The zettabyte era is here. Is your datacenter ready?

Move to 25GbE/50GbE with confidence

The projected annual traffic for the year 2020 is 15 trillion gigabytes of data. Where does this phenomenal growth in data come from? A self-driving car is estimated to generate 1 gigabyte of data every second. That's 2 million gigabytes of data per car per year. A Virtual Reality headset, for a 360-degree digital experience, uses 20 times the size of an HD-video, which in itself consumes about 11 gigabytes an hour. Self-driving cars, virtual reality and HDR videos are just a few of the data intensive applications which are rapidly becoming part of our every day lives. Exponential growth in cloud-based services resulting from data-intensive applications has increased the demand for ever-higher bandwidth networks. This has led hyper-scale web service providers as well as enterprises to push the boundaries of technology to achieve higher speeds at a lower cost per bit. Workloads in today's datacenter are converging to Ethernet and are surpassing the capacity of existing 10G networks. Datacenter operators must migrate from 10G to higher speed at the compute and storage endpoints to keep up with this demand on capacity.

25G Technology Overview and Industry Standards

The combination of 1 Gigabit/second and 10 Gigabit/second speeds has existed for over a decade and is predominantly used in today's datacenters at the server level. Both 1G and 10G data rates utilize the Small Form Factor Pluggable (SFP) connector at the server and switch level, allowing for backwards compatibility. The SFP form factor offers the smallest size and lowest power, resulting in the highest system density. Cabling infrastructure and physical layout is well understood, and as a result, migration for many datacenters from 1Gbps to 10Gbps has been seamless and cost effective.

Before a 25G data rate existed, the only available option for higher speed server connectivity was to deploy 40G Ethernet. When top-of-rack switches moved to high-density 40G Ethernet, the switches could still be connected to 10G servers using breakout cables, as 40G Ethernet consisted of four lanes of 10G. However, since the underlying technology for 40G Ethernet is simply four lanes at 10G speed, it does not offer the cost per bit, power consumption or server rack density advantages that are the necessary enablers for widespread speed transition. With the emergence of 25G technology, the IEEE has defined 100G Ethernet based on four lanes each at 25G. With the success of 100G Ethernet in datacenter networks for inter-switch links, 25G has proven to be a reliable and cost effective technology. The latest merchant silicon switches deliver up to 32 ports of 100G with single lane serializer-deserializer (serdes) architectures with 128 lanes of 25G, making 25G a logical speed choice for migration of next generation server and storage end points.

Utilizing 25GbE offers significantly higher bandwidth at lower cost per bit and lower power than 40G.

Table 1 shows the total bandwidth achievable with various top-of-rack server speeds. With 100G Ethernet as top-of-rack connecting to 25G servers, the bandwidth increase is 2.5x. For network designs where the use of 100G to 25G breakout cables is not feasible, a 25G top-of-rack switch connecting to 25G servers offers higher speed, lower cost per bit and lower power compared to a 10G server solution.

Table 1: Comparison of 10G, 40G and 25G Servers

Top of Rack Switch	48 x 10G	32 x 40G	32 x 40G	32 x 100G	48 x 25G
Servers	48 x 10G	128 x 10G	32 x 40G	128 x 25G	48 x 25G
Total Bandwidth	0.48 Tbps	1.28 Tbps	1.28 Tbps	3.2 Tbps	1.2 Tbps
Total Switch + Cable Power/Gigabit	X	0.4X	0.25X	0.25X	0.5X
Benefits of 25G	-	-	-	Higher Speed Highest bandwidth Lowest cost/bit Lower power	

To achieve an industry standard and interoperable Ethernet speed, several industry leaders, including Arista, formed the 25G consortium (http://25gethernet.org/). Subsequently, the IEEE formed a working group to define 25G Ethernet, and the specification was officially included in IEEE 802.3by. The IEEE also included Forward Error Correction (FEC) and Auto-negotiation in the specification, which are essential for achieving error-free links in a multi-vendor deployment scenario.

