Fiber Optic Communication Technologies: What's Needed for Datacenter Network Operations

Cedric F. Lam, Hong Liu, Bikash Koley, Xiaoxue Zhao, Valey Kamalov, and Vijay Gill, Google Inc.

ABSTRACT

In this article we review the growing trend of warehouse-scale mega-datacenter computing, the Internet transformation driven by mega-datacenter applications, and the opportunities and challenges for fiber optic communication technologies to support the growth of mega-datacenter computing in the next three to four years.

INTRODUCTION

The last decade has seen tremendous growth in the deployment of broadband access networks around the world. The proliferation of access bandwidths offered by technologies such as fiber to the home (FTTH) has led to the mushrooming of many new web applications, from traditional search to online interactive maps, social networks, office software, video streaming, mobile Internet, and so on. Most of these free-of-charge applications are running in datacenters that are transparent to end users. Datacenter computing and content streaming are becoming more and more popular, and rapidly growing. This trend is driving a transformation of our modern Internet.

The two diagrams in Fig. 1 were extracted from the ATLAS Internet Observatory 2009 annual report [1]. Figure 1a is the hierarchical view of the early Internet with a few large core service providers at the top level providing backbone interconnects to the regional access providers and local Internet service providers (ISPs) at the lower levels of the hierarchy. Today's Internet, which is shown in Fig. 1b, has transformed into a more meshed connectivity model. The central core of the Internet, which was dominated by traditional backbone providers, is now connected by hyper giants offering rich content, hosting, and content distribution network (CDN) services. It is not difficult to imagine that the network is moving toward more and more direct connection from content providers to content consumers, with the traditional core providers facing disintermediation.

Table 1 lists the ATLAS top 10 interdomain autonomous systems (ASs) in the public Internet in 2007 and 2009. We see that content providers such as Google and Comcast, which were not ranked in 2007, occupied prominent places in 2009. It should be noticed that these reports only account for publicly measurable bandwidth between ASs where the measurements were taken. Left uncounted here are three types of traffic:

- Traffic inside datacenters
- The backend bandwidths used to interconnect datacenters and operate the content distribution networks
- Virtual private network (VPN) traffic

These data demonstrate the transformation from the original focus on network connectivity by traditional carriers to a focus on content by the non-traditional companies. New Internet applications such as cloud computing and CDN are now reshaping the network landscape. Content providers and cloud computing operators such as Amazon, Microsoft, and Google have now become the major driving forces behind large-capacity optical network deployments. Most of these network demands, although invisible to market research firms, have been growing very quickly, as demonstrated by the ATLAS report.

According to TeleGeography Research's report [2], between 2002 and 2008, total private network deployments increased at a compound annual rate of 47 percent. At the end of 2008, private networks accounted for 20 percent of international bandwidth usage. The report also says: "increasingly, these entities have capacity requirements that are similar to those of the largest carriers."

The rest of the article is organized as follows. We give a high-level view of Internet-style computing and its benefits. We begin with a brief description of warehouse-scale computers (WSCs), and discuss the needs and challenges of intra-datacenter optical interconnects to support WSC infrastructure. We focus on the characteristics and requirements of an inter-datacenter long-haul optical network. Lastly, we conclude with the optical communication technologies that should be considered in order to sustain the growth of warehouse-scale computing.

0163-6804/10/$25.00 © 2010 IEEE

Internet-Style Computing

Figure 2 shows a high-level view of today's Internet-style computing. At the center is a cloud of geographically distributed mega datacenters connected by a large-capacity network. A user accesses the cloud through his/her local network service provider's carrier network, which interconnects with the datacenter network through Internet points of presence (POPs). As far as a user is concerned, the datacenter network appears as a single computer. The cloud infrastructure, which consists of the network and multiple warehouse-scale datacenters, is transparent to the user.

Internet-style computing provides many benefits. Scalability is the first one. All user data resides in the network, and are accessible anywhere in the world as long as there is a network connection. The data are automatically backed up, so there is no need to worry about losing data to facility or hardware failure. Internet-style computing also provides a platform for easy data sharing and collaboration. Google Documents for example, allow multiple collaborators to view and edit the same document simultaneously. Since end users need not be concerned with maintaining their own computing equipment, there is operational expense and capital reduction [3]. Internet-style computing results in better utilization of computing resources via technologies such as virtualization.

Without a good reliable and scalable network infrastructure, none of the above benefits can happen.

Intra-Datacenter Communications

A datacenter is a massively parallel super-computing infrastructure [4], which consists of clusters with thousands of servers networked together. To optimize performance and power in a cost-efficient manner, datacenter equipment is usually made of commercial off-the-shelf components so that they can take the advantage of the economy of scale of consumer commodities. Figure 3 shows a typical datacenter cluster, with servers arranged into racks of 20-40 machines each. Servers within the same rack are connected through a top of rack (TOR) switch. Rack switches are connected to cluster switches which provide connectivity between racks and form the cluster-fabrics for warehouse-scale computing.

Figure 4 shows pictures of a WSC: an overview of a WSC datacenter, servers in equipment racks, and interconnect cables in a WSC. Obviously, datacenter operators would welcome innovations that could simplify all the wiring shown in Fig. 4c. (For a good introduction of WSC, please refer to [4].)

Ideally, one would like to have a fully meshed intra-datacenter network that directly connects every server to every other server in a datacenter so that application software does not need to be concerned with the locality of the machines they use to distribute computation jobs. However, such a design would be prohibitively expensive.

