molex

DEVICES FOR DATA CENTRES

BY RANG-CHEN (RYAN) YU, VICE PRESIDENT OF BUSINESS DEVELOPMENT, GM OF OPTOELECTRONIC SOLUTIONS AT OPLINK, A MOLEX COMPANY

EFFICIENT 100G/400G OPTICAL TRANSCEIVER SOLUTIONS FOR HYPERSCALE DATA CENTRES

According to business information provider IHS, a decade of continued high growth of global network data centre traffic shows no sign of abating anytime in the foreseeable future. The phenomenal rise in the popularity of smartphones and other mobile devices, social media and apps, streaming video, augmented and virtual reality - garnering new users, more devices per user and rising data usage per device - account for a significant boost in data centre traffic. By 2020, analysts predict there will be 200 billion internet connected devices globally. Recent evidence suggests that the maturing consumer electronics markets may be just the tip of the iceberg. Growth rates for data bandwidth in cloud computing and machine-to-machine deployments are outpacing consumer data traffic and driving massive demand for high capacity data centre infrastructure.

DATA CENTRE AND OPTICAL INTERCONNECT GROWTH TRENDS

Over the last decade, top internet web companies such as AWS, Microsoft, Google and Facebook have been busy deploying larger data centres to meet customer demand, with some of them now containing over 100,000 computer servers per building. These hyperscale data centre providers leverage economies of scale by consolidating processing power in sprawling data centres near locations where real estate and energy supplies are abundant and less expensive. By 2020, nearly half of installed servers across all data centres will be housed in hyperscale data centres, according to Cisco. Those servers will represent 68% of processing power and 53% of total data centre traffic.

With more mission critical business applications and time-sensitive consumer applications powered by the cloud, more data centres are being deployed closer to population centres around the globe. Increasingly, web companies are building data centres with multiple buildings in close proximity and interconnected with massive bandwidth. Building data centres on separate power grids in higher populated areas can additionally lower latency and improve the consumer experience. The strategy can also overcome the risks and limitations of a larger data centre relying on a single power grid.

Inside each hyperscale data centre building, there may be tens of thousands to hundreds of thousands of computer servers interconnected by tiers of Ethernet switches to form a collective computing capability for serving web companies' own service (e.g., Google or Facebook), or for renting out to enterprise customers (e. g., Amazon AWS or Microsoft Azure). While there are many variations of schemes to interconnect computer servers, a typical 2018 hyperscale data centre networking connection is characterised by servers connected to a top-of-the rack (ToR) switch within a few metres at 25 or 2x25 Gbps with DAC (Direct-Attached Copper) cable. ToR switches are then interconnected via a massive switching fabric, often called leaf-spine architecture, by a large number of 100 Gbps optical links. Depending on the size of these data centres, typical optical interconnect ranges can be covered to a maximum of 500 metres, but large data centres require distances of up to 2km.

The current generation of 100G optical transceivers are based on 4 channels of optical transmitters and receivers with each running at 25 Gbps in parallel to achieve 100 Gbps aggregate. There are two types of 100G optical transceivers.

For those users willing to deploy more fibre and get lower cost per transceiver, a PSM-4 (Parallel-Single Mode-4) type transceiver is suitable. For those users who wish to deploy less fibre, a CWDM- 4 (Coarse WDM-4) type transceiver is preferred. Both types of 100G optical transceivers are being deployed in high volume today.

UPCOMING 100G/400G TRANSITION AND 100G PAM4 TECHNOLOGY

Current hyperscale data centre networking is characterised by a much faster pace of interconnect speed transition, typically occurring at three-year intervals. Innovative 100G interconnects are going mainstream and have been in deployment for the last two years, while the next speed transition is looming on the horizon. Although 200 Gbps is being considered, there is industry consensus that 400 Gbps will be the natural next step.

Current 4x25G based 100G technology is complex to package and is not scalable to 400G. In order to reduce 100G cost and support 400G optics economically, the industry is moving to adopt a new technology with optics encoded with PAM-4 (4-level Pulse Amplitude Modulation) at 50 GBaud, enabling 100G per channel and later 400G with 4x100G aggregation. The 100G Lambda MSA (Multi-Source Agreement) was formed to define this new industry standard and supported by 23 promoting companies, representing a broad industry ecosystem including companies making semiconductor integrated circuits, optical transceiver modules and networking systems, as well as end-user web companies.