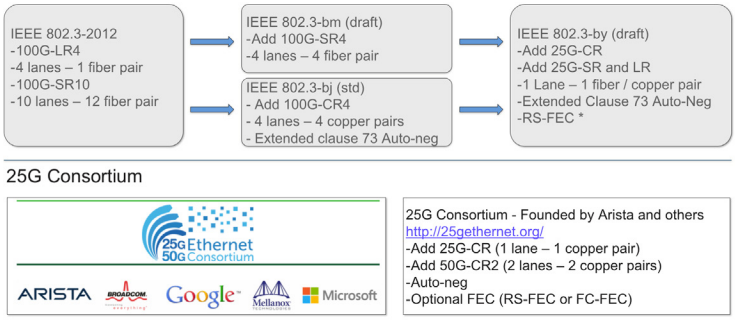

Figure 1: Summary of 25G Ethernet industry standards

25G Familiar SFP Form Factor as 1G and 10G: 25G Cables and Optics

Historically, every new Ethernet speed has gone through multiple pluggable form factor migrations to achieve higher density and lower power consumption goals. For example, 10G pluggable optics started with 1st generation XENPAK optics, which allowed for up to 8 ports per 1U. 10G moved successively to the X2 and XFP form factors before finally converging on the SFP form factor that allows for up to 48 ports per 1U. Similar form factor transitions happened for 40G (CFP to QSFP) and 100G (CFP, CFP2, CFP4 and QSFP), in achieving the highest density and lowest power.

With 25G Ethernet, a new speed was introduced, for the first time, leveraging the smallest widely deployed form factor at the first generation. The use of a single-lane 25G serdes technology allows for the re-use of the well-known 1G and 10G small form factor pluggable (SFP) as the form factor for the physical medium dependent (PMD) devices. Mechanically, the "25G SFP" is similar to both 1G SFP and 10G SFP+, with the exception of an enhanced electrical connector to handle the higher data rate of 25G. The form factor is also referred to as SFP28, due to its capability to handle data rates up to 28 Gbps.

A wide range of 25G copper cables and optics have been defined by both the IEEE and other multi-source agreements (MSA) providing a wide choice of connectivity options for Server to Top of Rack or End of Row (EoR) and for leaf-spine connections. Below is a list of some of the commonly available transceivers and cables for 25G connectivity:

- 25GBASE-CR: Twinax copper direct attach cable (up to 5 meters)
- 25GBASE-SR: Short reach optical transceiver for up to 70m over OM3 and 100m over OM4 multimode fiber
- 25GBASE-AOC: Active Optical Cable (up to 30 meters)
- 25GBASE-LR: Long reach optical transceiver for up to 10km over singlemode fiber

Figure 2: Wide range of 25G pluggable optics and cables

Familiar Infrastructure: Seamless Migration

Implementing 25G at the server offers the benefit of re-using the same server rack designs and works with familiar cabling infrastructure. One of the major hurdles with any speed upgrade is the amount of CAPEX required to upgrade the equipment, re-design the network and refresh the cabling infrastructure. When datacenter operators migrate from 10G based servers to 25G-based servers, the only associated cost is that of upgrading the equipment. The rack design and server capacity and leaf to spine fiber-cabling infrastructure can remain exactly the same. If the current design is a 32x40G leaf switch with 10G servers connected via 40G to 4x10G break-out direct-attach cables (DAC), the same design can be used with 32x100G leaf switches and 25G servers and 100G to 4x25G break-out DACs (Figure 3). If the leaf is a high-density 10G switch with 40G uplinks to the spine, the same density is available with 25G servers and 100G uplinks allowing for a simple and seamless migration.

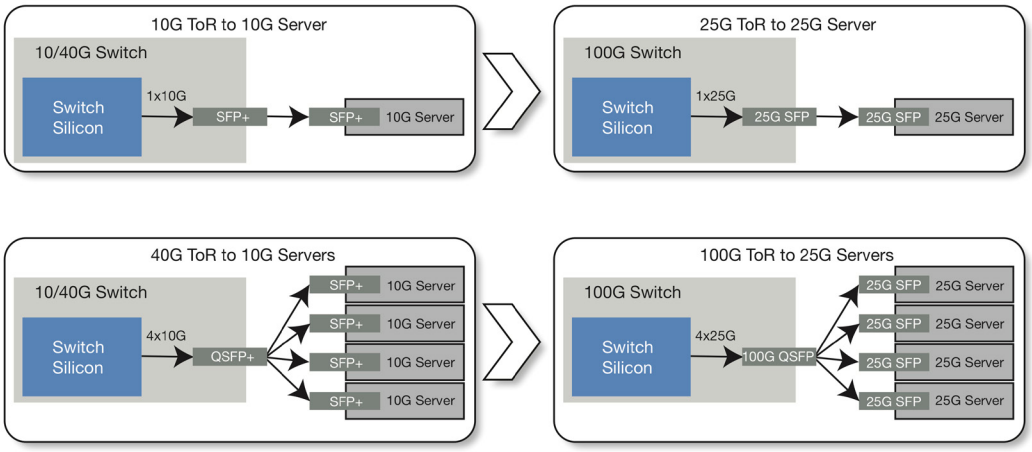

Figure 3: Identical rack-design and cabling infrastructure when migrating from 10G to 25G

Lower OPEX: 25G Ethernet offers higher bandwidth at the same power consumption compared to 10G solutions, and 25G solutions use SFP connections providing higher density than a comparable 40G solution (see Table 1). Both the power savings and higher density result in lower cooling requirements and lower operational expenditure for datacenter operators.

