# From Scaling Disparities to Integrated Parallelism: A Decathlon for a Decade

Peter J. Winzer, *Fellow, IEEE*, and David T. Neilson, *Fellow, IEEE*

*(Invited Tutorial)*

***Abstract*—Based on a variety of long-term network traffic data from different geographies and applications, in addition to long-term scaling trends of key information and communication technologies, we identify fundamental scaling disparities between the technologies used to generate and process data and those used to transport data. These disparities could lead to the data transport network falling behind its required capabilities by a factor of approximately 4 every five years. By 2024, we predict the need for 10-Tb/s optical interfaces working in 1-Pb/s optical transport systems. To satisfy these needs, multiplexing in both wavelength and space in the form of a wavelength-division multiplexing × space-division multiplexing matrix will be required. We estimate the characteristics of such systems and outline their target specifications, which reveals the need for very significant research progress in multiple areas, from system and network architectures to digital signal processing to integrated arrayed device designs, in order to avoid an optical network capacity crunch.**

***Index Terms*—Access, arrays, datacenters, digital signal processing, integration, microprocessors, mobile, network traffic, routers, Shannon limit, space-division multiplexing (SDM), storage, superchannels, supercomputers, technology scaling, video, wavelength-division multiplexing (WDM).**

## I. Network Traffic Growth

NETWORK traffic has been increasing at an enormous pace, consistently over decades and across all segments of the network, from the core to fixed and mobile access. This growth has been driven by a continuous stream of newly emerging and largely unanticipated digital applications and services. Many have failed at predicting the next "killer-application" or potential bounds on future traffic demands, as these predictions are typically based on extrapolating the bandwidth needs of already existing applications and services into the future. For example, some studies assume various higher-definition (4k or 8k) evolution steps of today's video services [1], while others examine the bandwidth of the human visual nerve [2], which multiplied by the number of potential video users might yield upper bounds for network traffic [3]. The problem with such approaches is that bottom-up estimates are fundamentally rooted in our experience with existing applications, yet network traffic has often been driven by totally unanticipated disruptions. As Internet pioneers Roberts and Wessler already observed back in 1970 [4], *"The kinds of services that will be available and the cost and ultimate capacity required for such service is difficult to predict."* And indeed, new kinds of traffic have emerged to dwarf previously dominating classes of traffic, such as (user-driven) data traffic overtaking voice traffic in the early 2000s [5], only to be then overtaken by video streaming and now machine-to-machine traffic, which dominates the traffic of webscale operators today and with the emergence of the Internet of Things (IoT) is expected to dominate mobile data traffic in the near future [6]. It appears that the success and demand for existing applications continuously drives scale and capacity of the underlying network infrastructure to points where further applications are enabled, renewing the cycle.

The compound annual growth rates (CAGRs) of network traffic have been extracted from actual long-term measurements by many individuals and organizations, and extrapolations into the future have been abundantly made [5], [7]–[10], revealing traffic growth rates typically between 25% and 80% per year, depending on operator, network segment, traffic type, and geographic region. In this context it is important to note that the Internet is not a single network but is composed of many diverse networks, some of which overlap geographically or share common links and are connected in a variety of ways with traffic joining or leaving at multiple locations. Today's Internet also contains substantial private networks that are used to interconnect data centers and deliver content; function and capacity of these networks is generally visible only to their operators. Variability in traffic and traffic growth rates on these diverse networks is expected, and measuring the total network traffic and its growth is practically impossible. Nevertheless, there is a variety of data that can be used to understand growth rates on the Internet, a representative collection of which is presented in this paper. While forward-looking traffic growth predictions always carry a natural amount of model dependent uncertainty, backward-looking long-term traffic logs portray historic reality and as such cannot be contested, but can be interpreted with respect to their implications on future network planning.

### A. Examples of Long-Term Traffic Growth Trends

Fig. 1 shows a collection of long-term historic traffic data; (a) shows the monthly averaged traffic passing through the

Manuscript received November 11, 2016; revised January 10, 2017; accepted January 18, 2017. Date of publication February 1, 2017; date of current version March 24, 2017.

The authors are with the Optical Transmission Systems and Networks Research Department, Bell Labs, Alcatel-Lucent, Holmdel, NJ 07733 USA (e-mail: peter.winzer@ieee.org; david.neilson@nokia-bell-labs.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2017.2662082

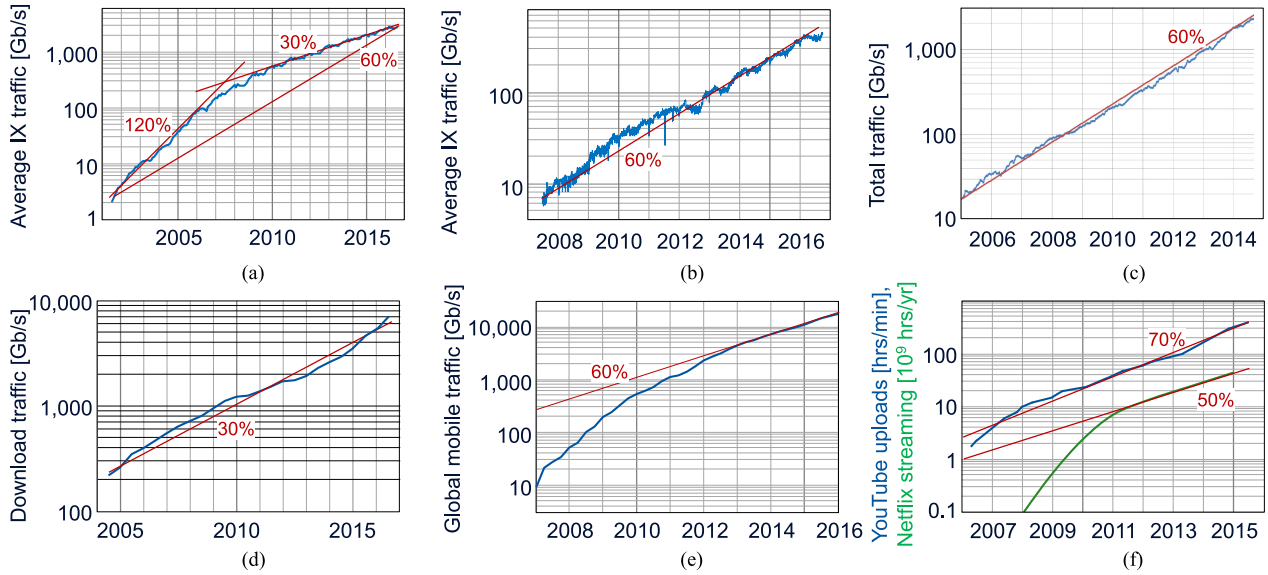


Fig. 1. Various long-term network traffic growth statistics from different geographical regions and applications, showing CAGRs between 30% and 70%. (a) Amsterdam Exchange (AMS-IX) [11]. (b) Seattle Exchange (SIX) [20]. (c) Total broadband demand (UK ISP) [3]. (d) Total broadband downloads in Japan [21]. (e) Global mobile data traffic [22], [23]. (f) Video traffic (YouTube, Netflix) [24], [25].

