CORNING

To Infinity, and Above!
A 400G story

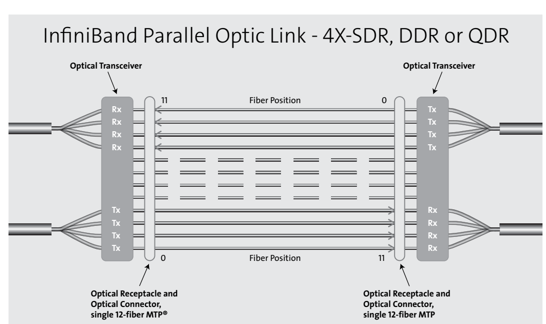
Just like the internet it enables, the data center industry never sleeps. Our insatiable appetite for bandwidth continues to challenge technology developers to keep pushing data faster and more efficiently than ever before. They are continually designing more engaging applications, processes, and products to enrich our social and professional engagements, de-humanising the everyday mundane, complex tasks and enveloping them in the digital bubble we live in today. This is Demand and Supply 101. We demand it and boy, are they supplying it in such a way that it creates a super dependence we never knew we had. My wife often jokes she doesn’t know how I survived before the search engine!

What’s behind this drive for faster, more efficient network speeds? Basically, it’s us. The human race. We get bored easily. We are constantly searching for new, more interactive ways to entertain ourselves, a more exciting and engaging method of communicating like never before and we expect it all to happen now. This instant. We don’t use the phone to talk to people anymore. Text messages are old school, and picture messages are rapidly becoming yesterday’s news. If you want to be down with the kids, it’s all about streaming ultra-rich content at every possible opportunity. Day and night. But it goes further than this. We want to connect more devices to the information super highway, we want improved analytics, we want more machines doing more for us than we ever imagined possible and we want it at our fingertips like ... yesterday.

I remember back in 2010 when the IEEE approved 40 Gb and 100 Gb and thinking there’s no way we would ever need to call on these colossal speeds. My broadband connection back then was pathetic by today’s standards – like trying to blow up a medicine ball with drinking straw – but at the time it served me well because I wasn’t consuming anywhere near the amount of bandwidth that I can’t live without today. But once my kids started using their cell phones, tablets, and online gaming platforms, my poor little copper connection started to groan with the demands that were placed on it. Thankfully, today I have fiber delivering over 30x the speed I used to have, but with more and more devices at home, I could always do with more. Much, much more!

Let me give a little bit of background on the technology road map to bring us right up to speed (ahem) on where we are today. Up until 2010, life was simple. We were running 10 Gbs across our fiber networks using 1 fiber to transmit and 1 fiber to receive. It didn’t matter whose active devices hung off the end, or which transceiver was deployed, because they all operated the same. And then IEEE ratified 40 Gb using a different physical media dependent (PMD) interface with parallel optics.

Parallel Optics



Unlike serial (duplex) transmission, parallel optics uses 8 fibers or more instead of the traditional 2 fibers. We transmit 10 Gb per fiber over 4 fibers and receive 10 Gb per fiber over the other 4 fibers to reach our aggregated 40 Gb speed (known as 40G-SR4). And because the fastest switch application specific integrated circuits (ASICs) available in 2010 were 10 Gb, to reach 100 Gb (known as 100GBASE-SR10), also ratified at the same time, we needed to have 20 fibers, 10 each to transmit 10 Gb and the other 10 to receive 10 Gb. Thankfully, in 2015 the IEEE ratified 100G-SR4 transmitting 25 Gb over 4 fibers and receiving 25 Gb over the other 4 fibers to reach our aggregated 100 Gb speed using the same 8-fiber platform as 40GBASE-SR4, and this consistency of fiber counts and electrical lanes continues as we move toward 200G and ultimately the 400G we saw announced last year.

So now this always-on industry presented network managers and facility owners with a choice: As more and more options are added, critical decisions need to be made around what infrastructure solution you should deploy to ensure seamless migration path to higher speeds. Let me break them down for you.

BiDi

Three years after IEEE approved 40 and 100 Gb, late 2013 saw the introduction of 40G BiDi from Cisco using a technology developed by Avago Technologies, now part of Broadcom's Optical Systems Division (OSD). This pioneering technology was developed to enable the existing two MM fiber networks that were installed in data center facilities using wavelength division multiplexing (WDM) techniques. To achieve 40 Gb, 2 x 20 Gb signals on the same multimode fiber but at different wavelengths, or colors of light, transmitting and receiving 20G at 850 nm and the other at 910 nm 'bidirectionally,' hence the name BiDi. We saw the release of 100G BiDi in 2018.