Backward compatibility: A majority of the 25G and 100G capable switch silicon offers backward compatibility to both 10G and 40G speeds, which provides ultimate flexibility to mix and match port speeds with a gradual migration to higher speed servers. All 100G QSFP and 25G SFP ports on Arista switches can also be used as 40G or 10G ports, respectively. With a minimal cost premium for 100G and 25G based systems compared to 40G and 10G systems, it becomes an easy choice to deploy 100G and 25G based solutions, continuing to operate at lower speeds and migrate on a per server basis to higher speeds when required.

Forward Error Correction (FEC) Requirement for 25G

To correct for errors and protect the data integrity, Forward Error Correction (or FEC) is used in 25Gb and higher speeds to improve data transfer throughput without the need for re-transmitting the data. FEC is a digital processing technique that greatly reduces the number of uncorrected errors and helps to extend the reach capability of both Copper and Optical PMDs. There are a number of FEC modes, offering differing degrees of error detection and recovery. The IEEE 802.3by standard has defined Clause 91 with Reed-Solomon FEC (RS-FEC) for 25G Ethernet to support most of the Copper and Optical physical media dependent devices (PMDs). Clause 74 of the 100GbE standard specifies BASE-R FEC also known as fire-code FEC, and is available in most 100G capable switch silicon. FC-FEC offers a weaker error correction but with lower latency compared to RS-FEC. To achieve error-free communication, the same type of FEC has to be enabled on both ends of the link. However, the use of FEC is not mandatory, and there are certain optics and cables options to avoid the use of FEC, either to bypass the lack of support at the switch/server level or to eliminate the additional latency added by the FEC. The dominant high-speed connection type for server to switch connectivity in rack applications is direct-attach copper (DAC) cables. IEEE 802.3by has specified three different 25G cable types, with different loss characteristics each with an expected FEC mode requirement, as described in Table 2. While using no FEC is possible if the cable is designed to a low loss specification, other cable types have an expectation of FEC to operate reliably.

Table 2: 25G Cable Types and FEC Requirements

25G Copper Cable Type	Loss Specification	Minimum FEC Requirement
CA-N	12.98 dB	No FEC
CA-S	16.48 dB	BASE-R FEC
CA-L	22.48 dB	RS-FEC

In addition to copper cables the IEEE specified 25GBASE-SR, industry specified 25GBASE-LR optical transceivers and the 25G active optical cables also require RS-FEC. Table 3 provides a summary of the 25G optics and cables from Arista along with the associated minimum FEC requirement.

Table 3: 25G Optics and Cables from Arista and Minimum FEC Requirements

SKU	Description	Minimum FEC Requirement
CAB-S-S-25G-xM	25G SFP to SFP twinax copper cable, 1, 2, 3 meters	No FEC (CA-N type)
CAB-Q-4S-100G-xM	100GBASE-CR4 QSFP to 4 x 25GbE SFP Twinax Copper Cable, 1, 2 meter	No FEC (CA-N type)
CAB-Q-4S-100G-3M	100GBASE-CR4 QSFP to 4 x 25GbE SFP Twinax Copper Cable, 3 meter	BASE-R FEC (CA-S type)
SFP-25G-SR	25GBASE-SR SFP transceiver up to 70m over parallel OM3 or 100m over OM4 multimode fiber	RS-FEC
SFP-25G-LR	25GBASE-LR SFP transceiver, up to 10km over single-mode fiber	RS-FEC
AOC-S-S-25G-xM	25GbE SFP to SFP Active Optical Cable, 3m to 30m lengths	RS-FEC

Interoperability

One goal of the 25G Ethernet consortium was to set an industry-standard interoperable specification that enables early market adoption, while reducing connectivity costs. With the IEEE also adopting and standardizing the specifications, there has been significant cross-vendor effort in the industry to test for interoperability between Network Interface Controllers (NIC) and switches. Arista Networks attended a 25G plug-fest, organized by University of New Hampshire Interoperability Lab (UNH-IOL). Key vendors in the 25G Ethernet ecosystem participated to test, demonstrating the strong desire to ensure interoperability. Results of the plug-fest showed broad availability of multi-vendor interop with switches, cables and NICs. With 25G technology maturing in the near future, end users can expect seamless operation between the main components within the datacenter network.