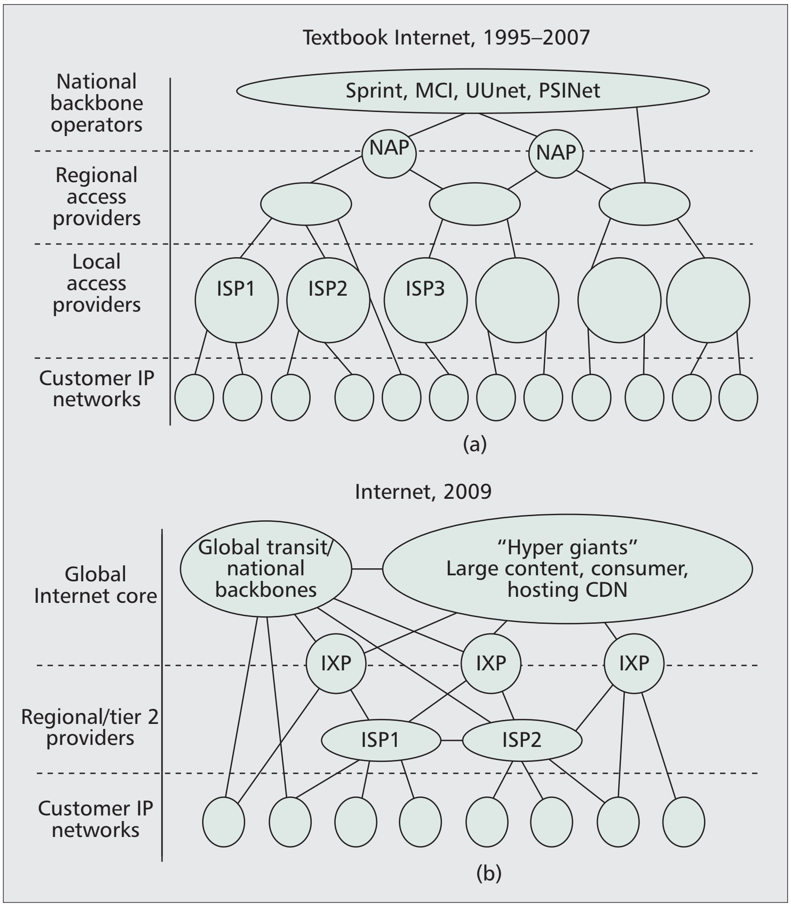

Textbook Internet, 1995–2007
National backbone operators
Sprint, MCI, UUnet, PSINet
NAP
NAP
Regional access providers
Local access providers
ISP1
ISP2
ISP3
Customer IP networks
(a)
Internet, 2009
Global Internet core
Global transit/ national backbones
"Hyper giants" Large content, consumer, hosting CDN
IXP
IXP
IXP
Regional/tier 2 providers
ISP1
ISP2
Customer IP networks
(b)

Figure 1. *Internet evolution [1]: a) 1995-2007; b) 2009.*

(a) Top ten, 2007			(b) Top ten, 2009		
Rank	Provider	%	Rank	Provider	%
1	Level(3)	5.77	1	Level(3)	9.41
2	Global Crossing	4.55	2	Global Crossing	5.7
3	ATT	3.35	3	**Google**	5.2
4	Sprint	3.2	4		
5	NTT	2.6	5		
6	Cogent	2.77	6	**Comcast**	3.12
7	Verizon	2.24	7		
8	TeliaSonera	1.82	8		
9	Savvis	1.35	9		
10	AboveNet	1.23	10		

Table 1. *ATLAS top 10 public Internet bandwidth generating domains [1].*

In reality, cluster interconnections are aggregated with hierarchies of distributed switching fabrics (Fig. 5) [5]. Intra-datacenter switching fabrics have ultra-large capacity compared to metro and long-haul networks. In addition, the