The strong benefit of adopting single channel 100G optics includes a much reduced optics element count for lower cost, thereby laying a foundation for economical 400G and the elimination of an inverse gearbox when

Originally published in opticalconnectionsnews.com (Issue 13, Q2 2018)

molex

the electrical interface migrates to 100G serial interface in the future. According to one estimate, PAM-4 100G delivers a 60% reduction in component count and a 33% reduction in power requirements.

The 100G Lambda MSA has recently released to the public a first draft of specifications that define 100G FR (2km), 100G LR (10km) and 400G FR4 (2km) and potentially also defines 400G LR4 (10km). With advances in digital signal processing and high speed optoelectronic device technologies such as high-speed silicon photonics, we anticipate rapid industry adoption and implementation with field deployments possibly starting as early as 2019.

INTERDATA CENTRE DCI SOLUTIONS

Hyperscale data centres are deployed near population centres globally and interconnected with ultra-high bandwidth. While many ultra-high speed optical fibre links are deployed across continents and oceans, the majority of these links are between data centre buildings within a data centre campus, or data centres within the same metropolitan area. These data centre buildings are interconnected with massive bandwidth, which can be up to tens of terabits per second.

For interconnected data centres within a few kilometres on one another, an operator may choose to deploy simple 100G CWDM4 (2km) or 100G LR4 (10km) type optical transceivers and migrate to 100G FR/LR (utilising PAM-4 technology), with several hundred pairs of fibre. If fibre is not adequate and adding more fibre becomes too costly, operators may choose to deploy DWDM (Dense Wavelength Division Multiplexing) optical transceiver solutions in order to aggregate up to 40x100G per pair of fibre. For these on-campus short distance interconnections, single channel 100G PAM-4 with direct detection is a much more economical and attractive solution versus more complex coherent transmission technology, which requires both amplitude and phase modulation/decoding with polarisation multiplexing/ demultiplexing and coherent detection using a precisely controlled optical local oscillator.

For interconnect data centres up to 80km apart, 100G PAM-4 DWDM with advanced digital signal processing technology may still be cost favourable, even with the added tunable dispersion compensation requirement which can however be shared among all DWDM channels. Coherent detection will be used to cover distances greater than 80km distance. When data centres transition to 400G, DCI solutions will scale accordingly, but 4x100G PAM-4 can still be used to cover relatively short reach DCI applications and coherent 400G will extend coverage for the remaining inter-data centre connections.

OPTICAL TRANSCEIVER FORM FACTORS

For 100G data centre applications, the industry has overwhelmingly adopted the QSFP28 (Quad Small Form-factor Pluggable) transceiver module. As the industry is preparing to transition from 100G to 400G, several emerging MSA form factors are contending for adoption. The leading candidate is the QSFP-DD (Quad Small Form-factor Pluggable Double Density), which is derived from the QSFP28 and has twice the electrical data connections and an only slightly longer mechanical length, thereby preserving compatibility with the QSFP28. In conjunction with an improved thermal design, a QSFP-DD transceiver module and cage configuration can support power dissipation beyond 12W.

The second contender is the OSFP (Octal Small Form-factor Pluggable) optical transceiver, which is slightly larger and longer than the QSFP-DD interface. The key advantage of the OSFP module is the larger form factor which allows for higher power dissipation of up to 16W. The disadvantages are the lack of backward compatibility with the QSFP28 and the slightly larger size which translates into a lower faceplate density.

A third MSA called COBO (Consortium of On-Board Optics) has defined a form factor with the electrical interface located away from the system faceplate and directly onto the system PCB. The advantage of this configuration is the placement flexibility of the transceiver module which can be closer to a switch IC higher speed interface, making it easier to deal with signal integrity issues. Since the COBO transceiver module is mounted on a 2D PCB surface, there is also more room for heat sink implementation, thereby supporting a potentially higher power dissipation rating.

Hyperscale data centres are quickly becoming critical for major web companies, which are investing heavily and at a fast pace to keep up with technology developments and service innovations. The development of faster electrical and optical signalling technologies will continue to accelerate massive data aggregation at hyperscale data centres globally. The latest 100G and 400G optical technology developments provide a broad range of efficient hyperscale data centre connectivity solutions to support an ever-increasingly rich mix of data intensive applications.

Order No. 987651-8668 USA/0k/GF/2018.06