Amsterdam Internet Exchange (AMS-IX) over the past 15 years [11], with a CAGR of 120% between 2001 and 2007, flattening to a CAGR of 30% around 2007, a long-term average growth of 60% over the past 15 years. A very similar behavior is observed for the monthly averaged traffic passing through all Internet exchange points (IXPs) that are part of the Euro-IX group [12]. The change in slope around 2007 might be due to certain traffic types (such as video) bypassing IXPs and being served from content delivery networks (CDNs) and local server caches directly within the provider's network [13]–[15]. Also, many content providers are now performing private peering rather than going through a public IXP, especially for high peering bandwidths [16]. This is seen, e.g., in the large number of private peering points listed by major providers in peering databases [17], [18]. Additionally, peering with a large number of smaller providers is achieved using IXP route servers [19]. A different growth situation is found for the Seattle Internet Exchange (SIX) [20], which has seen a rather constant CAGR of 60% over the past 9 years, cf. Fig. 1(b).

Drilling down into the types of traffic within the networks, Fig. 1(c) shows the 10-year evolution of the total broadband demand at peak time for a major UK-based internet service provider (ISP) [3], with a consistent CAGR of 60%. Fig. 1(d) focuses on the total amount of broadband downloads in Japan, showing a CAGR of 30% over a 12-year window [21]. A major portion of data traffic now originates from mobile sources, and Fig. 1(e) shows the global mobile data traffic over the past 10 years [22], [23], whose growth has reached a steady CAGR of 60% for the past 3 years.

In terms of video traffic, driving much of the network traffic growth over the past years, the blue curve in Fig. 1(f) shows the hours of uploaded YouTube videos per minute over the past 10 years, with a CAGR of 70% [24]. Another big network traffic growth factor, shown by the green curve in Fig. 1(f), is the amount of Netflix streaming video, which after an initial start-up phase with more than tenfold growth from 2008 to 2012 has been growing at a constant rate of 50% per year [25]. In 2014 Netflix represented 33% of downstream Internet traffic in the United States during peak hours, accounting for over 10 Tb/s of network traffic [26]. In early 2014, Netflix had 18 Tb/s of capacity from deployed edge servers [27]. Note that growth in the volume of content, especially video, may not result in an equivalent amount of traffic growth throughout the network. Apart from growing video resolution and improved video compression, which factor into an assessment of network traffic growth, content may be cached closer to the end user to alleviate the load on the longer-haul links. We will discuss the possibility of caching and the resulting network traffic growth implications in Section II–C.

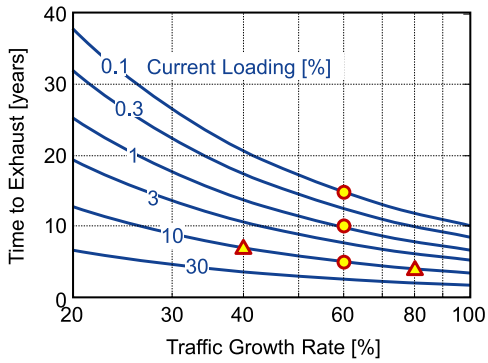


Fig. 2. Time it takes to exhaust the full capacity of a (transport or switching) system as a function of the traffic growth rate and the current system loading [28].

### B. *Capacity Exhaust and Exact Traffic Growth Rates*

In order to see what the above growth numbers imply for the required evolution of networking hardware, Fig. 2 plots the time it takes to exhaust the full capacity of a given (transport or switching) system based on a certain traffic growth rate and

TABLE I
LONG-TERM SCALING TRENDS FOR VARIOUS INFORMATION AND COMMUNICATIONS TECHNOLOGIES

| | | | |
|---|---|---|---|
| Generation & Processing | Supercomputer performance | 1995-2010 | 90% |
| | | 2010 – 2016 | 45% |
| | Microprocessor performance | 1980 – 2016 | 40% - 70% |
| | Core router capacity | 1987 – 2016 | 45% |
| | Router blade capacity | 1997 – 2016 | 40% |
| | Memory and storage capacity | 1980 - 2016 | 60% |
| | Wireless interface speed | 1995 – 2016 | 60% |
| | Fixed access interface speed | 1983 – 2016 | 40% - 50% |
| Transport | Router interface speed | 1980 – 2005 | 70% |
| | | 2005 – 2016 | 20% |
| | Transport interface speed | 1985 – 2016 | 20% |
| | WDM capacity per fiber | 1995 – 2000 | 100% |
| | | 2000 - 2016 | 20% |

a current fractional system loading, following the simple exponential growth equation [28]

$$C_{\mathrm{System}} = C_{\mathrm{Current}}(1 + r_{\mathrm{Growth}})^{\mathrm{TimeToExhaust}}, \quad (1)$$

where $C_{\mathrm{Current}}$ is the current traffic on the system, $r_{\mathrm{Growth}}$ is the traffic growth rate, and *TimeToExhaust* is the time it takes to fully exhaust the installed system capacity $C_{\mathrm{System}}$. The current fractional system loading is then given by $C_{\mathrm{Current}}/C_{\mathrm{System}}$. The circles indicate three operating points, all at a traffic CAGR of 60%. Let us first consider the point at 0.1% of current system loading, which corresponds, e.g., to a 10-Tb/s system that is currently only carrying 10 Gb/s worth of traffic; it would take 15 years to fully exhaust the capacity of this system. If a full 100-Gb/s of traffic were running on that system today (1% current loading), it would take 10 years to fully exhaust the capacity of the system; and if 10 channels at 100 Gb/s each were operating on the system today (10% current loading), it would take only 5 years to fully exhaust the system capacity. Quite a few operators today are in this kind of a situation.

Interestingly, we note that within the regions of typical traffic growth, the curves in Fig. 2 are rather flat, which implies that the exact traffic growth rate does not make much difference with respect to the time it takes to exhaust the full system capacity. For example, as indicated by the triangle markers in Fig. 2 for a system that is currently 10% loaded, it would take between 4 and 7 years to fully exhaust the system if the traffic growth rate varied between 80% and 40%. In the context of the massive scalability problems we are discussing in this paper, these few years of difference are of little relevance. Hence, we conclude that while there is massive historic support for a continuing ∼60% traffic CAGR, knowledge of the exact traffic growth rate is not key to the problems discussed in this paper.