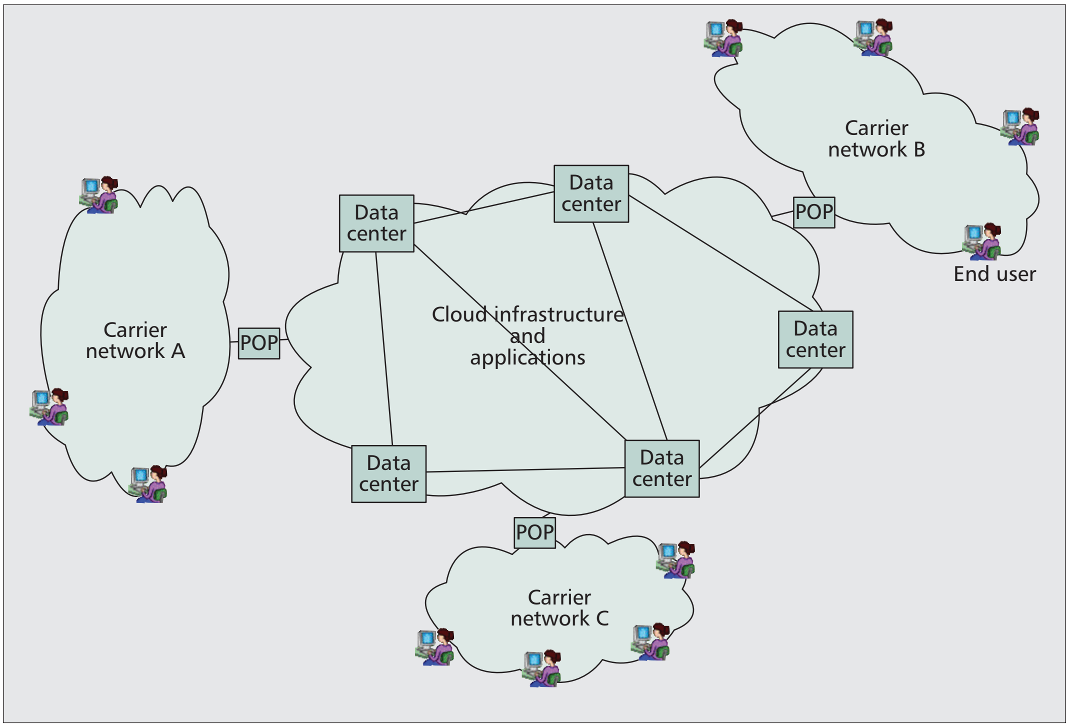


Figure 2. *High-level view of Internet-style computing.*

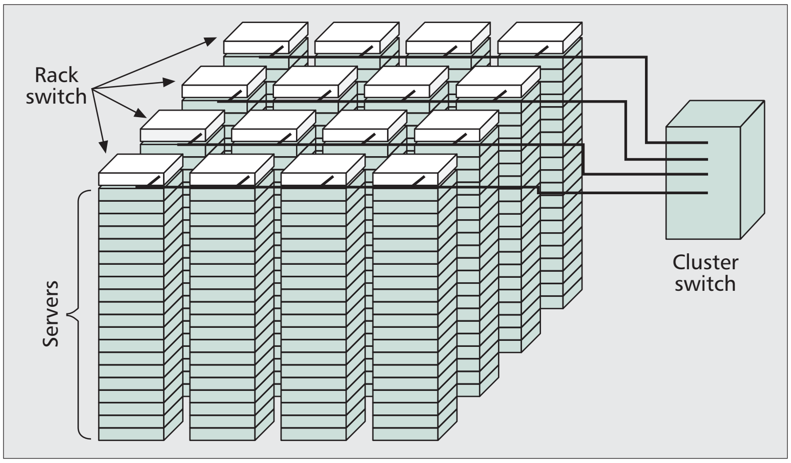


Figure 3. *Typical view of a cluster inside a warehouse-scale datacenter.*

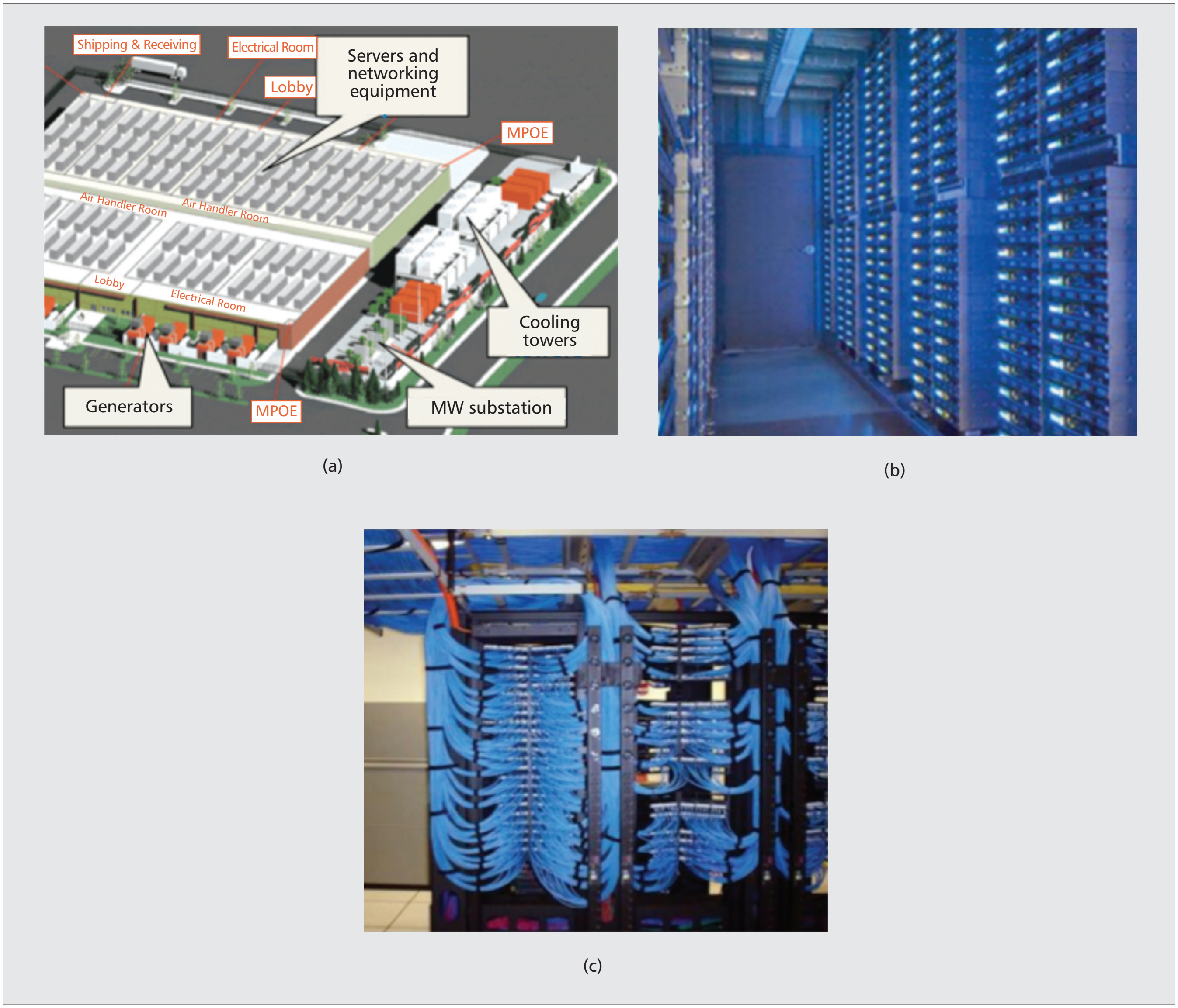
intra-datacenter environment is a fiber-rich environment, where spectral efficiency is often not an important concern. Notice that a mega datacenter could consist of either one building or multiple buildings, as shown in Fig. 5. The reach requirements for intra-datacenter optical interconnect ranges from 10 m to 10 km.

There are multiple topologies in which intra-datacenter cluster fabrics can be formed [6–8]. Commonly seen topological structures for cluster interconnects include torus, hyper cube [6], fat tree [7], and flattened butterfly [8]. No matter what topology is chosen for intra-cluster connectivity, the port density will eventually limit the number of servers that can be connected to a switch, and the scaling of the centralized controller will limit the number of nodes that can be managed in one cluster. Therefore, maximizing the switch port density and the number of nodes of a cluster is an important concern for intra-datacenter networking equipment design.

