Silicon Photonics

THE KEY TO DATA CENTRE CONNECTIVITY

ROBERT BLUM

Data centre traffic growth is driving the need for high-speed connectivity between servers and switches. Silicon photonics will be a key enabling technology to meet the future demands, writes Intel's **Robert Blum.**

Significant innovations in optical connectivity are required for data centres to scale compute and storage functions so that they can continue to support future bandwidth-intensive and compute-intensive applications.

This is where silicon photonics comes into play with its potential to bring electronics-type cost and scale to the optics industry – an industry that has traditionally been focused on lower volume or longer distance applications with limited ability to scale to the requirements of modern data centres. Silicon photonics has now been productised and will be a key enabling technology to meet the future demands for bandwidth in data centres.

Networks inside data centres are often based on Clos topologies (a type of non-blocking, multistage switching architecture that reduces the number of ports required) and Hyperscale data centres will typically contain tens of thousands of Ethernet switches to interconnect the racks of servers through a leaf and spine network architecture.

A typical data centre today has one or two 10GbE based network interface controllers deployed at the server, which is then aggregated to 40GbE at the top of rack (TOR) switch. The connections between server and TOR are usually made through direct attach copper (DAC) cables, since these are the most cost effective solutions at these data rates for distances of a few metres. But the uplink from TOR to the next tier switch is almost always optical.

Smaller data centres will typically use VCSEL-based transceivers over multi-mode fibres. These 40G transceivers aggregate four 10G lasers and can transmit over distances of up to 300 metres. Higher tier switch interconnects (leaf to spine and above) usually require the use of single mode fibres since distances between the switches will often exceed 300m.

100G UPGRADES AND SINGLE MODE

As data centres transition to 100GbE, the move to single mode becomes inevitable for several reasons: First, the transmission distance over standard multimode fibre is limited to about 70m at 25G, which is too short a distance for most data centres. Second, various multi-source agreements optimised for 100G data centre connectivity have been formed, resulting in a much lower cost differential between multimode and single mode transceivers. Third, in anticipation of the increased demand for bandwidth, several large data centre operators had already proactively installed large amounts of single mode fibre in their trunk cabling, so there is not even a one-time "upgrade" cost to migrate to single mode connectivity.

INTEL'S APPROACH

One of the key benefits of Intel's approach to silicon photonics is that all the functionality needed for a 100G transmitter – in the case of a CWDM4 chip, this includes 4 lasers with different wavelengths, 4 modulators and an optical multiplexer – can be integrated on a single silicon die. This is made possible by the unique hybrid laser and results in significant advantages in manufacturability and product cost. For example, this level of integration enables on-wafer testing of the complete transmitter such that only known-good-die are passed on to the backend assembly, resulting in improved yields due to lower complexity and a reduced number of assembly steps. Similar transceivers based on traditional optics or discrete lasers require the assembly of multiple optical components, typically involving multiple bonding steps and multiple active alignment steps. And discrete designs also drive larger footprints, which limits their scalability to future form factors.

SWITCHING TRENDS

The TOR, leaf and spine switches being deployed today are typically 3.2Tb/s Ethernet switches in a 1RU chassis, though some 2RU systems are also available at 6.4Tb/s switch capacity.

These switches have 25G SERDES and work well with the 100G QSFP28 transceivers. As switches move from 3.2T/6.4T to 12.8T, line rates will move to 50G SERDES using PAM4 modulation. At these data rates, new transceivers based on 50G electrical I/O are needed.

To support 12.8T of switch capacity in a single RU, 400G transceivers in a form factor about the size of the QSFP are needed and two MSAs have formed to do just that: the QSFP-DD (DD stands for double density) and the OSFP (O stands for octal). Both MSAs support 8 lanes of electrical I/O at 50G PAM4 and can therefore support 400G optical interfaces. The challenge will be which optical interface will be suitable for 400G data centre connectivity.

The IEEE has standardised a DR4 interface, which is similar to 100G PSM4 but using 100G PAM4 optical modulation instead of 25G NRZ on 4 parallel fibres. Using PAM4 modulation results in a significantly reduced link budget and higher power consumption due to additional ICs- as well as increased complexity, but it does allow the reuse of existing fibre infrastructure already deployed for 100G PSM4. For duplex fibre, no uncooled option currently exists and it's to be expected that MSAs will form to address the need for a low-cost and manufacturable duplex fibre solution.

BEYOND 400G

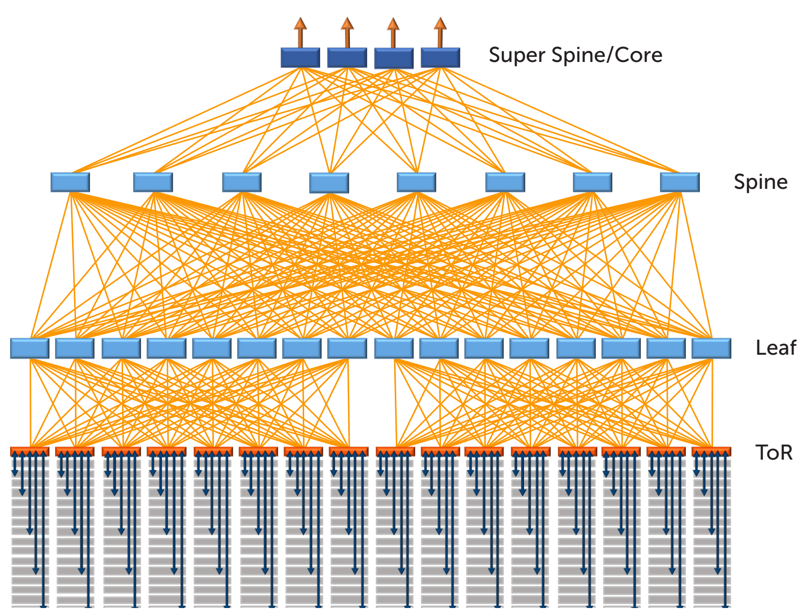
As data centres evolve beyond 12.8T switches and 400G pluggables, integration of optics with networking silicon will eventually become inevitable. This is indeed one of the benefits of integrating lasers, modulators and optical multiplexers on a single chip: It enables the continued evolution towards higher bandwidth and smaller form factors.

This integration of photonics and networking silicon will likely be done in a multi-chip package, leveraging the developments made in semiconductor packaging and taking advantage of using the best process nodes for each component.

Today's operational focus remains on delivering 100G transceivers for data centres which already drive volumes in the millions of units per year. The next generation 400G pluggable form factors are defined and the industry is preparing to demonstrate 12.8T switches with optical interfaces in the near future. As we look beyond 12.8T, pluggable transceivers are likely running out of steam – they are too large and too power hungry – and we expect a paradigm shift in how optics will be deployed in the data centre.

Robert Blum is Director of Strategic Marketing and Business Development for Intel's Silicon Photonics Division.

Intra Data Centre Network Topology



The optical connections (drawn in orange lines) can be hundreds of meters and are migrating to single mode as data centres upgrade to 100G connectivity.