## II. TECHNOLOGY SCALING DISPARITIES

In order to understand why networks have been able to support ∼60% of traffic CAGR over the past decades but are starting to fall behind in their capabilities to support such growth rates in the future, Table I summarizes the long-term scaling trends for several important communications technologies, revealing a divide between technologies for the *generation and processing of data*, which are growing at a CAGR in the 60% range (more accurately, between ∼40% and ∼90%), and technologies for the *transport of data*, which are growing at a CAGR of around 20%. This significant scaling disparity manifests itself in a ∼4x walk-off over the course of 5 years and in a ∼18x walk-off in 10 years and is at the heart of the network scalability problem that is threatening to result in an optical networks capacity crunch [29]. The long-term growth trends of Table I are extracted from a multitude of detailed statistics shown in Fig. 3, which we will discuss in more detail in the following subsections.

### A. Supercomputers and Data Centers

Accurately tracked by top500.org [30], Fig. 3(a) shows the evolution of supercomputer performance in terms of Tera Floating Point Operations Per Second (TFLOPS). The three data series represent the respectively leading supercomputer (#1), the respectively last one on the list of the top 500 supercomputers (#500), and the sum of all 500 computers on the list for an average performance measure. The average supercomputer performance shows a very consistent CAGR of 88% over more than 2 decades, with a hint of a slope change to 45% around 2013; this slow-down is not seen in the #1 trend, tracking an 88% CAGR, though the data is inherently step like making trend changes less obvious. Supercomputer #500 exhibits a CAGR of 95% from 1993 to around 2010, when the growth rate reduces to a consistent 45%.

Note that #500 on the list as of June 2016 [31] has 5,440 cores on 340 blades, each hosting 2 Intel Xeon 8-core processors that are also commonly used for servers in *data centers,* and each blade is interconnected using 40-Gb/s Ethernet. Some of the larger cloud data centers can be more than 100 times larger, with 50,000 to 80,000 servers [32]. Cloud data centers do not take part in the top500 benchmarking (so are missing in this scaling evolution), yet represent an increasingly attractive alternative for high-performance computing, without the need to construct dedicated supercomputers [33].

Cloud data centers, which are the next-generation mega data centers, are approaching the scale of several hundred thousand servers, that may be a single location or a cluster of tightly coupled smaller data centers [34]. Assuming 250,000 servers connected at 40 Gb/s (or 400,000 servers connected at 25 Gb/s), we find a total intra-datacenter traffic of 10 Pb/s. Assuming further that between 1% and 10% of that traffic actually leaves the data center, we arrive at datacenter interconnection capacities of between 100 Tb/s and 1 Pb/s, *per data center*. Hence, the scaling of supercomputing and the related scaling of cloud data centers will have a massive impact on network traffic scaling in the future.

### B. Microprocessors

The performance of microprocessors, rooted in the increase in the number of transistors per chip at a long-term consistent 45% per year following Moore's Law, cf. Fig. 3(b), is somewhat harder to uniquely quantify owing to various different benchmarks that may be used. Industry estimates range from 40% to 70% per year [35]–[38]. A representative collection of microprocessor performance data is given in Fig. 3(b) [35], [39],

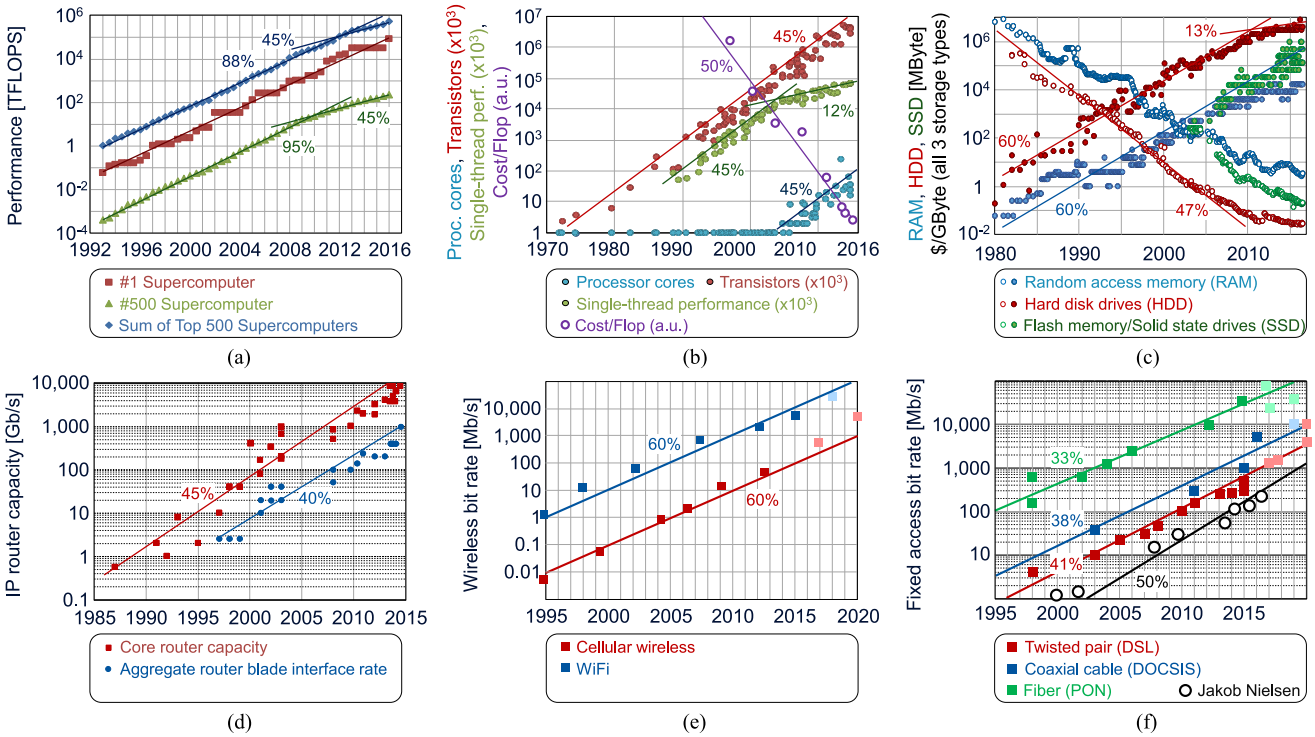


Fig. 3. Long-term scaling trends for various information and communications technologies. (a) Supercomputer performance, Top500 [30]. (b) Microprocessor performance and cost [35], [39]–[42]. (c) Storage capacity (•) and cost (◦) [47], [48]. (d) Core router and router blade capacities. (e) Wireless peak bit rates [57], [58]. (f) Fixed access bit rates [61], [62].

