

## Q&A: Ethernet Evolution—25 is the New 10 and 100 is the New 40

[Electronic Design](#)

[William Wong](#)

Thu, 2015-10-29 09:23



[Mellanox](#) has been known for its support of InfiniBand, but it's also a major player in the high-speed Ethernet market. One of the latest technologies is 25-Gbit/s Ethernet. Dell'Oro predicts that it will have the most rapid adoption compared to any previous speed. In fact, it expects two million adapters to be shipped in 2018.

I spoke with Kevin Deierling, Vice President of Marketing at Mellanox, about this trend and the part Mellanox expects to play in the market.

**Wong:** When you talk about your Ethernet products, you say “25 is the new 10.”

What do you mean by that and how did 25GbE come into being?

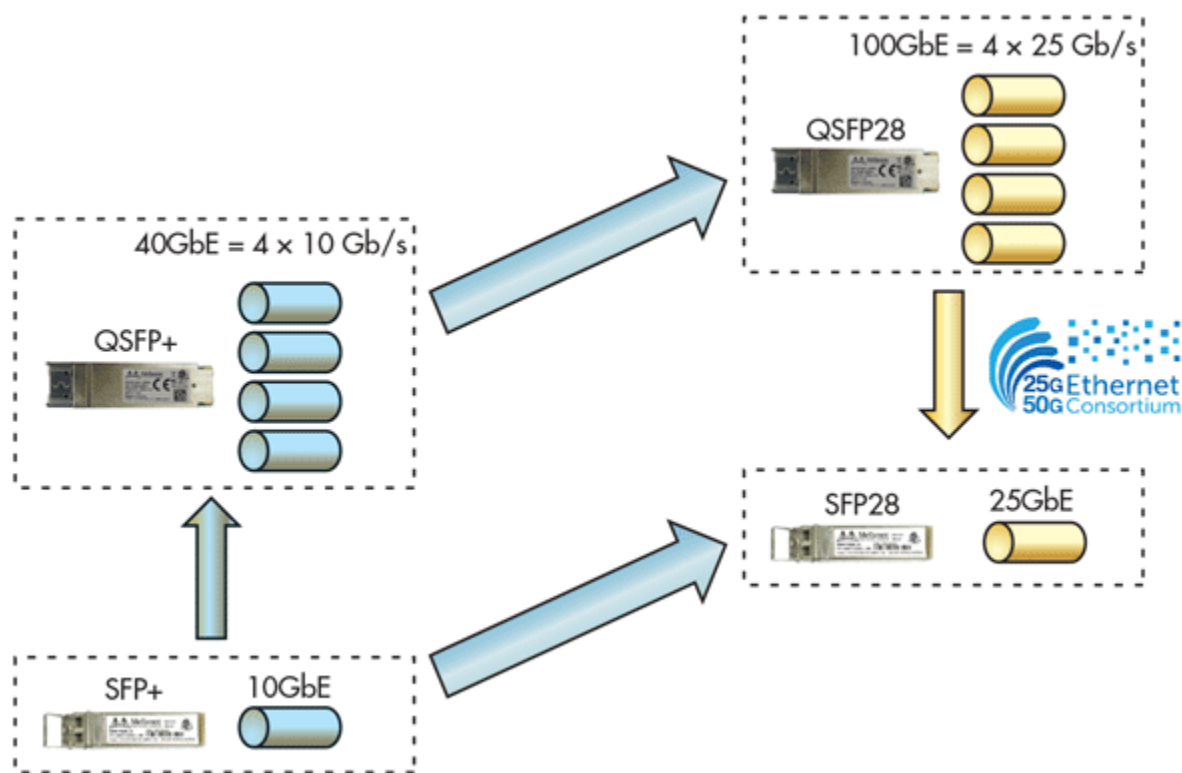
Related

[Justifying 2.5G/5G Ethernet](#)

[Emulation Fast-Tracks Networking Products to Market](#)

[Data-Center Networking: What's Next Beyond 10 Gigabit Ethernet?](#)

**Deierling:** Ethernet has traditionally increased by 10X in speed every generation, meaning a jump from 10GbE to 100GbE is the next step. In fact, we are now seeing the first 100-Gb/s Ethernet devices come to market, but there are also additional intermediate speeds including 25, 40, and 50 Gb/s. Many of the largest data centers in the world have already moved to 40 Gb/s, which consists of four 10-Gb/s lanes. These data centers are well-equipped to move to 100-Gb/s Ethernet solutions.