UNIV

Shortly after the BiDi release, Arista and Juniper launched their universal transceiver for 40 Gb (UNIV) using yet another WDM technique to operate across a single pair of fibers. This time it launched 4 x 10 Gb in the traditional single-mode 1310 nm region, with center wavelengths of 1271, 1291, 1311, and 1331 nm unidirectionally yet still across a pair of MM fibers or SM fibers (hence the universal name).

SWDM4

To round out our third WDM alternative, we look at SWDM4. Using a similar 4-wavelength approach to UNIV, the key difference here is that they launch in the more traditional MM 850 nm region with center wavelengths of 850, 880, 910, and 940 nm with options for both 40 Gb and 100 Gb.

There are WDM options for 200 Gb as well, but the key takeaway at this juncture is that, unlike parallel optics which is a fully ratified IEEE standard, these technologies remain proprietary or as part of a multisource agreement (MSA) today and are not interoperable with each other or any standards-approved transceiver type. BiDi must connect to another BiDi, UNIV to UNIV and so on. This has the potential to severely limit the options available to network managers as they look to migrate to the next speed, often "vendor-locking" them to a single manufacturer and limiting those critical infrastructure choices I referenced earlier.

It's very clearly documented that parallel optics, using lower cost components, are always released into the market first, with their WDM counterparts following 3-5 years later. If you decide to follow a 2-fiber network strategy with WDM and your business/clients demand the next technology upgrade, what choice are you left with if it's not yet available?

Either stall until it's released, but this has revenue and reputation impacts, or upgrade the network to support a parallel optic deployment. However, if you have the fibers in place today (even if you are still following the WDM path), you have the option to use both technologies across the same cabling plant without any unplanned upgrades.

Modulation Techniques

Traditional nonreturn to zero (NRZ) modulation uses two values of 0 and 1. Think of it like a lightbulb being the transmitter and your eyes being the receiver. You simply turn the light, or optical signal "on and off," but if your eyes are as bad as my old things this becomes more difficult to identify or receive as the speeds increase. For 100G, PAM4 (pulse amplitude modulation) was developed which doubled the symbol rate of NRZ and again to simplify the methodology, it uses four discrete values 00, 01, 10 and 11 – turning the light slightly on, fully on, slightly off, and fully off. PAM4 is used on some, but not all, 100G options – it will, however, become the norm for 400G. It's worth noting that, as speeds increase, we may see the introduction of other modulation techniques down the line.

Now to keep this article shorter than a bestselling novel, I've only referenced Ethernet and the different MM options. Please be aware that single-mode (SM) is by no means any easier. For as long as I've walked on this Earth, SM has always been 2 fibers, just like our traditional MM brethren until the advent of 40G. And just like MM, we started to see parallel optic transceiver options for SM from 40G and above as users needed longer distance, but still wanted to take advantage of port breakout capabilities – taking a larger speed port and breaking it into four smaller speeds to leverage a much better total cost of ownership. The shorter distance of 500 m PSM4 vs traditional SM distances of 10 km results in a lower cost.

The current industry trend is that once the volumes start to take hold, the cost per port to deploy a higher speed parallel optic and break it out, vs. using the same native speed optics (1 x 100G broken into 4 x 25G rather than 4 x native 25G, as an example), becomes lower and the higher the optic speed, the more cost effective it becomes. In addition, breakout offers significant density savings, as well as reduced energy costs for both power and cooling.

400G – Where We Are Today

As I mentioned earlier, our insatiable appetite for bandwidth has driven the industry to develop even faster, more latency-independent networks. The proliferation of artificial intelligence (AI) and machine learning (ML), serverless computing, distributed flash storage and faster server CPU/GPU/FPGA are all factors driving 400G. As an example, Broadcom's Tomahawk III ASIC supports 12.8 Tb/s, or 128 x 100G optical signals, and there are developments for 25.6 Tb/s (256 x 100G), 51.2 Tb/s (512 x 100G) and 102.4 Tb/s (256 x 400G). Compressing more and more throughput into the same footprint provides the opportunity to process data much faster than ever before, fulfilling the "Need for Speed" at our fingertips. This doesn't come without its challenges and you'll see in the Table below that we are being presented with several different options to support length, lane and fiber types.