[40], showing that the benchmarked single-thread performance (green) has slowed down from a CAGR of 45% to 12% due to the stagnation in clock frequencies around 2005. Other benchmarks [37] show single-thread performance having slowed down from between 50% to 60% per year to a CAGR of around 20%. Irrespective of the amount of slow-down in single-thread performance growth, *parallelism* is making up for much of it, with the number of processor cores increasing at ∼45% per year since 2005 (blue data in Fig. 3(b)). The combination of more processor cores and increasing single-thread performance results in a continued increase in microprocessor performance at around 60% to 70% per year. Equally important is the corresponding price erosion, which according to several sources [41], [42] is around 50% per year when normalized to processing power and adjusted for inflation, cf. open purple data points in Fig. 3(b). This balance between performance increase and cost reduction has let the absolute cost of microprocessors remain roughly constant over decades (in fact, slightly decreasing due to an over-compensation in cost reduction), despite orders of magnitude of performance improvements, a general phenomenon that our digital information society has gotten quite used to.

### C. Memory and Storage

The importance of the scaling of memory and storage as far as communication networks are concerned lies in the trade-off between caching and transport. In fact, caching content close to the end user has been necessary to scale the Internet over the last two decades, by eliminating the need to retrieve identical information multiple times from a central repository across the entire network. Initially, large scale distributed caching networks were envisioned to deliver web content [43], [44] and have now grown to be more general purpose CDNs. These represent hundreds of thousands of globally distributed cache servers peering with or directly connected to thousands of networks, which typically deliver content supplied by multiple different content providers. In addition, there are CDNs that are dedicated to or that belong to a single major content provider [45]. However, increasing local caching to accommodate the rapid growth of content (CAGR of 70%, cf. Fig. 1(f)) simultaneously requires increasing the long-haul transport from central storage repositories to (a) deliver the content to the cache and to (b) accommodate access to the very significant portion of locally *un-cached* content. Content requests are often long-tailed, with the proportion of traffic from less popular content still being a substantial fraction of the total traffic; for example [26], [46], up to ∼10% of the content may account for ∼80% of the requests, while the remaining ∼90% of the content is still being requested ∼20% of the time. The effectiveness of caching to off-load traffic from the transport network thus depends on achieving a relatively high cache hit ratio (typically $> 80\%$), and therefore requires a significant fraction, e.g., ∼10% of the content being available on all local servers [15]. This also implies that for large content sets there will be cache misses that can turn into bursts of network traffic. With respect to user generated video such as YouTube, one might expect the relevance of a video to diminish with time and therefore reduce its need to be cached. However, in terms of the possibility to discard "old content", it is important to note that the average age

of content in an exponentially growing database ($C_0\, e^{\rho t}$) is well approximated by $1/\rho$, which for a 70% annual growth rate implies that the average age of YouTube videos remains constant at around 22 months and the number of videos younger than this constant average age (around 63% of the database) grows at the CAGR of the uploads. Therefore, while caching can reduce the fraction of content traffic in the long-haul network, it is unlikely to significantly reduce its long-term growth rate.

Fig. 3(c) shows the long-term evolution of storage, taking into account hard disk drives (HDDs, red), random access memory (RAM, blue), and flash memory/solid state drives (SSDs, green). Solid symbols represent storage capacity and open symbols are storage cost in $/GByte, in this case not adjusted for inflation. All data are taken from [47], [48]. The chart reveals a reasonably good fit to a 60% CAGR for both hard disk and memory capacity between the 1980s and ~2010, saturating into a ~13% growth rate around 2010. With the advent of flash memory/SSDs, the ~60% CAGR trend for memory continues to this day, maybe even at a slightly accelerated pace, albeit on a different technology and cost basis. Matching the overall scaling of memory and processing has been vital to maintaining balanced computer architectures [49]. Similar to microprocessors, storage costs have been falling at ~47% per year, over-compensating the increase in storage capacity. While the cost of RAM and HDDs is typically two orders of magnitude apart, flash memory/SSD technology lies in between and offers a denser storage option than HDDs at higher memory access rates.

With the continued scaling of server technologies, including microprocessor network cards and both flash and HDD storage, single servers fitting in two rack units can support more than 20 Gb/s of interface bandwidth today, equivalent to around 10,000 video streams, and store more than 60 TBytes of content. Higher-capacity storage servers with up to 300 TBytes in two rack-units are also possible [50]. While this represents a significant fraction (~10%) of the Petabyte-sized video library of a provider such as Netflix [26], it represents only 0.1% of the total uploads to YouTube. (From the data in Fig. 1(f), one can estimate the total size of the video database of YouTube to be around 400 PBytes in 2015.) From descriptions presented by Netflix [25]–[27] it is possible to service all their video requests using four racks of servers plus switches deployed at strategic network hubs with no need for a wider backend network. YouTube [51] appears to be served using a combination of local caching and connectivity to a much larger distributed network of servers, reflecting the difference in scale of their content sets.

Importantly, the rate at which hard disk (now only 13% CAGR) and flash storage (continuing at 60% CAGR) is scaling is less than the scaling of user generated content (70% CAGR), which means that any gap will tend to widen. The widening gap will affect local caching much more than central content repositories and impact the core network in terms of an increased amount of content traffic.

In addition to the challenges in managing the quantity of cacheable content, there continues to be live or other dynamic content that cannot be cached, which requires alternative methods such as replicating single streams based on user demand [52].

TABLE II
CHIP TECHNOLOGY AND COMPUTATIONAL SPECIFICATIONS OF THREE GENERATIONS OF ALCATEL PACKET PROCESSER CHIPS [53]

| Chip | Year | Line rate (Gb/s) | CMOS node (nm) | RISC cores | Core frequency (MHz) | Instr./s (billions) |
|---|---|---|---|---|---|---|
| FP | 2003 | 10 | 180 | 30 | 190 | 6 |
| FP2 | 2007 | 100 | 90 | 112 | 840 | 100 |
| FP3 | 2011 | 400 | 40 | 288 | 1000 | 308 |

### D. Core Routers and Router Blade Capacities

Fig. 3(d) examines the growth of Internet routers in terms of overall single-rack routing capacity (red) and router blade capacity (blue). Apart from some outliers in the early 2000's, core router capacities have been on a consistent 45% CAGR trend. This is not surprising, as routers build on custom routing ASICs, which follow the same growth trends as (very advanced) multi-core processors; Table II summarizes the chip technology and computational specifications of three generations of Alcatel packet processer chips [53].