For illustration purposes, an example of a state-of-the-art cluster switch (DS3456) made by Sun Microsystems (now part of Oracle) is shown in Fig. 6 [9]. This switch, which occupies a space of roughly two racks, contains 1152 Infini-Band (IB) ports. Each IB port uses a 60 Gb/s CXP module, which employs 12 5 Gbs/s (DDR rate) parallel optical lanes and ribbon fiber connections. Each CXP can be fanned out to three 20 Gb/s DDR IB links to connect three server nodes via a 4 × 5 Gb/s QSFP transceiver. In total, 3456 server nodes can be connected to such a switch. This switch has a total bisectional raw bandwidth of 55 Tb/s. In a datacenter environment such switches can be further connected in stages to form larger fabrics using the various topologies mentioned in the last paragraph. One striking feature of the DS3456 switch is that the whole front face is almost fully covered with IB transceiver modules. Transceiver port density therefore limits the size of the switches we can build and the number of hosts that can be connected to a switch.

Integration can help improve port density and system throughput. Figure 7 shows the evolution of short-reach optical transceivers [10] from a single 10 Gb/s SFP+ to a 4 × 10G QSFP module with four 10 Gb/s parallel optical channels, and a CXP active optical cable (AOC) with 12 parallel 10 Gb/s channels.

Ideally, to take full advantage of the integration, as the size of modules scales down, the power of the modules should also scale down proportionally so that the total power dissipation per unit space does not change when we cram more transceivers into the same chassis space. Otherwise, scaling the size of the modules themselves would not give us the needed port count



(a)

(b)

(c)

Figure 4. *Views of a warehouse-scale computer: a) overview of a WSC datacenter; b) servers in equipment racks; c) interconnect cables in a WSC.*

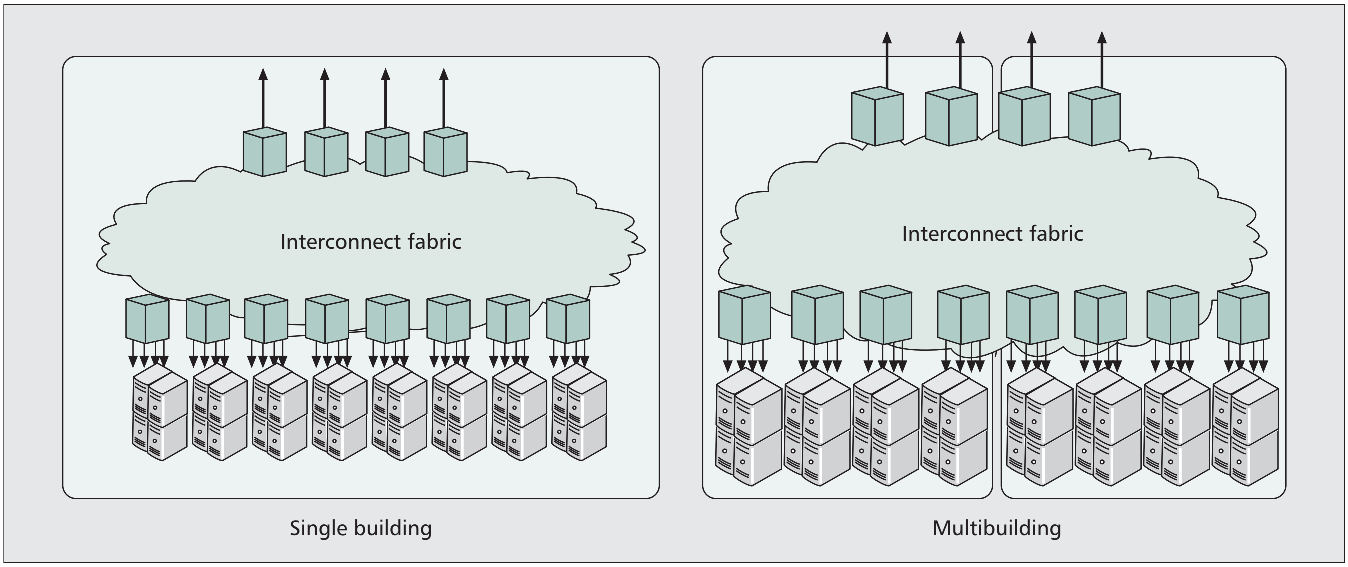


Figure 5. *Hierarchies of intra-datacenter cluster-switching interconnect fabrics.*

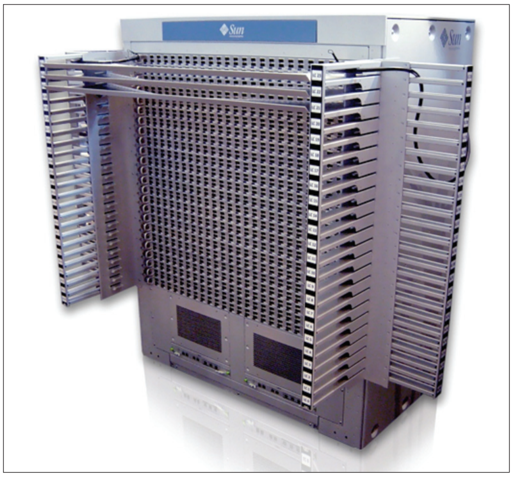


Figure 6. *Sun DS3456 cluster switch.*

benefit. In Fig. 7d the dotted line shows 1:1 power scaling of the modules with port speed. The solid line shows the actual power scaling. The integration of electronic driving, receiving and management circuits has clearly helped to scale the power to be less than 1:1.

Parallel optics has been playing an important role in intra-datacenter communication. It takes advantage of the low cost and low power consumption of vertical-cavity surface-emitting lasers (VCSELs), along with the natural parallelism of data lanes in computer architecture and warehouse scale computing. In addition, the electrical signal rate aligns well with the optical signal rate; no expensive and power-hungry SerDes or gear box is needed for data-rate conversion. The disadvantage is that the MTP/MPO fiber termination cost is expensive, and ribbon fibers are required for external connections. To further scale the density and reach a higher data rate, a photonic integration circuit (PIC) using either wavelength-division multiplexing (WDM) or silicon photonics is inevitable.

