturized in time to fit into the higher volume form factor to provide a complete solution. In either case, the lowest cost points and most frictionless development paths always won. Many reasons exist for the industry to repeatedly converge toward the same form factor.

A network architecture also needs to stick with same switch port density per rack unit to maintain proven fabric designs because networking equipment is designed around them to achieve certain densities. Maintaining this consistency limits the impact on the network ecosystem in that location. For example, if the switch port density changes from 16 down to 8, a location needs other infrastructure pieces added to it to maintain the same connectivity levels or total system throughput.

Significant experience and reuse benefits both system and module suppliers; it helps minimize costs and accelerate development. And finally, the backward compatibility use-case advantage of using the same form factor is significant because it enables multi-speed port options on switches. The slower optics in higher speed ports are often required for seamlessly migrating to higher port speeds.

Adoption timeframes cut in half

Shifting market dynamics are behind the market urgency for the general availability of 400 GbE. With the introduction of 10 GbE, early adopters deployed before high-volume data centers. When 100 GbE was introduced, again early service provider adopters deployed ahead of high-volume data center markets even though market demands required the difference in adoption cycles to be slashed in half.

As the number of bandwidth-hungry apps and workloads explode, adoption of 400-GbE technology in key environments such as cloud providers, telcos, and hyperscale data centers may happen almost simultaneously. According to projections for optical modules from LightCounting, the five-year adoption of 400GbE is expected to be 20 times faster than 100GbE (see Figure 1).[2] This opportunity creates an unprecedented industry investment by both established players and startups.

The market demand requires vendors to find the optimal solution to meet all those needs now. They can't wait for multiple form factor iterations to get to the lowest cost points. The QSFP-DD form factor is being adopted by the industry as the high-volume form factor for 400 GbE. It can achieve the multiple goals of technical feasibility, backwards compatibility, and breadth of reach support.

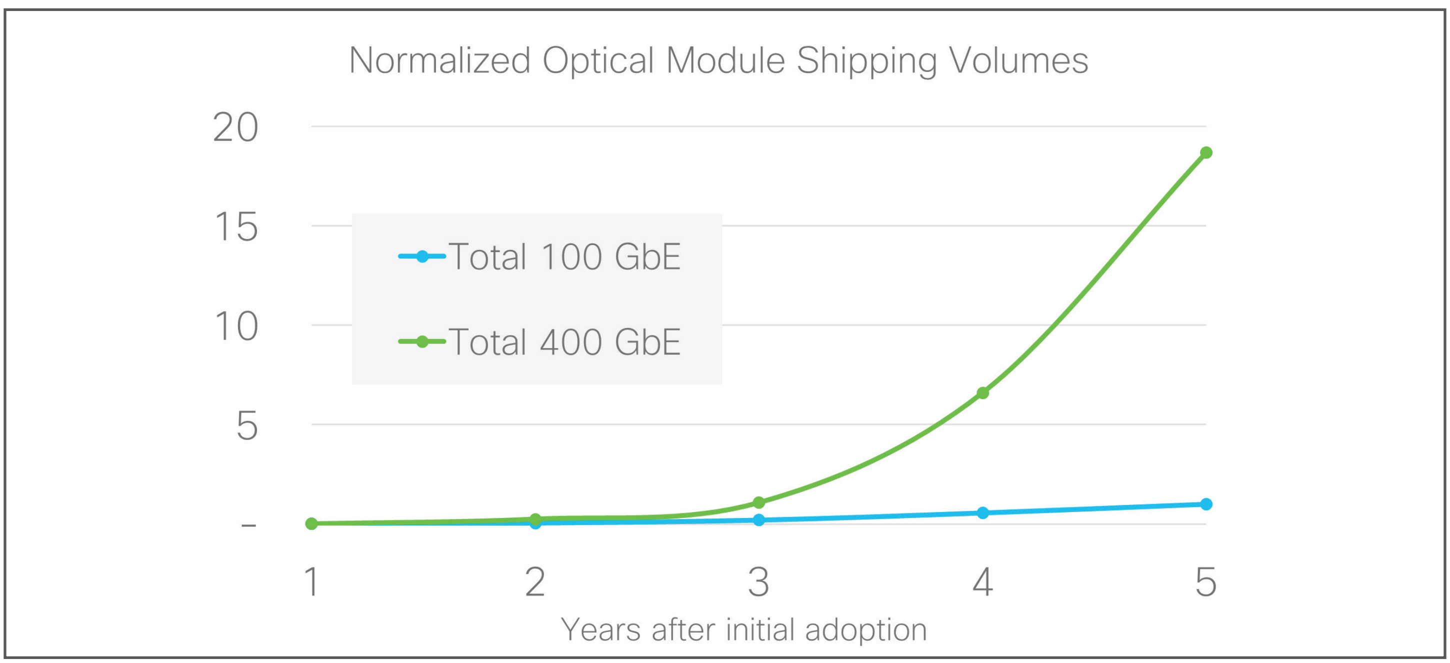
QSFP-DD is ready for current and future market dynamics