50G Ethernet

The stated objectives of the 25G Ethernet consortium include 50Gb Ethernet as an interface made up of 2x25G lanes. While this has not been adopted by the IEEE yet, there are several 50G Network Interface Controllers available today that can be connected to 100G switches. Arista 100G switch ports can be configured for 2x50G mode and connected directly to 50G servers. A 50G QSFP cable has been specified by the 25G consortium as a 2-way break-out twinax cable with one end of 100G QSFP and 2 separate 50G QSFP at the other ends. This enables connectivity between a 100G switch port and a 50G server, as shown in figure 4. Since the 50G specification utilizes the same 25G serdes technology and interoperates with a broad range of 100G switches, mega-scale datacenters can achieve higher efficiency with 50G attached servers compared to 40G.

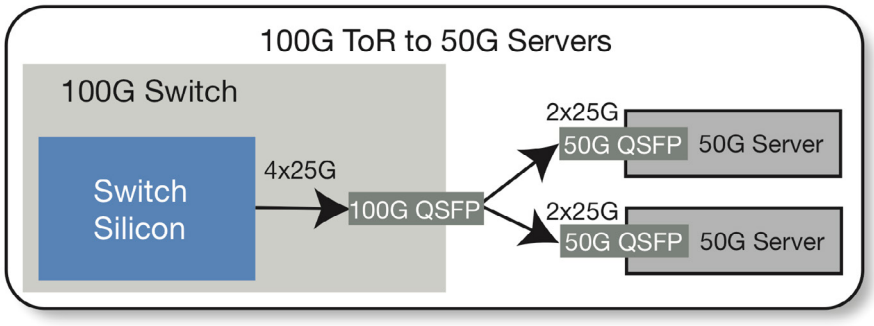

Figure 4: 100G switch to 50G server connectivity using depopulated copper cable

Arista 100GbE and 25GbE Systems

As a founding member of the 25G Consortium, it is not unexpected that the Arista Networks leaf and spine switches include a wide range of systems with support for 100GbE using QSFP ports that can be converted to 4 lanes of 25GbE and a growing range of systems with support for 25GbE ports using the SFP connectors. Each system that supports 100G or 25G mode can also support both 40G or 10G, respectively. The multi-speed capability provides the backward compatibility necessary to allow for easy and cost effective upgrades as well as a mix and match of speeds during extended upgrade cycles.

The Arista range of fixed leaf and spine platforms offer a wide range of systems with full support for the IEEE 25GbE 802.3by standard, including the 7050X3, 7060X2, 7160 and 7280R2 Series.

Summary

The 25G Ethernet technology is ready for primetime. It offers higher speed, lower power consumption and lower cost per bit compared to 10G Ethernet. 25G Ethernet is available with the familiar small form factor (SFP) pluggable, mitigating concerns around transitions. It utilizes familiar cable types and connects seamlessly to widely available 25G and 100G Ethernet switches. The performance, density and cost benefits of 25G and 50G Ethernet are clear, and the migration makes absolute sense. Multiple market forecasts show 25G Ethernet deployments becoming the second highest server port deployment after 10G Ethernet in the very near future. The zettabyte era is here. Is your datacenter ready?

Santa Clara—Corporate Headquarters
5453 Great America Parkway,
Santa Clara, CA 95054

Phone: +1-408-547-5500
Fax: +1-408-538-8920
Email: info@arista.com

Ireland—International Headquarters
3130 Atlantic Avenue
Westpark Business Campus
Shannon, Co. Clare
Ireland

Vancouver—R&D Office
9200 Glenlyon Pkwy, Unit 300
Burnaby, British Columbia
Canada V5J 5J8

San Francisco—R&D and Sales Office
1390 Market Street, Suite 800
San Francisco, CA 94102

India—R&D Office
Global Tech Park, Tower A & B, 11th Floor
Marathahalli Outer Ring Road
Devarabeesanahalli Village, Varthur Hobli
Bangalore, India 560103

Singapore—APAC Administrative Office
9 Temasek Boulevard
#29-01, Suntec Tower Two
Singapore 038989

Nashua—R&D Office
10 Tara Boulevard
Nashua, NH 03062