To satisfy the continuing growth of bandwidth demand in datacenter operation, the data rate, power, and space density all need to scale. In the next three to four years, datacenter operators would like to see the speed of optical transceivers scaled up by a factor of 4 while keeping the power and space profile unchanged. Such scaling, however, will demand new technologies to be implemented for short-reach interconnect. Traditional on-off-keying (OOK) modulation will perform up to 20 Gb/s for VCSELs. Beyond that, electronic and optical component performances become demanding in both bandwidth capability and dispersion. To keep the 4× and 10× transmission-speed scaling, new technologies are needed in transceiver modules. Examples of these technologies include electronic dispersion compensation, integrated low-power silicon photonics, and new signal modulation schemes.

In the long run, according to the International Technology Roadmaps for Semiconductors (ITRS) projection [11], the processing power on digital complementary metal oxide semiconductor (CMOS) chips will significantly outstrip the total off-chip input/output (I/O) bandwidth in 2022 because of the limitations in pin count, clock speed, and power dissipations. Photonic integration holds the promise to bridge this gap. Optical interconnect can not only run at much higher clock rate, but also make use of the third dimension for interconnect. Hybrid optical interconnect at 10 Gb/s speed on a multi-chip-module carrier has already been demonstrated by IBM Labs [12], which extends interconnects to the third dimension. Researchers around the world are investigating the techniques to incorporate a photonic layer on future silicon chips, which will help alleviate the bottleneck of off-chip bandwidth.

Inter-Datacenter Communications

As shown in Fig. 2, datacenters are geographically distributed. They are usually located near power stations in rural areas where the land is cheap, and the power is abundant and readily available at low cost. Distributed datacenters not only provide redundancy, but are also needed for load balancing the computation needs, and improve customer experience by reducing transit latency.

Long-haul optical networks are needed to get traffic from datacenters to population centers where users are located. These networks are built to reduce interconnect bandwidth costs, move data between remote datacenter locations, and ensure scalable capacity for datacenter operations. The growing deployment of new broadband access network infrastructures such as FTTH will only exacerbate bandwidth requirements for inter-datacenter long-haul and metro networks.

Unlike in the intra-datacenter environment where fiber is readily available, in long-haul networks fiber is very scarce between datacenter locations (partly due to the remote locations where they are situated). Long-haul transmission fibers are very expensive and time consuming to build or acquire. The traffic between datacenters is mostly machine generated by cloud computing applications, and the volume is huge. Therefore, capacity and spectral efficiency are very important for inter-datacenter communication links. Second, as datacenters are usually located in remote areas, ultra-long transmission distances, from 500 to 6000 km with minimum regeneration are needed. Regenerators costs money and increases operational expenditure. Sometimes, there is simply no space and power for regeneration purposes between remote datacenters.

Third, unlike traditional telco networks that require a lot of intermediate add/drops, datacenter optical networks are mostly point-to-point *fat-pipe* connections with few intermediate add/drops. Capacity and reach are much more important than optical layer flexibility.

The common requirement shared by carrier networks and datacenter networks is smooth upgrade of long-haul capacity.

From an economic perspective, there is tradeoff among cost, reach, capacity, and spectral efficiency. Where fiber is abundant, there is no point in paying for the premium of new cutting-

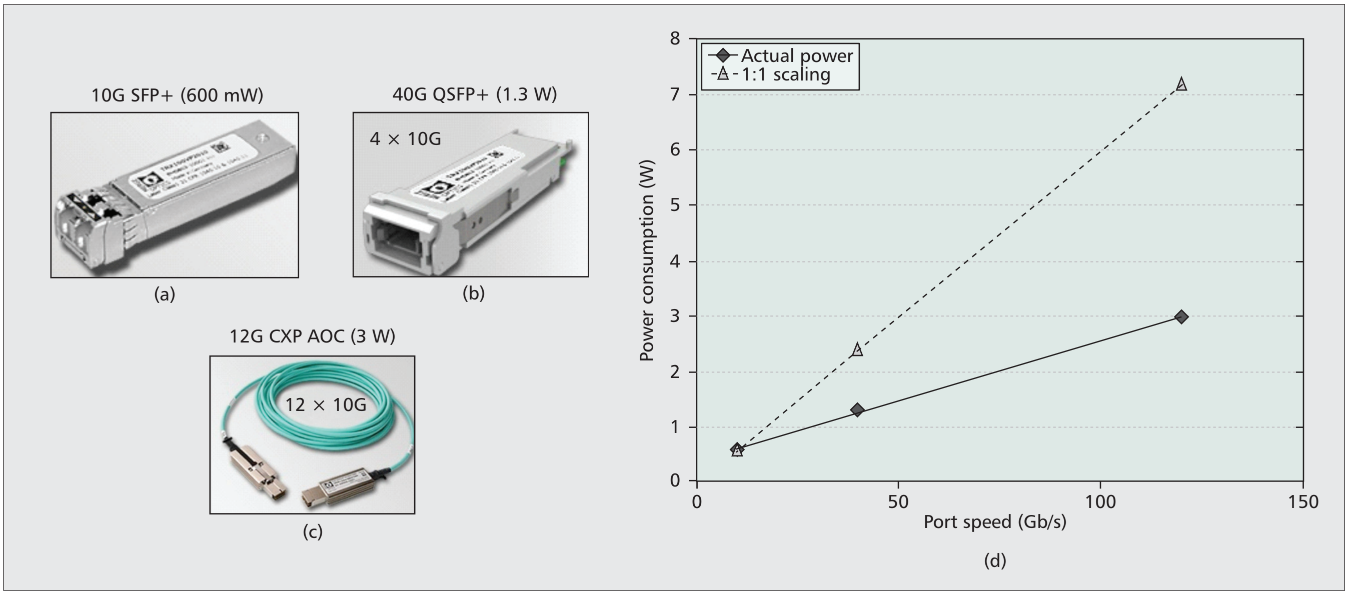


Figure 7. *a) A single 10G SFP+ module; b) 40 Y 10G QSFP module with four 10 Gb/s parallel optical channels; c) a CXP active optical cable with 12 parallel 10 Gb/s channels; d) power scaling with respect to port speed (courtesy of Merge Optics).*

edge technology to achieve the ultimate transmission bandwidth and distance. For long-haul networks where fiber is scarce, it makes sense to adopt the latest transmission technology. The coherent optical receiver is especially important for sustaining the growth of next-generation datacenter long-haul networks. Compared to traditional optical modulation formats, digital signal processing (DSP)- enabled coherent optical receivers provide not only better spectral efficiency but also better tolerance to amplitude spontaneous emission (ASE) noise, chromatic dispersion, and polarization mode dispersion (PMD). This helps maximize both system reach and capacity. Such systems also enable the use of legacy *bad* fibers with excessive PMD values.