To establish itself as the dominant and successful pluggable form factor, QSFP-DD has been designed to meet all of the market needs and optimized system designs for 400 GbE. The beauty of the QSFP form factor is that the industry has had more than a decade of experience in building solutions around this form factor. This background enables rapid innovation and rapid maturity. Any form factor faces three key challenges. Can the electrical connector work at this higher speed? Can the system cool the module? And can the optics fit? Although this paper only discusses the thermal aspect, QSFP-DD successfully meets all three challenges.

Over the past decade, the ratio of cost between the hardware platforms and the optics has shifted substantially towards the optics. This trend will only increase with 400 GbE, but generational compatibility will help offset the impact of this trend. Over 24 million QSFP modules that represent an $8 billion investment will be deployed by the end of 2019. Even as 400 GbE rolls out, QSFP28 100 GbE will continue strong growth that is driven by the emergence of 100-GbE servers and the increase in bandwidth throughout enterprise and service provider networks.

Adding new equipment and running the same network faster isn't sufficient. You need to consider multiple aspects of backward compatibility, including reusing existing modules and the continued investment in 100 GbE. QSFP-DD leads the industry by being backwards compatible with 100-GbE QSFP28 and even 40-GbE QSFP+. Backwards compatibility should be maintained whenever possible because extending compatibility across three or four generations allows service providers and enterprises to maintain a positive return on investment (ROI) for installed platforms.

Figure 1. Comparing the first five years of 100GE vs 400GE. LightCounting projection for 400GE module volume



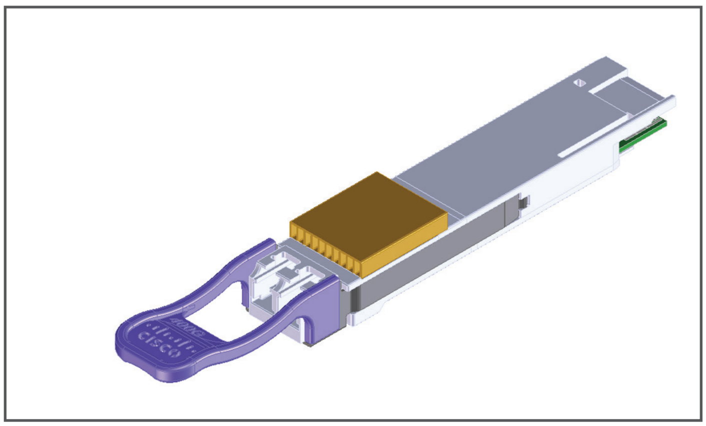
If hardware and optics vendors expect service providers or enterprises to be willing to adopt new equipment earlier, they need to protect ROI. Hardware and optics vendors must ensure compatibility with the cooling architecture, for example, top to bottom or side to side. QSFP-DD has the flexibility of separating the module and the heat sink, which enables equipment providers to build solutions that can use the newer 400-GbE technology for its needs regardless of the installed infrastructure.

Achieving backward compatibility with QSFP-DD required solving a wide range of challenges, including component size and layout, module and system cooling, and electrical connectors that support 4 and 8 channels using 56G serial-deserializers (SerDes). These factors are tightly coupled and need to be considered along with other system components such as high-power ASICs. Naturally, these challenges can be reduced in a unique module design (such as OSFP) except that it abandons compatibility with earlier generations. The market importance of achieving backwards compatibility led many companies to embrace the innovation challenge around QSFP-DD.

Prepare for power dissipation requirements that exceed 20 W

Thermal dissipation is one of the most visible technical challenges in 400 GbE. The initial 400-GbE modules were anticipated to require 12 W to support many of the client optics. Given that QSFP28 only supports approximately 4 W, some people considered 12 W a step too far. By applying a cumulative decade of experience building systems and solutions around QSFP modules, the industry was able to focus their innovation to meet the market needs. Through advanced module cage design, heatsink design, and system mechanical and thermal design, advances were made to achieve the cooling of a 15-W modules, fully populated in the most challenging system thermal designs. The success in exceeding the initial goals led to even more ambitious outcomes. The 400ZR/ZR+ coherent modules planned for 2020 has power targets that could go up to 20 W. Building on the knowledge and innovation for the 15-W solutions, further innovation has made cooling 20 W possible, and standards groups will soon ratify these solutions for QSFP-DD. This step has been accomplished by adding an integrated heat sink onto the nose of the module as shown in Figure 2. Thermal innovations have already been adopted for cage and heat sink designs within the systems.