However, many data centers are only cabled to support 10-Gb/s Ethernet. This would require new optical cabling infrastructure, since 10-Gb/s optical cables won't support 100-Gb/s rates today. That is where 25-Gb/s Ethernet comes in. It uses the same LC fiber cables and the SFP28 transceiver modules are compatible with standard SFP+ modules. This means that data-center operators can upgrade from 10 to 25 using the existing installed optical cabling and get a 2.5X increase in performance. For this reason, industry analysts, including both Dell'Oro and Crehan, are forecasting that 25-Gb/s Ethernet will have the most rapid adoption of any previous networking speed, with Dell'Oro predicting two million adapters shipping in 2018.

It seems like 25GbE has come out of nowhere, but the technology actually came into being as the natural single-lane version of the IEEE 802.3ba 100-Gb/s Ethernet standard. The 100-Gb/s Ethernet standard uses four separate 25-Gb/s lanes running in parallel, so defining a single lane makes it a straightforward and natural subset of the 100-Gb/s standard. It is different from the previous 10- and 40-Gb/s generations, where the single-lane 10-Gb/s standard came first and was then followed by the four-lane 40-Gb/s version.

The standardization of 25GbE has taken a somewhat different path—both within an industry consortium focusing on interoperability and within the IEEE with a focus on backwards-compatibility and detailed end-to-end connectivity, including cabling and noise budgets. The IEEE initially voted against standardizing 25-Gb/s Ethernet. As a result, the 25GbE industry consortium (25gethernet.org) was formed in July of 2014 with Mellanox, Google, Microsoft, Broadcom, and Arista as founding members. Along with the original five founding members, Cisco and Dell joined as promoters of the technology, with over 40 other members at the adopter level. Shortly after the formation of the 25G Consortium, the IEEE decided to form the 802.3by 25 Gb/s Ethernet Task Force to standardize the technology, with the last changes expected in November 2015 and the ratified standard expected in Q3 2016.

**Wong:** What about 100-Gb/s Ethernet? How is this related to 25-Gb/s Ethernet and where might I use these speeds?

**ierling:** The 100-Gb/s Ethernet speeds are also expected to grow very rapidly for two reasons. The first is top-of-rack (TOR) switches as a means to aggregate multiple 25GbE ports to the servers, and connect TOR to the aggregation or spine switches to create massive scale-out clusters in the data center. The 100-Gb/s links use only a single cable and thus reduce the interconnect cost and complexity.

The second driver for 100 Gb/s concerns server connectivity for new, advanced platforms—particularly for software-defined storage (SDS), sometimes called Server-SAN storage. Good examples include Ceph, GPFS, VSAN, ScaleIO, Storage Spaces, and Nexenta. These scale-out storage systems used a converged interconnect that can merge front-end and back-end connections.

Furthermore the adoption of flash-memory-based SSDs is driving the requirement for higher speeds. In fact, [Microsoft demonstrated that by using RoCE-based NICs, they were able to achieve 100-Gb/s line rate performance with just three NVMe Flash SSD devices.](#)

**Wong:** What are the key features I should I look for in a 25-Gb/s Ethernet NIC and switches?

**Deierling:** As always, the basic features of size, weight, and power, and of course, cost, are critical to consider when evaluating networking equipment. These “SWaP+C” features really drive the decisions to move to any new technology. Systems such as the Mellanox SX1200 at 40 Gb/s and SN2100 at 100 Gb/s are about half the size and consume only half the power, which helps to reduce total cost of ownership compared to other industry offerings. In addition, white-box “Open Ethernet” switch solutions decouple the hardware from the software, resulting in additional cost savings.

