White Paper

Getting the Most out of 400GE

Data Center Operators (DCOs) are in hyper-growth mode to keep up with demand for social media applications and cloud computing. This growth has pushed DCOs to add storage and transport capacity at an unprecedented rate, and to bolster Data Center Interconnect (DCI) capacity to alleviate the data crush.

Rather than deal with the complexity of DWDM, many DCOs have opted to simply run Ethernet and fiber between two locations, albeit Ethernet at 400Gps. However, 400GE is not just another line rate. To achieve such high transmission rates, the 400GE standard has some unique features that must be considered when testing and troubleshooting.

Originally known as IEEE 802.3bs, 400GE was officially approved in December of 2017 and is part of a broader family of related themes such as 200G, next generation 100G, and 50G Ethernet. To achieve such high transmission rates, 400GE relies on more sophisticated optics, more complex modulation, and forward error correction (FEC), which in turn result in higher power requirements and residual heat. These five factors must be considered when testing 400G links and choosing 400G test equipment.

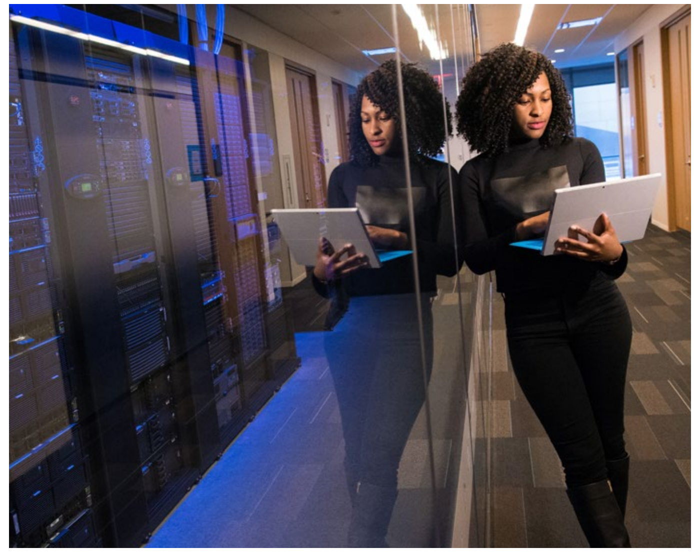
Pluggable Optics

When the industry moved from 10Gbps pluggable optics to 100Gbps, it meant shifting from one lane and laser to multiple lanes and lasers, for example from SFP to QSFP. More lasers mean higher performance, but with more complexity and a higher failure rate as a tradeoff. The increased complexity and failure rates required network users and operators to test the multi-lane optics in the field especially after traditional throughput and latency tests like Y.1564 failed.

Now, as the industry moves from 100Gbps pluggable modules to 400Gbps, the complexity is again ratcheted up, with more components in the same physical footprint. The QSFP-DD, for example, is a marvel of sophistication, with integrated lasers and drivers, high performance photodiodes and microcontrollers integrated into a very small form factor. With so much data traversing these critical and complex network elements, testing pluggable optics is a must, especially to help isolate a fault.

QSFP-DD vs. OSFP

Currently, the two pluggable optics module types for data center access and interconnect use cases are QSFP-DD and OSFP. Of the two, QSFP-DD has been built with backwards-compatibility in mind. If a 400G switch supports QSFP-DD, the ports on that switch can support QSFP+, QSFP28, and QSFP56. In other words, a 100G or 40G optical module may plug into 400G QSFP-DD port and pass traffic through it. Data center operators like that backwards-compatibility which gives them flexibility and investment protection. QSFP-DD has broad industry support and has a greater share of shipped ports.



OSFP modules are also available, and the form factor is being used by some cloud giants. Because 400G requires more power than 100G, OSFP can draw up to 15W. Because of the greater power draw, the OSFP form factor is slightly wider and deeper that QSFP and includes a built-in heat sink. OSFP ports require special adapters is for backwards compatibility with 100G QSFP. However, many hyperscale ICPs are not looking backward, but forwards, to 800GE (for which standards do not yet exist). While QSFP-DD is backwards compatible with 100G, its physical design provides less growth potential to higher rates, namely 800Gbps. OSFP on the other hand, is 800GE-ready, and although 800GE infrastructure is not available, some industry giants are already betting on it with their pluggable module strategy.

