

Coherent Transceivers at a Low Latency

Latency, the time it takes for data to travel from its source to its destination, is a critical metric in modern networks. Reducing latency is paramount in the context of 5G and emerging technologies like edge computing. Smaller data centers placed locally (also called *edge data centers*) have the potential to minimize latency, overcome inconsistent connections, and store and compute data closer to the end-user. Various trends are driving the rise of the edge cloud:

- **5G technology and the Internet of Things (IoT):** These mobile networks and sensor networks need low-cost computing resources closer to the user to reduce latency and better manage the higher density of connections and data.
- **Content delivery networks (CDNs):** The popularity of CDN services continues to grow, and most web traffic today is served through CDNs, especially for major sites like Facebook, Netflix, and Amazon. By using content delivery servers that are more geographically distributed and closer to the edge and the end user, websites can reduce latency, load times, and bandwidth costs as well as increasing content availability and redundancy.
- **Software-defined networks (SDN) and Network function virtualization (NFV):** The increased use of SDNs and NFV requires more cloud software processing.
- **Augment and virtual reality applications (AR/VR):** Edge data centers can reduce streaming latency and improve the performance of AR/VR applications. Cloud-native applications are driving the construction of edge infrastructure and services. However, they cannot distribute their processing capabilities without considerable investments in real estate, infrastructure deployment, and management.

Several of these applications require lower latencies than before, and centralized cloud computing cannot deliver those data packets quickly enough. Let’s explore what changes are happening in edge networks to meet these latency demands and what impact will that have on transceivers.

The Different Latency Demands of the Cloud Edge

As shown in Table 1, a data center on a town or suburb aggregation point could halve the latency compared to a centralized hyperscale data center. Enterprises with their own data center on-premises can reduce latencies by 12 to 30 times compared to hyperscale data centers.

Type of Edge	Data center	Location	Number of DCs per 10M people	Average Latency	Size	
On-premises edge	Enterprise site	Businesses	NA	2-5 ms	1 rack max	
Network (Mobile)	Tower edge	Tower	Nationwide	3000	10 ms	2 racks max
	Outer edge	Aggregation points	Town	150	30 ms	2-6 racks
	Inner edge	Core	Major city	10	40 ms	10+ racks
Regional edge	Regional	Major city	100	50 ms	100+ racks	
Not edge	Hyperscale	State/national	1	60+ ms	5000+ racks	

Table 1: Types of edge data centers and their characteristics. Source: *STL Partners*

Cisco estimates that 85 zettabytes of useful raw data were created in 2021, but only 21 zettabytes were stored and processed in data centers. Edge data centers can help close this gap. For example, industries or cities can use edge data centers to aggregate all the data from their sensors. Instead of sending all this raw sensor data to the core cloud, the edge cloud can process it locally and turn it into a handful of performance indicators. The edge cloud can then relay these indicators to the core, which requires a much lower bandwidth than sending the raw data.

Using Coherent Technology in the Edge Cloud

As edge data center interconnects became more common, the issue of interconnecting them became more prominent. Direct detect technology had been the standard in data center interconnects. However, reaching distances greater than 50km and bandwidths over 100Gbps required for modern edge data center interconnects required external amplifiers and dispersion compensators that increased the complexity of network operations.

At the same time, advances in electronic and photonic integration allowed longer-reach coherent technology to be miniaturized into QSFP-DD and OSFP form factors. This progress allowed the Optical Internetworking Forum (OIF) to create the 400ZR and ZR+ standards for 400G DWDM pluggable modules. With modules that are small enough to pack a router faceplate densely, the datacom sector could profit from a 400ZR solution for high-capacity data center interconnects of up to 80km. The rise of **100G ZR technology** takes this philosophy a step further, with a product aimed at spreading coherent technology further into edge and access networks.

How Does Latency in the Edge Affect DSP Requirements?