For some deployments, it is important to have a complete networking stack bundled with the switch. However, Open Ethernet also gives customers the option to use a third-party stack from their preferred Network Operating System software supplier. This gives customers the flexibility to choose the best hardware and software for their specific application needs.



After these basic considerations, you need to dig into the technical details, such as switch performance. Here, throughput, latency, and packet performance across all packet sizes are keys. The state of the art is the Mellanox SN2700 with latency of 300 ns, which is about half that of other offerings. Even more important than latency is the ability to switch packets at full wire speed for all packet sizes. So it is important to test switches to ensure they can maintain full-wire speed switching on all ports without dropping packets.

Obviously, at these speeds, packet loss is unacceptable in the data center, so it is surprising that many switches drop packets when faced with a demanding workload. The Tolly Group’s [basic RFC2544 testing of several 40Gb/s switches](#) showed that switches based on the Broadcom switch ASIC actually drop between 4% and 20% of the packets. Many of these switches are advertised as full wire speed\* with limitations exposed in a footnote, or in some cases, the limitations not disclosed at all. So it is important to look carefully for the asterisk and footnotes to ensure that a switch is truly able to operate at full wire speed across all packet sizes. Mellanox switches based on the SwitchX-2 at 40 Gb/s or the Spectrum at 100 Gb/s are able to operate full wire speed, for all packet sizes, with [zero packet loss](#).

**Wong:** What about InfiniBand? How is InfiniBand doing against competitive offerings, and is it still growing in

high-performance computing (HPC) market?

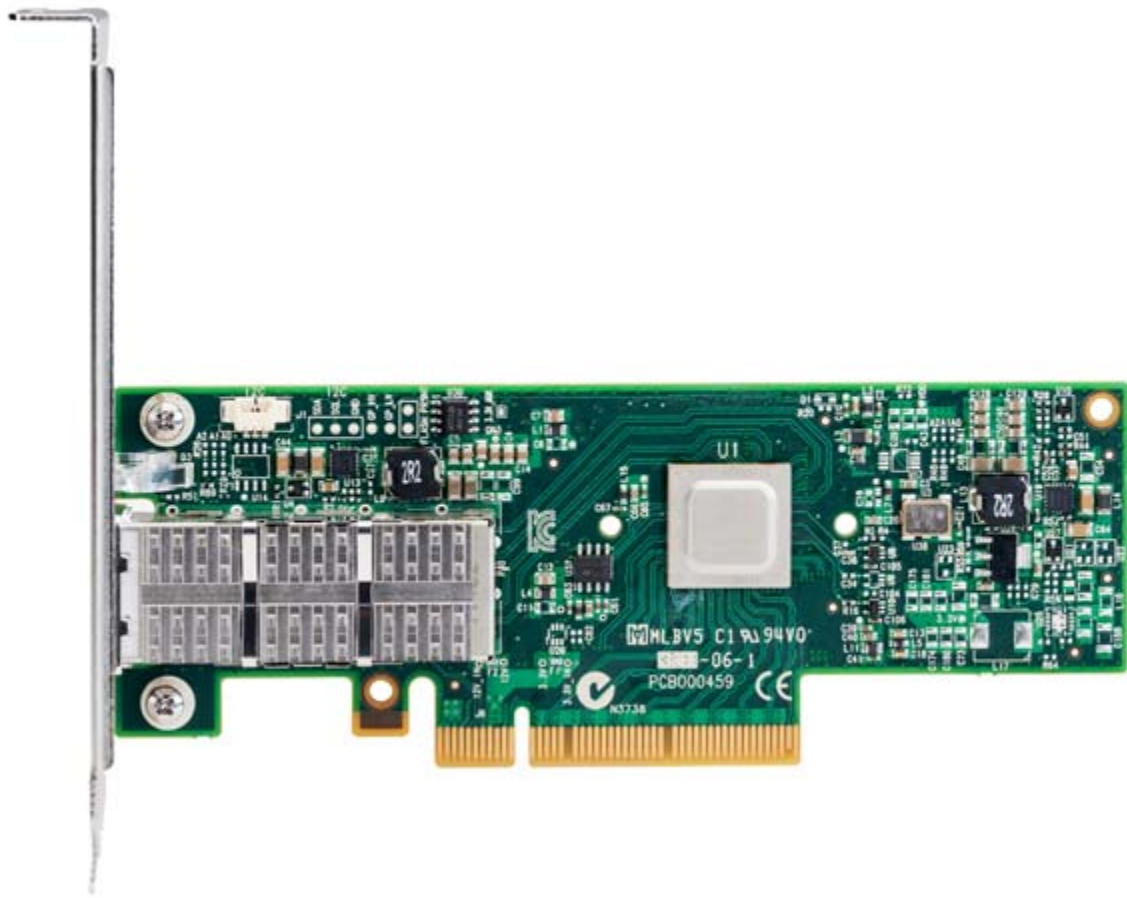
**Deierling:** InfiniBand technology continues to grow in the HPC market and capture market share. At the International Super Computing show this year, for the first time [InfiniBand powered over half of the world's fastest computers](#). InfiniBand-based system delivered 40% to 50% better efficiency than systems based on other interconnects. Being able to perform more work with the same server hardware makes a pretty compelling business case and explains the adoption of InfiniBand. We are just at the beginning of the transition from FDR to EDR (100 Gb/s), and for the first time saw EDR systems crack the Top 500 list. The shift to EDR delivers even higher performance and efficiency, and will drive continued growth in the HPC market.