For any network operator, the ultimate design goal is to maximize the total capacity and minimize the cost per bit. Shannon showed that in an additive white Gaussian noise (AWGN) limited channel, information capacity is linearly proportional to channel bandwidth but logarithmically proportional to signal-to-noise ratio (SNR) [13]. It is therefore much easier to increase the capacity by exploring more fiber spectrum in fiber than increasing the SNR or using advanced modulation formats with high-order constellations such as *M*-quadrature amplitude modulation (QAM). Most of today's commercial WDM fiber optic transmission systems make use of the C-band for transmission because nature has blessed C-band with the lowest fiber loss and the most mature low-noise wide-band optical Erbium doped fiber amplifier (EDFA) at very low cost. These properties serve to maximize the signal-to-noise ratio in the Shannon formula

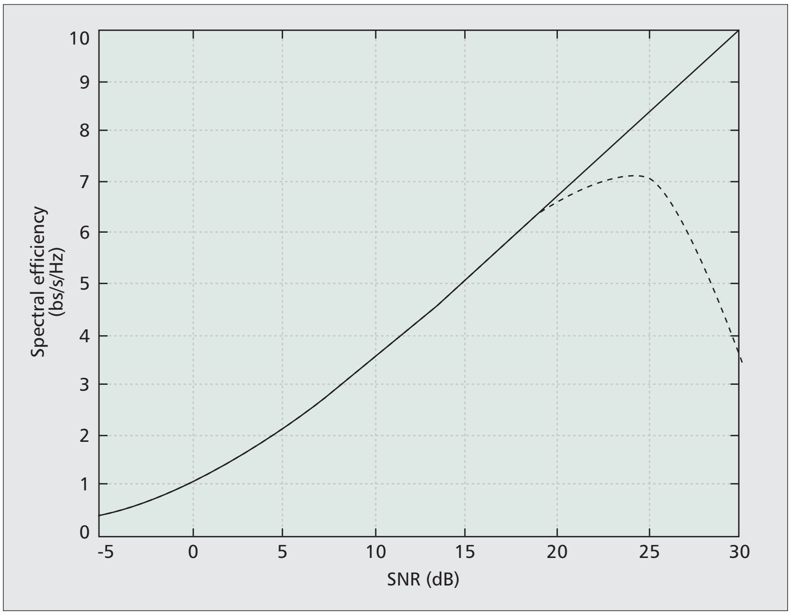
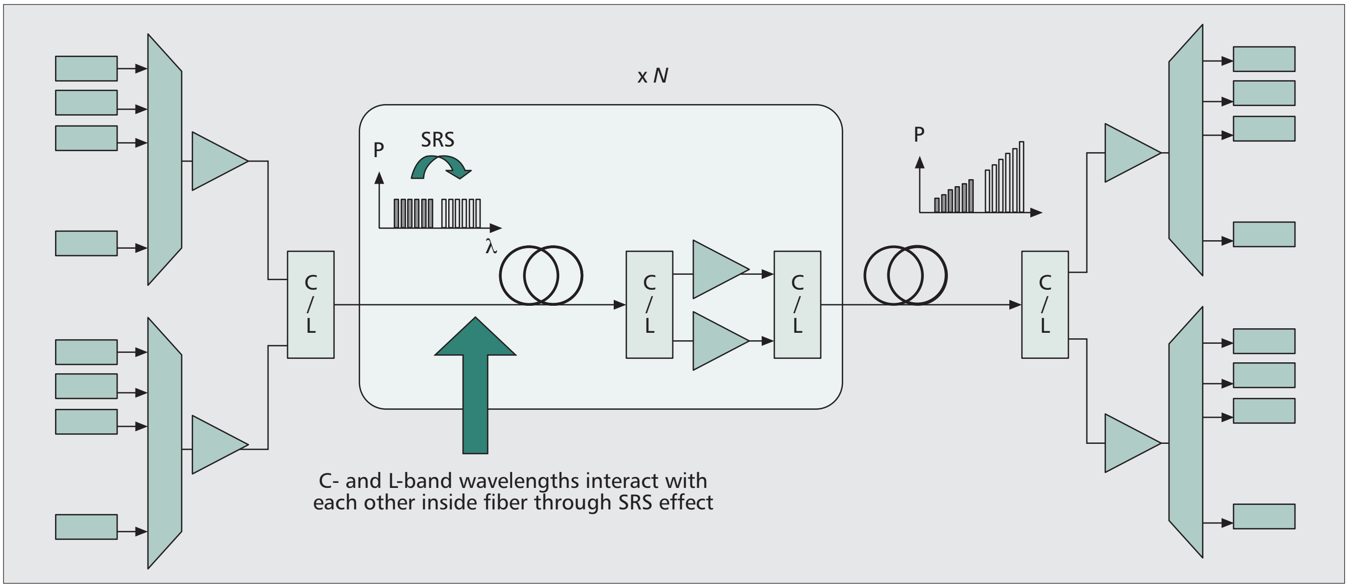


Figure 8. *Spectral efficiency vs. SNR: Shannon limit (solid line) and fiber nonlinearity limit (dotted line).*

L-band is the next natural spectral region to use besides C-band and is nothing new to fiber optic transmission engineers. EDFAs can be made to operate at L-band, and fiber loss in L-band is still very low. In fact, recent laboratory experiments with record transmission capacities were performed using both C- and L-bands [14, 15]. Besides, there is a significant embedded base of deployed large effective area fiber (LEAF) in the field. LEAF has very small chromatic dispersion at C-band wavelengths, which leads to high nonlinear penalty. As a result, L-band wavelengths perform better than C-band wavelengths in LEAF fiber because of the higher dispersion in L-band and thus lower nonlinearity.