Here is a list of all the current 400G optic PMDs (physical media dependent) that have either been released or are still in development as an IEEE standard:

400G Optics						
IEEE Standard	**Fiber Type**	**Number of Electrical Lanes**	**Number of Fibers**	**Connector**	**MM Link Length**	**SM Link Length**
400GBASE-FR8	Single-Mode	8 WDM	2	Duplex LC	-	2 km
400GBASE-LR8	Single-Mode	8 WDM	2	Duplex LC	-	10 km
400GBASE-SR16	Multimode	16 Parallel	16+16	32 F MPO	100 m	-
400GBASE-DR4	Single-Mode	4 Parallel	4+4	8 F MPO	-	500 m
400GBASE-DR4+/eDR4	Single-Mode	4 Parallel	4+4	8 F MPO	-	2 km
400GBASE-SR8	Multimode	8 Parallel	8+8	16 F MPO	100 m	-
400GBASE-SR4.2	Multimode	8 Parallel	4+4	8 F MPO	100 m	-

In addition to the above, 2016 saw the formation of two MSAs looking at transceiver form factors for 400G with eight electrical lane connections. The form factor MSAs were developed to look at what the connectivity options would be, including the packaging for the transceiver. These are QSFP-DD (quad small-form-factor pluggable – double density) and OSFP (octal small-form-factor pluggable). Corning is a member of both these MSAs.

QSFP-DD - http://www.qsfp-dd.com/

QSFP-DD uses the same physical footprint as QSFP28, so it's fully backward compatible when using the same chassis or line card. In terms of optical interfaces, the most recent specification (Rev 5.0 July 2019) lists the following options:

QSFP-DD	
Male MPO	MPO-12 Single Row
	MPO-16 Single Row
	MPO-12 Double Rows
Duplex LC	
Dual CS	
MDC*	
SN*	

*MDC and SN are a new generation of very-small-form-factor connectors that were launched early in 2019, to be covered in a separate article. These are being driven by hyperscale operators who would like to break out 400G to 4x100G optical circuits directly at the transceiver.

OSFP - https://osfpmsa.org/

The OSFP form factor was designed with higher speeds in mind (however both QSFP-DD and OSFP can handle 800G heat dissipation), and as such has a slightly larger footprint than the QSFP-DD. This is mainly due to an additional set of cooling fins along the top of the body, designed to dissipate higher temperatures generated as a result of the higher power required (note that QSFP-DD has now added to the design to dissipate more heat). These temperature increases are consistent for both QSFP-DD and OSFP with a typical requirement for multimode between 8-15 W. To offer backward compatibility to QSFP28, you need to use an adapter.

The current OSFP specification from January 2019 lists:

OSFP	
Male MPO	MPO-12 Single Row
	MPO-16 Single Row
	MPO-12 Double Rows
Duplex LC	
Dual CS	

The more astute among you may have noticed that the table of PMDs does not align with the options listed within either MSA, and you would be completely accurate. The reality of the matter is that while the different optical interfaces are listed inside the MSA, there are currently no identified optics for MDC or SN, and no identified solution for MM LC, just SM for 500 m, 2 and 10 km.

"So when will we see 400G optics coming to a store near us soon?", I hear you all cry! We saw announcements from Arista and Cisco in 2019 with expectations to have product ready for shipping by the end of 2019 or early into 2020. Arista have announced options for both QSFP-DD and OSFP, and Cisco just for QSFP-DD.

Conclusion

To Infinity and Above, or rather 400G, may seem quite a ways off for many of you reading this article. What we've all come to realize, however, is that we as an IT industry will never stop pushing the envelope and we must all accept that what we love today will soon become the memory of yesterday.

We have experienced more network speed increases in the last nine years than we have in the previous 30 (with more to follow). With so many options, planning the right technology migration path with the most flexible and scalable fiber infrastructure possible should be at the top of your priority list. If you don't, you could find yourself having to continue investing in additional fiber, because the optics to support your next technology refresh are not yet in development.

CORNING

Corning Optical Communications LLC • 4200 Corning Place • Charlotte, NC 28216 USA
800-743-2675 • FAX: 828-325-5060 • International: +1-828-901-5000 • www.corning.com/opcomm