Router blades have almost kept up with router capacities, at a CAGR of around 40%. Originally, router blade capacities represented the maximum logical capacity that could act as a single port on a router, called clear channel interface rate. One would ideally want to have the option of a single high-speed interface that can handle the entire traffic from one router blade, though today standards-based interfaces are limited to 100 Gb/s. While the highest-capacity router blade available today can process 1.2 Tb/s [54], it is actually composed of 12 individual channels of 100 Gb/s, each with its own interface and network processor. The router blade with the highest logical single-port capacity of 400 Gb/s has also been demonstrated with a single 400-Gb/s clear channel interface [55], featuring an integrated dual-carrier transponder.

Large core routers are designed to allow expansion to multiple racks to support more line cards in parallel. This is achieved by adding high-density very-short-reach interfaces onto the switch cards. Pairs of fabrics can be directly interconnected to provide a doubling in router capacity compared to a single rack. Larger numbers of racks, which allow even greater router scaling, can be accommodated through the use of separate switch fabric racks in a Clos network configuration. While it is possible to continue to scale routers to as large a capacity as required, this results in additional footprint, power and cost per bit over the conventional scaling of a single-rack router.

Regarding the *cost evolution* of core routers, [56] estimates a per-year cost-per-bit erosion of 26% during the period 2006–2013. This cost erosion is less than what would be required to balance the increase of router capacities for constant cost per bit. With an estimated 48% CAGR of router capacity sales [56], the overall cost of routing in the network increases at about 10% per year, i.e., $(1 - 0.26) \times 1.48 \sim 1.1$.

The *energy-per-bit* of core routers has been falling at a similar rate as the cost-per-bit, around 25% per year. As with costs, this implies that the overall energy consumption of routers in the network has been increasing at around 10% per year.

### E. Wireless Access

With reference to Fig. 3(e), wireless access bit rates have been growing at a CAGR of around 60% since the mid 1990's [57], [58], both for cellular technologies (red) and WiFi (blue). WiFi operates at ∼100x the speed of cellular access, at the expense of a reduced coverage radius on the order of ∼10 m instead of ∼100 m. Global mobile traffic in itself is also growing at a 60% CAGR, cf. Fig. 1(e). Future wireless networks such as 5G will see an increase both in aggregate data rates (10–100x) and in the number of wireless access points, the latter owing to the need for an increasing density of cells with decreasing sizes. This asks both for more backhaul capacity and for more high-capacity links to be deployed, potentially with more frequent switching of traffic as users move between cells. Since 5G offers the prospect of access rates comparable to current WiFi connected to fixed access, but with greater coverage and seamless hand-off, one may expect high bandwidth applications that currently dominate fixed access to migrate to the mobile environment, thereby driving both wireless and overall traffic growth. In addition to the high backhaul and transport capacities that 5G wireless networks will require, *low latency* will be a key 5G requirement as well, both to support centralized radio processing *architectures* (e.g., cloud radio access networks, C-RANs [59]) and to enable real-time applications from augmented reality to autonomous driving [60]. From an optical networking point of view, low latency requirements translate into high capacity requirements. This is either through optically routed, circuit-switched connections with minimal opto-electronic regeneration and through-traffic packet processing or by reducing the utilization and therefore congestion on packet based systems.

### F. Fixed Access

The growth of fixed access rates is less uniform than that of wireless access rates. Fig. 3(f) shows various long-term scaling trends across the industry. The red, blue, and green data, respectively, represent the evolution of Digital Subscriber Line (DSL) access rates over twisted copper wires, Data Over Cable Service Interface Specification (DOCSIS) access rates over coaxial cable, and passive optical networking (PON) access rates, growing at 41%, 38%, and 33% [61]. The open circles represent the long-term data collected by Jakob Nielsen since 1983 [62], together with his corresponding 50% long-term trend line, representative of the actually experienced single-user fixed access rates obtained by a North-American Internet user, and sometimes dubbed Nielsen's Law.

As an interesting (yet universally familiar) economic observation, the average cost per household spent on communications services relative to the disposable household income has remained roughly constant for decades, with detailed data available for both the United States and Australia [63]–[66], while access rates have increased by orders of magnitude. The same observation is made by network service providers around the world, who have experienced flat average revenue per user (ARPU) over years [56]. As discussed above for microprocessors and memory, flat costs despite dramatically increasing performance requires proportional cost/bit reductions, which access technology has been expected to provide.

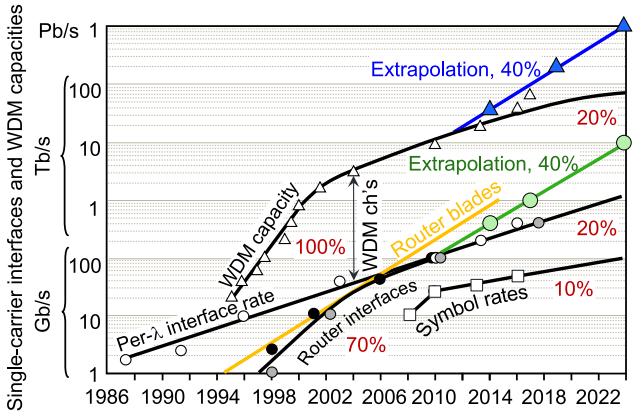


Fig. 4. Long-term scaling trends for optical communications technologies.

Increasing access speeds, especially through the continuous migration to highly scalable next-generation PON solutions, is likely to increase data consumption of end-users and of an increasing number of end-user hosted 'things' in two ways. First, the benefit of faster upload and download speeds results in shorter wait times for the end user, which is likely to increase the amount and the size of transferred files. Second, once access bandwidths cross certain thresholds that enable new services, these will then start driving new end-user traffic. A prominent example are video streaming services: Netflix video streaming started to become an increasingly attractive alternative to mailed DVDs once actually experienced single-user fixed access rates exceeded 10 Mb/s around 2008, compare Fig. 1(f) and Nielsen's data points in Fig. 3(f).