Another significant change introduced at 400G (technically 200G) and beyond for pluggable optics is the introduction of CMIS which stands for Common Management Interface Specification. One of the key purposes is to serve the growing number of applications supported by such optics. CMIS is a requirement to help manage the growing number of lanes, FEC types, and modulation types in optics; it is also used to manage breakout applications.

PAM4 Modulation

Another 400GE feature that separates it from 100G is its modulation scheme, both electrically and optically for most client interfaces. 400G Ethernet is so fast that it has outpaced the ability of a conventional laser on/laser off binary modulation scheme to keep up. To compensate, PAM4 modulation has been developed, which utilizes four amplitude levels rather than two in order to double the overall bit rate. Since the gap between signal levels is now much smaller, PAM4 is more susceptible to noise. To state that differently, the physical layer on a 400GE link will always have errors and it's important to expect that and know what level or errors is acceptable. As a result, understanding and controlling the error rate at the physical level grows in importance.

Forward Error Correction (FEC)

400GE also uses FEC technology. While FEC has been widely used in OTN technology for many years, its use in general-purpose client interfaces is relatively new. In fact, we now have the expectation that PAM4-based links will run with an error floor (potentially as high as 10^-4) with the FEC layer correcting this to give an effective error-free link at packet level.

The increased speeds and use of FEC technology mean some modules with higher raw error rates will operate error-free post-FEC and others will not. A more sophisticated understanding of the error distribution and statistics is required to determine acceptable from unacceptable error patterns and determine true root causes. For example, FEC works in 5440-bit blocks broken into 10-bit symbols. More than 15 symbol errors in 1 block is considered unacceptable as this results in uncorrectable errors with dramatic effects. As 400GE has a heavy reliance on error correction, it is key to understand errors both at the pre-FEC and post-FEC levels. This is especially important when testing pluggable optics, that is QSFP-DD or OSFP, and understand both pre-FEC and post-FEC error levels. That is because this is what the fiber infrastructure and the Ethernet equipment will experience respectively.

Power

400G infrastructure will draw more power than 100G. However, in most data centers, the real heat and power issues come from the servers, not the top of rack switches or your leaf/spine boxes. As data center operators plan migrations to 400G, they must be aware of the increased power draw as it may or may not have a material effect on power and cooling infrastructure or require significant changes or upgrades.

Power draw for 400G test gear is a more important consideration. Unless a 400G test head is rack-mounted, it will likely require battery power. A portable 400G test unit requires a large power draw to generate 400Gbps of network traffic, which drains battery power quickly. Depending on the type of network throughput test running, a soak time of minutes to hours might be required to achieve the required level of accuracy. A portable test unit that burns through battery power is not actually very portable. Therefore, battery life can be a serious concern when choosing 400GE test gear.

Heat

Heat dissipation has always been an issue for data center operations, and a migration to 400GE will certainly not relieve that. However, greater attention must be paid to the selection of 400GE test units as they tend to run hotter relative to 100GE test gear. The problem is that 400GE optics constantly subjected to high temperatures will suffer more wear and tear, and ultimately have a shorter life span. There is great variation in the heat performance of 400GE test gear across manufacturers, so it is a problem that DCOs can avoid with proper awareness.

In fact, one area where both power and heat are quite important is with pluggable digital coherent optics. New interface types have become available most notably in the form of ZR and ZR+. A key goal with such interfaces is to achieve longer transmission distances via the integration of technology such as advanced modulation, coherent receiver integration, and the use complex FEC algorithms. This enables to more easily interconnect data centers and offers the option of either using unamplified or amplified optical links. As the circuitry in such optics, such as QSFP-DD or OSFP, becomes so dense the availability of ample ventilation and sufficient cooling becomes a requirement to ensure durability. ZR and ZR+ interfaces introduce key requirements like being able to control the transmission frequency and modulation type.

VIAVI Solutions helps you get the most from your cloud/data center ecosystem so you can turn up infrastructure, successfully scale, and profitably innovate — today and tomorrow.

We help you successfully transition to next-generation technologies and implement major transformations with greater confidence. VIAVI solutions enable the most complex and demanding communications and data networks that evolve with networking innovations. To ensure that our solutions can support next-gen networks, we stay ahead of the curve on technology advancements, engaging early with standards bodies, equipment manufacturers, and software vendors. Find out more at viavisolutions.com/hyperscale.

Contact Us **+1 844 GO VIAVI**
(+1 844 468 4284)

To reach the VIAVI office nearest you,
visit viavisolutions.com/contact

getting-the-most-out-of-400GE-wp-xpf-nse-ae
30193291 900 0221