The traditional disadvantage of coherent technology vs direct detection is that coherent signal processing takes more time and computational resources and, therefore, introduces more latency in the network. Adapting to the latency requirements of the network edge might, especially in shorter link distances, might require digital signal processors (DSPs) to adopt a “lighter” version of the signal processing normally used in coherent technology.

Let’s give an example of how DSPs could behave differently in these cases. The quality of the light signal degrades when traveling through an optical fiber by a process called *dispersion*. The same phenomenon happens when a prism splits white light into several colors. The fiber also adds other distortions due to *nonlinear optical effects*.

These effects get worse as the input power of the light signal increases, leading to a trade-off. You might want more power to transmit over longer distances, but the nonlinear distortions also become larger, which beats the point of using more power. The DSP performs several operations on the light signal to try to offset these dispersion and nonlinear distortions.

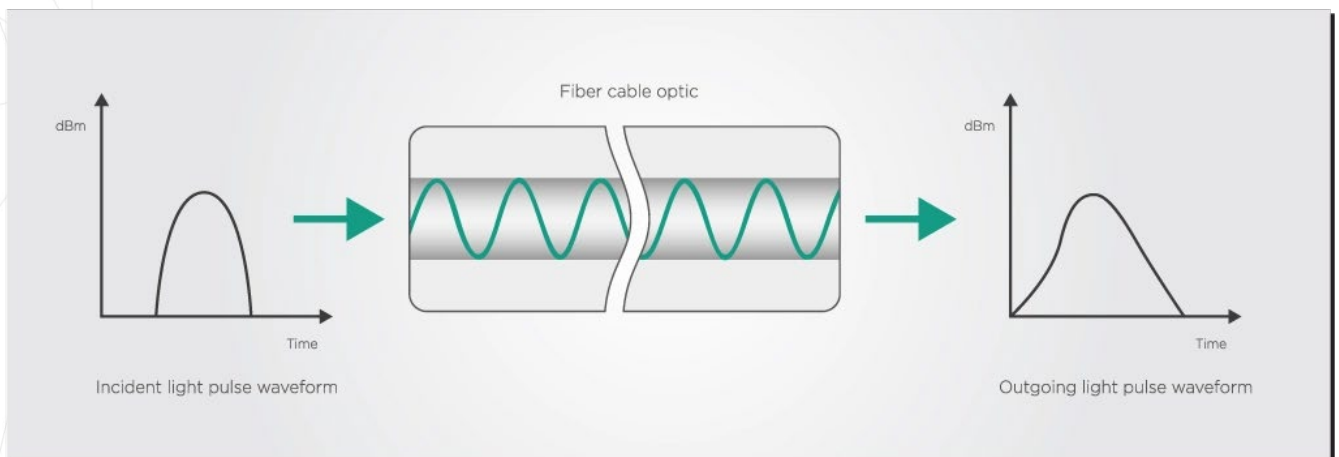


Figure 1: Explaining dispersion compensation in optical fibers. A pulse of light will spread out and distort due to dispersion effects.

However, shorter-reach connections require less dispersion compensation, presenting an opportunity to streamline the processing done by a DSP. A lighter coherent implementation could reduce the use of dispersion compensation blocks. This significantly lowers system power consumption and latency too.

Another way to reduce processing and latency in shorter-reach data center links is to use less powerful forward error correction (FEC) in DSPs. You can learn more about FEC in [one of our previous articles](#).

Takeaways

The shift towards edge data centers tries to address the low latency requirements in modern telecom networks. By decentralizing data storage and processing, closer to the point of use, edge computing reduces latency but also enhances the efficiency and reliability of network services across various applications, from IoT and 5G to content delivery and AR/VR experiences.

The use of coherent transceiver technology helps edge networks span longer reaches with higher capacity, but it also comes with the trade-off of increased latency due to more signal processing from the DSP. This scenario means that DSPs will have to reduce the use of certain processing blocks, such as dispersion compensation and FEC, to meet the specific latency requirements of edge computing.

More Articles from EFFECT Photonics