### G. Optical Line Interfaces

Fig. 4 shows the long-term evolution of single-wavelength optical interfaces used in optical transport equipment (SONET, SDH, OTN; white circles), revealing a consistent CAGR of 20% since the mid 1980's, limited by the scaling of cost-effective high-speed electronics and opto-electronics as opposed to the Moore's Law scaling of economically viable digital processing. Current commercial wavelength-division multiplexed (WDM) systems operate at single-carrier bit rates as high as 400 Gb/s [67]. An important observation in this context is that sustaining this 20% scaling has only been possible through the use of *parallelism* in the form of higher-order quadrature amplitude modulation (QAM) and polarization-division multiplexing (PDM): The 20% per year increase in symbol rates that drove single-carrier interface rates up to 40 Gb/s in the early 2000's, has been replaced by a mere 10% annual advance in high-speed CMOS ASIC-integrated digital-to-analog (DAC) and analog-to-digital (ADC) converters that are at the heart of high line rate transponders (cf. square markers in Fig. 4). The remaining 10% in interface rate scaling since ∼2004 are due to the 4-fold multiplexing inherent to coherent detection (2 quadratures × 2 polarizations) as well as an increase in amplitude levels per

degree of freedom, which fundamentally results in a reduced transmission reach. Note in this context that trading interface rates for transmission reach is a viable path for access systems, such as in DSL, where it is permissible to place the hand-off point between fiber and twisted copper pairs closer and closer to the home [61], or in wireless access, where it is permissible to increase the density of base stations and thereby reduce transmission distances [58]. However, in core networks the trade-off between interface rate and transmission reach is not as flexible, as these systems are expected to connect locations at fixed, geographically dispersed distances corresponding to the locations of data centers and population hubs. Some moderate trade-off is possible only for traffic that can use caching, at the expense of network storage, as discussed in Section II–C.

Router interface rates (black circles in Fig. 4) and the corresponding Ethernet standards (gray circles) advanced at a CAGR of ∼70% in the late 1990's and saturated into the capabilities of cost-effective high-speed electronics and opto-electronics in 2010, the remarkable year where 100-Gb/s was standardized both as OTN and as Ethernet, and first commercial solutions for single-carrier 100-Gb/s OTN line transponders and for 100-Gb/s Ethernet clients were available. However, neither 40-Gb/s nor 100-Gb/s Ethernet use single-carrier transmission at the full line rate today; instead, inverse multiplexing over typically 4 lanes or 4 wavelengths is being employed, in addition to multiple amplitude levels that have more recently been introduced to commercial client interfaces [68]. The 400-Gb/s Ethernet standard is anticipated for late 2017, which puts it right on the 20% CAGR trend shown in Fig. 4.

Another interesting observation from Fig. 4 is the changing need for inverse multiplexing, i.e., the transport of a single high-speed signal through multiple parallel lower-speed signals. From the orange curve in Fig. 4, which replots the evolution of router blade capacities from Fig. 3(d), we see that up until the late 1990's, multiple client interfaces were needed to carry the traffic of a single router blade, and traffic from multiple router blades could be aggregated onto a single optical wavelength for long-haul transport. By ∼2006, client interfaces had caught up with router blade capacities as well as with optical transport, eliminating the need for inverse multiplexing and wavelength aggregation. The current 20% growth of client as well as of optical interface rates, though, lets transport increasingly fall behind the continuing 40% capacity growth of router blades, increasing the need for inverse multiplexing both on the client side and on the optical transport (line) side. The observed technology scaling disparity results in a walk-off by about a factor of 2 every 5 years. At this scaling, we should expect to see 20-Tb/s router blades needing twenty 1T Ethernet interfaces mapped onto twenty optical wavelength carriers for long-haul transport by 2024.

### H. WDM System Capacity

Also shown in Fig. 4 is the aggregate per-fiber capacity of commercially available WDM systems (triangles). In the 1990's, commercial WDM capacities were growing at a CAGR of 100%, i.e., one could keep buying a WDM system of twice the capacity of last year's system for almost a decade. The over-provisioning in capacity supply beyond the 60% growth in bandwidth demand was one of the factors contributing to the 'telecom bubble' in the year 2000 [5]. Since then, WDM capacity growth has slowed dramatically to 20% per year, essentially reflecting the strongly saturating spectral efficiencies achievable by commercial systems, as WDM systems are closely approaching their respective Shannon limits [69], [70] while optical amplifier bandwidths have remained fixed.

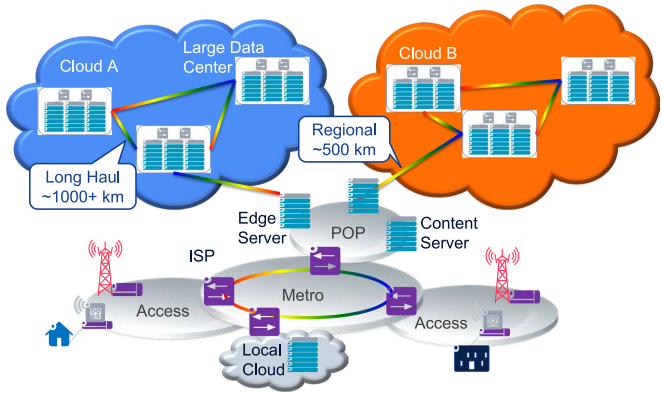


Fig. 5. Future network architecture. Long-haul networks interconnect cloud data centers at very high capacities. Cloud data centers will provide regional feeds to servers located in network POPs where they will interface with ISPs, who will then route traffic over metro networks for delivery to access networks.

The fact that the scaling trends for single-carrier interfaces and WDM capacities have been running in parallel over the past decade also implies that the number of WDM channels per system has stayed constant, at around 100 wavelengths. Increasing the number of wavelength channels at the expense of additional amplifier bandwidth has only been pursued recently on a commercial scale, with C+L-band systems starting to become available [67].

From a *cost-per-bit* point of view, [56] reports a cost erosion of 18% per year for long-haul WDM systems, compared to a CAGR in sold WDM systems of 37%, suggesting an annual cost increase of 12% for WDM transport in carrier networks.

The *energy-per-bit* for long-haul transport has been falling at around 20% per year [71], similar to the above numbers for costs, suggesting, as with core routers, a total increase in the cumulative energy consumption of installed transport equipment on the order of 10% per year.

## III. Optical Transport Systems in 2024

While we do not intend to present a complete network architecture for 2024 in this paper, we consider the way in which the overall network structure might impact optical transport networks, both in terms of capacity and switching flexibility. With reference to Fig. 5, we expect the networks of big cloud service providers (Clouds A, B, etc.) to each consist of a relatively small number (∼tens) of large mega-datacenters with hundreds of thousands of servers each. These mega-datacenters may be housed within individual warehouse-scale buildings, or may be distributed across multiple sites within a ∼100-km radius [34]. Mega-datacenters will be connected by private networks using

long-haul transport technologies (∼1000+ km), and will connect to many metro areas (one of them shown in Fig. 5) via regional (∼500 km) links, where they terminate on edge servers in network points of presence (POPs). At the POPs, content and services from both the data center clouds and other sources will then be delivered to ISP networks for further distribution to access networks over their metro WDM systems. There may also be local cloud data centers in the metro to provide functions whose privacy or latency requirements ask for proximity to the end user. Based on this structure we expect the network between mega-datacenters to consist of low-degree switching of very large capacity links, with switching occurring based on longer term requirements such as data replication or growth in datacenter sizes. The regional networks would require higher-degree connectivity since the number of metro areas would likely exceed the number of large datacenters. The metro space is where we would expect to see the need for most switching and reconfiguration: users will switch between the kinds of applications or content they use; the location of users changes as they move through the metropolitan space during the day and week, and the type of access changes with them, fixed or wireless.