**Wong:** With new higher rates, has the gap between Ethernet and InfiniBand narrowed or is InfiniBand still growing in other segments besides HPC?

**Deierling:** Actually InfiniBand continues to grow very nicely in many different market segments beyond HPC, including enterprise, storage, Web 2.0, and cloud segments. It turns out that InfiniBand is still far and away the highest-performance interconnect for building scale-out systems, whether for compute, cloud, or storage.

InfiniBand is at the heart of the most-efficient, highest-performance, most-reliable, and most-scalable enterprise appliances. These systems host the most demanding and mission-critical applications at the largest government and business enterprises in the world. These platforms are not just being deployed for enterprises, but also in the new generation of public cloud offerings. For example, Salesforce.com, the largest Software-as-a-Service provider, is hosted on Oracle ExaData platforms running on an InfiniBand backbone. Similarly, workloads like Parallel Data Warehouse, Bing Maps, Hadoop, advanced analytics, and other major Web 2.0 websites are being hosted on InfiniBand-based systems.

And it's not just database and analytics appliances, but enterprise storage, too. EMC's top-of-the-line VMAX3 platform, along with virtually all of the major fault-tolerant enterprise storage platforms in the world, has adopted InfiniBand for the core interconnect. These highly available storage systems combine fault tolerance with high performance and deliver worry-free solutions to the most demanding enterprises in the world.





InfiniBand continues to evolve with higher bandwidth, lower latency and new features being added all the time. Thus, its continued growth is ensured in HPC, enterprise, storage, Web 2.0, and cloud markets.

While the most efficient appliances and storage platforms run on InfiniBand, some users simply prefer Ethernet. So it's really not an "either-or" question. For those who want 25-, 50-, or even 100-Gb/s Ethernet, it's here today with end-to-end solutions available. And for those demanding the ultimate in performance, EDR InfiniBand solutions offer even lower latency, higher efficiency, and superior performance.

In many cases, system vendors use both InfiniBand and Ethernet depending on what their customers want. To make this simple, even Virtual Protocol Interconnect (VPI) products, such as the ConnectX family of adapters, offer the ultimate flexibility and can be configured to operate either as InfiniBand or Ethernet based on firmware. The choice is ultimately in the hands of the customer, but no matter what, big data and faster storage demand higher-performance networks.

**Source URL:** <http://electronicdesign.com/ethernet/qa-ethernet-evolution-25-new-10-and-100-new-40>