Eventually, fiber nonlinearity limits the achievable SNR and ultimate capacity [16] inside the optical fiber (Fig. 8). Today's state-of-the-art commercial systems with 80 channels of 50 GHz spaced 100GbE transmissions achieve a total capacity of 8 Tb/s, or an equivalent spectral effi-



C- and L-band wavelengths interact with each other inside fiber through SRS effect

Figure 9. *Stimulated Raman scattering induced spectral tilt in a C+L-band system.*

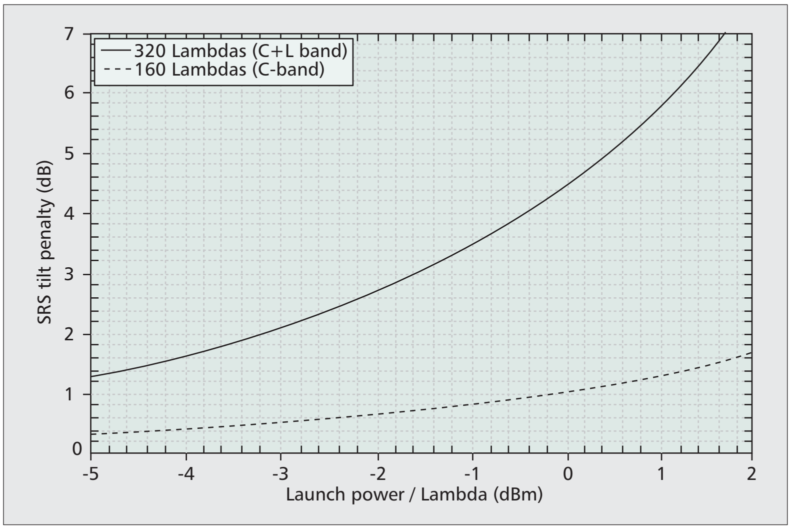


Figure 10. *Raman gain tilt vs. per wavelength launch power in a 70km span of LEAF fiber. Channel spacing is 25GHz.*

ciency of 2 b/s/Hz. Simulation results show that we are not far from reaching the ultimate nonlinearity limited C-band capacity [16].

One way to increase SNR, reach, and capacity, and lower fiber nonlinearity is to deploy low-loss large-core pure Silica optical fibers in new builds. Such fibers have been widely used in submarine transmission networks already.

Another key technology to preserve SNR and reduce fiber nonlinearity is distributed Raman amplification, which helps to extend unregenerated transmission distances. With Raman amplification, one can reduce the optical launch power and hence lower optical Kerr nonlinearity effects. One nice feature about Raman amplification is that it operates at any wavelength, and can be used to open up new transmission spectrum besides C-band and L-band.

However, the same Raman effect that gives us all these benefits can also cause a gain tilt problem if not managed properly. The Raman amplification effect happens between any two wavelengths in fiber. Shorter wavelength signals will amplify longer wavelength signals and cause spectral tilt (Fig. 9). Unlike other fiber nonlinear effects, stimulated Raman gain increases with channel spacing. It also increases rapidly with signal launch power and number of wavelengths in fiber.

Figure 10 shows the stimulated Raman scattering (SRS) tilt plotted against launch power for a 70 km, 25 GHz spaced LEAF link, with only C-band wavelengths and C+L-band wavelengths, respectively. It is striking that the introduction of L-band wavelengths significantly increases the Raman gain tilt. At –1 dBm launch power, the spectral tilt after propagating the 70 km span is > 3.5 dB when both C- and L-band wavelengths are present. Gain tilt control should therefore be taken into account in the initial design for a system intended for future L-band upgrade.

Research also indicated that 400 Gb/s per wavelength transmission might be the highest practical limit for single-carrier modulation [17]. To go beyond 400 Gb/s channel speed, one may have to resort to novel multicarrier modulation schemes such as orthogonal frequency-division multiplexing (OFDM). For example, in one demonstration researchers at Bell Labs demonstrated 1.2 Tb/s OFDM with 300 GHz bandwidth [18].

Current long-haul fiber optic transmission systems artificially slice optical spectrum into International Telecommunication Union (ITU) grids. To support future ultra-wideband multicarrier OFDM transceivers, it is important that the transmission line be free of bandwidth-limiting optical elements such as channelized reconfigurable optical add/drop multiplexers (ROADMs). Spectrum is a precious resource, as indicated by the linear relationship in the Shannon formula. So to ensure future capacity growth for the Internet, it is important that one does not limit himself/herself by putting bandwidth-limiting elements such as channelized ROADM devices on the transmission line [19, 20], especially since we are quickly approaching the capacity limit of C-band.

Conclusion

To conclude, fiber optic technologies provide the bloodstream for datacenter operations. Datacenter networks use the full spectrum of fiber optic technologies, from short reach to long haul. For short-reach intra-datacenter interconnects, power, cost, and space profiles are critical for the continual scalability of warehouse-scale computers. Potential technologies to realize next-generation intra-datacenter interconnects include photonic integrated circuits, advanced signal modulation, and electronic dispersion compensation, among others.

On the long-haul side, next-generation long-haul networks need to move away from channelized ROADM systems. To pave the way for future modulation schemes such as optical OFDM, an open wide-band spectral plan should be considered as opposed to the ITU grid channel plan. Commercial equipment vendors should increase their efforts in engineering transmission systems beyond C-band, and infrastructure providers should deploy low-loss, low-nonlinearity, and large-core optical fibers in new fiber builds.