Based on the native scaling of routers and router blades at 45% and 40% per year, respectively, cf. Section II–D and Fig. 3(d), we extrapolate the need for *higher-speed interfaces* in Fig. 4, green line, such that interfaces would track the scaling of router blades, with a ∼2-year delay, or equivalently with a constant need for 2:1 inverse multiplexing, starting our extrapolation in 2010. This extrapolation predicts the need for 400-Gb/s per-carrier WDM interfaces in 2014; these were in fact available on the market in 2016. While work on 400-Gb/s client interfaces is currently still ongoing, commercial interoperable interfaces will likely not be available much before the respective Ethernet standard towards the end of 2017. By 2017, though, our extrapolation asks for 1-Tb/s interfaces. From current industry trends it is highly unlikely that these will be commercially available by then, neither as single-carrier OTN nor as client interfaces. Even more unrealistic is the commercial availability of *single-carrier* 10-Tb/s optical line interfaces by 2024, clearly revealing the severe challenges that optical interfaces are meeting in trying to keep up with the CMOS based scaling of router blades.

Regarding the scaling of *system capacities*, the blue line in Fig. 4 extrapolates capacities at a modest traffic CAGR of 40%, also starting with 2010 as a baseline year. Following this extrapolation, commercial systems should have supported 40 Tb/s in 2014, which they actually did in 2016. Systems appearing on the market in 2017 will support ∼70 Tb/s over the C+L bands (cf. white triangle marker in Fig. 4), which is getting close to our extrapolation curve. However, both 200-Tb/s systems and 1-Pb/s systems that ought to be commercially available around 2019 and 2024 are well above reported single-fiber research capacity records of ∼100 Tb/s [72], [73] and are even beyond the estimated Shannon limit of a dual-band system, hence seem unrealistic for commercial implementation within the required time frame.

In the remainder of this section we will explore basic scaling options, both for high-speed interface rates and aggregate system capacities.

### A. Scaling System Capacity Through Improved Fiber

While deploying new low-loss and/or low-nonlinearity fiber will certainly help to increase the performance of commercial high-capacity WDM systems (as new fiber has always done along its history), improved fiber will not resolve system capacity scalability problems. To appreciate this, we note that the spectral efficiency (SE), or capacity per unit bandwidth, determining system capacity for a fixed optical amplifier bandwidth in a system with white Gaussian noise is given by Shannon's famous expression [69]

$$\mathrm{SE} = \log_2\left(1 + \mathrm{SNR}\right) \approx \log_2\left(\mathrm{SNR}\right), \tag{2}$$

where the approximation holds for high signal-to-noise ratios (SNR $\gg$ 1), the regime of interest for high-SE systems. Hence, in order to obtain a factor *K* of increase in the SE, one needs the *K*-th power of the original SNR (or *K* times its dB value),

$$K \cdot \mathrm{SE} \approx \log_2\left(\mathrm{SNR}^K\right) \propto K \cdot \mathrm{SNR}_{\mathrm{dB}}. \tag{3}$$

The overall SNR of an optical transmission line, including optical amplifier noise and nonlinear interference noise (NLIN), can be calculated using first-order perturbation theory [74]–[76] as

$$SNR = \frac{P}{\sigma_{\mathrm{OA}}^2 + \chi P^3}, \tag{4}$$

with the optical amplifier (OA) induced noise variance $\sigma_{\mathrm{OA}}^2$ proportional to the fiber's dB-loss coefficient $\alpha_{\mathrm{dB}}$ (exactly for ideal distributed amplification and to a good approximation for lumped amplification); the NLIN parameter $\chi$ is proportional to the fiber's squared nonlinearity coefficient $\gamma$ and, for ideal distributed amplification, is independent of the fiber's loss coefficient. At optimum signal launch power, the SNR obeys the proportionalities

$$SNR = \frac{1}{3}\sqrt[3]{\frac{4}{\sigma_{\mathrm{OA}}^4 \chi}} \propto \gamma^{-2/3}\alpha_{\mathrm{dB}}^{-2/3}. \tag{5}$$

Inserting this expression in (3) leads to the relations

$$\Delta SE = \frac{2}{3}\log_2\left(\frac{\gamma_0}{\gamma}\frac{\alpha_{\mathrm{dB},0}}{\alpha_{\mathrm{dB}}}\right) \tag{6}$$

and

$$2^{\frac{3}{2}(K-1)\cdot SE_0} = \frac{\gamma_0}{\gamma}\frac{\alpha_{\mathrm{dB},0}}{\alpha_{\mathrm{dB}}}, \tag{7}$$

with $\mathrm{SE}_0$, $\gamma_0$, and $\alpha_{\mathrm{dB},0}$ being the (single-polarization) spectral efficiency, nonlinear coefficient, and loss coefficient of the reference system over which a *K*-fold capacity gain (for a $\Delta$SE difference in single-polarization spectral efficiency) is to be achieved through improved fiber characterized by $\gamma$ and $\alpha_{\mathrm{dB}}$. For example, to achieve a doubling in capacity ($K = 2$) of a system with a (single-polarization) SE of $\mathrm{SE}_0 = 4$ b/s/Hz, typical of long-haul transmission research today, an improvement of either the nonlinearity coefficient or the dB-loss coefficient by a factor of 64 is needed (i.e., 0.01 $\mathrm{W}^{-1}\mathrm{km}^{-1}$ or 0.0023 dB/km when considering as a baseline some of today's most advanced fiber [77], [78] with a loss coefficient of 0.15 dB/km and a nonlinearity coefficient of 0.66 $\mathrm{W}^{-1}\mathrm{km}^{-1}$), or both coefficients need to be

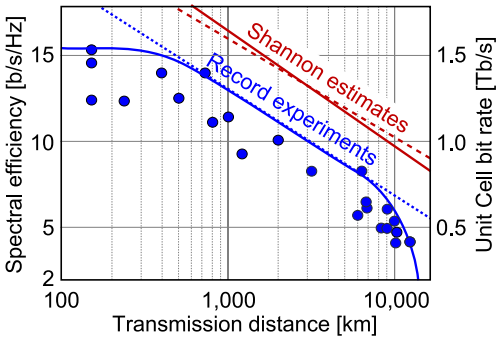


Fig. 6. Tradeoff between (dual-polarization) spectral efficiency and transmission reach in fiber-optic communication systems.