Figure 2. QSFP-DD module showing the integrated heatsink on the nose to enable 20-W system cooling for ZR+



As demonstrated at the Optical Fiber Communication Conference and Exhibition (OFC) 2019, this enhanced module design enabled QSFP-DD modules operating at 20 W to be effectively cooled by a significant margin (see Figure 3).

Another thermal consideration is that optics modules can't be viewed as a closed system. They must work within the overall design of the router, switch, or server. A valuable characteristic of QSFP modules over OSFP is that their smaller footprint allows for greater air intake. It benefits the rest of the system, which can be clearly seen in platforms that offer both options.

In many other areas, the work to achieve investment protection in QSFP-DD required a massive, industry-wide collaborative effort. Significant technical advances, many considered impossible at first, have been made at each step along the journey from 40 GbE to 400 GbE. Work has already begun to address these challenges for future Ethernet speeds. Repeating the argument that QSFP has reached its limits should be viewed with skepticism. History has demonstrated the ability to overcome previous perceived limitations.

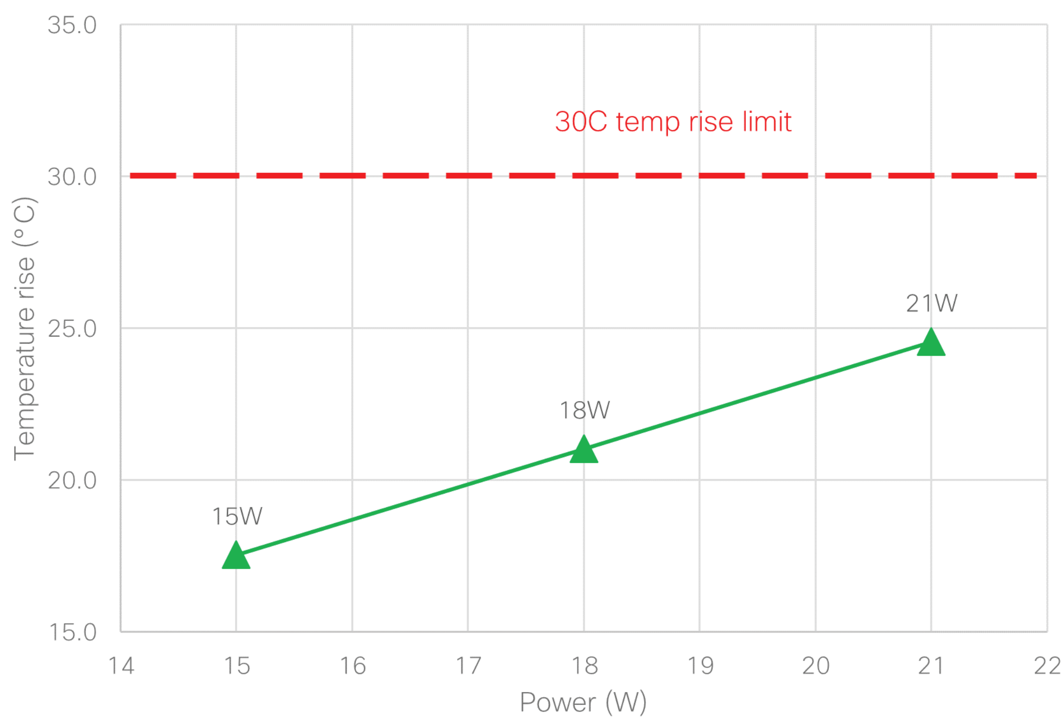
Ensure a long and useful life

The current landscape of double-digit Internet bandwidth growth is straining the economics of network infrastructures. As you budget for expansion, you need to maintain a positive ROI and ensure the longevity and useful life of the hardware and components. The QSFP-DD pluggable form factors are built to accommodate future bandwidth expansions, but more importantly, they provide backward compatibility to ensure interoperability with your installed platforms.

Learn more

To learn more about Cisco infrastructure products, please visit the Mass Scale Networking page.

Figure 3. Measured QSFP-DD module case temperature rise versus module power dissipation. Maintaining a case temperature rise below 30 degrees Celsius enables operation in any network environment.



1. Cisco Visual Networking Index, 2018. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html
2. LightCounting. "Ethernet Forecast Report." Sept 2018.