References

[1] C. Labovitz *et al.*, *ATLAS Internet Observatory 2009 Annual Report*; http://www.nanog.org/meetings/nanog47/presentations/Monday/Labovitz_ObserveReport_N47_Mon.pdf

[2] http://www.telegeography.com/products/global-bandwidth-research-service/analysis/supply-and-demand/index.html

[3] D. Patterson, “General Session Keynote, Cloud Futures 2010,” http://perspectives.mvdirona.com/2010/05/04/PattersonOnCloudComputing.aspx

[4] L. A. Barroso and U. Hölzle, *The Datacenter as a Computer — An Introduction to the Design of Warehouse-Scale Machines*, Morgan & Claypool, 2009. http://www.morganclaypool.com/doi/pdf/10.2200/S00193ED1V01Y200905CAC006

[5] B. Koley, “Requirements for Data Center Interconnects,” paper TuA2, *20th Annual Wksp. Interconnections within High Speed Digital Systems*, Santa Fe, New Mexico, 3–6 May 2009.

[6] K. Hwang, *Advanced Computer Architecture: Parallelism, Scalability, Programmability*, McGraw-Hill, 1993.

[7] C. E. Leiserson, *IEEE Tran. Computers*, vol. C-34, no. 10, 1985, pp. 892–901.

[8] J. Kim, W. Dally, and D. Abts, “Flattened Butterfly: A Cost-efficient Topology for High-Radix Networks” *ISCA’07: Proc. 34th Annual Int’l. Symp. Computer Architecture*, 2007, http://cva.stanford.edu/publications/2007/MICRO_FBFLY.pdf, pp. 126–37.

[9] J. Anderson and M. Traverso, “Optical Transceivers for 100 Gigabit Ethernet,” *IEEE Commun. Mag.*, vol. 48, no. 3, Mar. 2010, pp. S35–S40; see also http://www.oracle.com/us/products/serversstorage/networking/infiniband/036558.pdf

[10] http://www.mergeoptics.com

[11] D. Miller, “Device Requirements for Optical Interconnects to Silicon Chips,” *Proc. IEEE ’97*, 2009, pp. 1166–85.

[12] B. J. Offrein (IBM), ECOC 2009.

[13] C. Shannon, *Bell Sys. Tech. J.*, vol. 27, 1948, p. 379.

[14] A. Sano *et al.*, *OFC/NFOEC 2010*, paper PDPB7.

[15] X. Zhou *et al.*, *OFC/NFOEC 2010*, paper PDPB9.

[16] R.-J. Essiambre *et al.*, *OFC/NFOEC 2009*, Paper OThL1.

[17] T. Pfau *et al.*, *J. Lightwave Tech.*, vol. 27, no. 8, Apr. 2009, pp. 989–99; Also, see. T Pfau, “Hardware Requirements for Coherent Systems beyond 100G,” *ECOC 2009*, WS1, DSP & FEC.

[18] S. Chandrasekhar, ECOC 2009, paper PD2.6.

[19] C. F. Lam and W. I. Way, “A System’s View of Metro and Regional Optical Networks,” *Photonics West*, San Jose, CA, Jan. 29, 2009.

[20] M. Jinno *et al.*, *IEEE Commun. Mag.*, Nov. 2009, pp. 66–73.

Biographies

Cedric F. Lam (clam@google.com) is currently an optical network architect at Google. Before joining Google, he worked at OpVista Inc. as chief system architect, responsible for the development of an ultra-dense WDM transport system with integrated ROADM functionality. Prior to OpVista, he was a senior technical staff member at AT&T Labs-Research. His research covers broadband optical transport and access networks architectures, optical signal modulation and transmission, passive optical network, HFC, and more. His current focus is in FTTH and optical networking technologies for data center applications. He received his B.Eng. in electrical and electronic engineering from the University of Hong Kong with First Class Honors and his Ph.D. in electrical engineering from the University of California at Los Angeles.

Hong Liu is a member of technical staff at Google Platform Advanced Technology, where she is involved in the system architecture and interconnect for a large-scale computing platform. Her research interests include interconnection networks, high-speed signaling, and optical metro design. Prior to joining Google, she was a member of technical staff at Juniper Networks, where she was principally involved in the architecture and design of high-end physical interface cards, network core routers, and multi-chassis switches, including Juniper’s flagship core router T640, the world’s first OC768 line card, and the world’s very first switch-matrix, TX640. She received her Ph.D. in electrical engineering from Stanford University.

Bikash Koley is currently a senior network architect at Google, where he is focused on network infrastructure scaling, optimization, and reliability. Prior to joining Google, he was the chief technology officer of Qstreams Networks, a company he co-founded. He also spent several years at Ciena Corporation in various technical roles developing DWDM and Ethernet technologies. He has published many articles and papers, and regularly speaks at conferences and industry events. He received M.S. and Ph.D. degrees from the University of Maryland at College Park and a B.Tech. from the Indian Institute of Technology, all in electrical engineering.

Xiaoxue Zhao is a senior network engineer at Google, whre she focuses on network architecture, design, and planning. Her research interests also include various advanced optical technologies for datacenter networks. She was a research assistant at the University of California at Berkeley prior to joining Google. She received her Ph.D. degree in electrical engineering from the University of California at Berkeley.

Valey Kamalov is a staff optical transport engineer at Google, where he is focused on the performance of the optical network. Prior to joining Google, he spent seven years at Nokia Siemens Networks designing optical networks. He received B.S. and Ph.D. degrees from Moscow University in quantum electronics, and his Sc.D. degree in chemical physics from the Russian Academy of Sciences.

Vijay Gill is senior manager, Engineering and Architecture at Google. He is responsible for all network design, expansion and datacenter infrastructure for Google’s production network, as well as participating in various industry organizations and advancing the company’s efforts in the standards arena. He has co-authored a variety of RFCs on traffic engineering, multihoming, and routing. He has also given talks and presentations on network design, BGP scaling issues, and traffic engineering in forums such as NANOG and IETF. He is also currently serving on the IETF Internet Architecture Board (IAB). Prior to joining Google, he worked as senior technical manager for AOL Global Network Operations and was responsible for setting the technical direction and strategy for AOL production. Before AOL, he worked as manager of architecture at MFN/Abovenet where he participated in revamping the global backbone, standardization of routing policy, and product development. Earlier in his career, he worked as a senior engineer at UUNET, participating in the MPLS and multicast engineering projects.