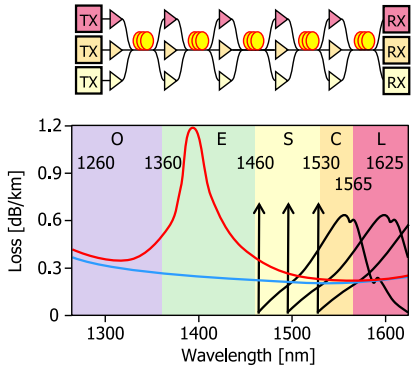


Fig. 7. Low-loss window of optical transmission fiber without (blue) and with (red) waterpeak. Also shown are the positions and gain spectra of Raman pumps that need to be placed in the S-band for Raman gain in the L-band.

simultaneously reduced by a factor of 8 (i.e., 0.083 $W^{-1}km^{-1}$ and 0.019 dB/km). Doubling the capacity of a more advanced system operating at a $SE_0 = 8$ b/s/Hz would require a reduction in either coefficient of 4096 (i.e., $1.6 \cdot 10^{-4}$ $W^{-1}km^{-1}$ or $3.7 \cdot 10^{-5}$ dB/km), or of both coefficients simultaneously by a factor of 64 (i.e., 0.01 $W^{-1}km^{-1}$ and 0.0023 dB/km). These numbers make even a doubling in system capacity a very serious research challenge and make a significant capacity gain from improved fiber parameters entirely unrealistic, even for the best hollow-core fiber one might imagine [79], [80].

### B. Scaling Interfaces and Capacity by Advanced Modulation

As discussed above, advanced modulation in the form of multi-level quadrature modulation combined with polarization multiplexing has been necessary to continue the scaling of single-carrier interface rates at 20% per year. However, higher-order modulation always comes at the cost of reduced transmission reach, as visualized in Fig. 6, showing two independent numerical estimates of the (dual-polarization) Shannon limit of standard single-mode fiber (SSMF) [69], [80], [81] as well as a collection of record experimental results (using a wide variety of fiber types). The slope of the Shannon limit is $-2$ b/s/Hz for each doubling in transmission distance [81] (solid), notwithstanding earlier numerical investigations reporting only $-1.65$ b/s/Hz (dashed [69], [80]) due to constraints on the modulation format. The $-2$ b/s/Hz slope for each doubling in transmission distance is predicted by the first order perturbation results discussed above: As both optical amplifier noise $\sigma^2_{\mathrm{OA}}$ [69] and NLIN parameter $\chi$ [74]–[76] are linearly proportional to the transmission distance $L$ (exactly for ideal distributed amplification and to a good approximation for lumped amplification), the SNR at optimum launch power, (5), decreases linearly with $L$, resulting in a 1-b/s/Hz reduction in SE per polarization for each doubling in transmission distance. Assuming instead of SSMF (0.2 dB/km, 1.27 $W^{-1}km^{-1}$) some of today's most advanced fiber (0.15 dB/km, 0.66 $W^{-1}km^{-1}$) [77], [78], the Shannon limit estimates in Fig. 6 would shift up only slightly, by about $\Delta SE = 0.9$ b/s/Hz (cf. Eq. (6)), or 1.8 b/s/Hz when considering both polarizations. Note from the experimental records shown in Fig. 6 that at the short-reach end, systems are limited by laser phase noise and ADC/DAC resolution [82], while in the ultra-long-haul regime additional effects manifest, such as nonlinear signal-noise interactions during fiber propagation [83], [84] or receiver-induced impairments such as electronically enhanced phase noise that degrades digital chromatic dispersion compensation [85], [86]. These effects let the experimentally achieved research records shown as blue markers in Fig. 6 depart from the blue dotted line (which has about the same slope as the Shannon limit estimates) at very long distances and at very high SEs.

Owing to the log-linear relationship between $L$ and SE, a small capacity improvement comes at the expense of a significant reach reduction, which puts a limit to advancing both single-carrier interface rates and WDM system capacities through higher-order modulation and lets the only possibility for significant system scaling be the use of parallelism in wavelength and space [87].

### C. Scaling System Capacity Through Wavelength Parallelism

One possible way to significantly scale system capacity is to increase the optical transmission bandwidth. To first order, an increased system bandwidth linearly increases system capacity. Fig. 7 shows typical loss coefficients across the low-loss window of commercial optical fiber with (red) and without (blue) the characteristic hydroxyl group (OH) absorption peak around 1380 nm. In principle, a factor of ~12 in bandwidth could be gained when operating deployed fiber from the O-band all the way to the L-band (~1260 nm–1625 nm, i.e., 53.5 THz), as opposed to using the C-band only (~1530 nm–1565 nm, i.e., 4.4 THz), as is done in today's commercial systems. However, it is very unlikely that this factor of 12 actually translates into a similar capacity gain due to several key disadvantages of ultra-broadband systems that make a factor of at most ~5 more realistic: For example, fiber has somewhat higher losses and optical amplifiers have higher noise figures outside the C-band; band-splitters are needed in multi-band systems to split the wavelength multiplex into appropriate amplification bands for parallel amplification, adding further loss to the spans. This lets different amplification bands exhibit different transmission reaches or SEs, or requires different amplifier spacings for dif-

ferent bands, leading to impractical system designs. Further, as soon as systems venture beyond the C+L band, Raman pumps that pump the L-band necessarily occupy parts of the S-band, directly competing for spectrum with S-band signals, as visualized by the Raman pump locations and gain spectra shown in black in Fig. 7. Even if backward-pumped, discrete or distributed pump reflections will cause significant crosstalk [88], which will particularly hurt the higher-order modulation formats one would want to use on capacity-optimized links [89]. Further, pump-pump and pump-signal interactions will become of increasing concern in such systems [90]. Also, practical concerns related to the maximum power that may be launched into deployed optical fiber without invoking fiber fuse problems come into consideration [91]. In addition, as ultra-wideband systems involve different component technologies across their bands, these systems will have a higher cost-per-bit than single-band (C-band) systems. Importantly, 'wavelength parallelism' is not true parallelism in the sense that a truly parallel system should deploy the exact same system components in parallel, which is key to bringing down costs through both *volume* and *integration*. The only real benefit of ultra-wideband systems is their re-use of the vast already installed fiber base (∼700 Million km in terrestrial long-haul networks, out of over ∼3 Billion km of globally deployed fiber [92], [93]). Deploying new fiber is very expensive, dominated not by fiber costs but by installation labor, and prices vary widely depending on the deployment scenario, with ∼$20,000 per km assuming available duct space being a realistic assumption; deploying a 1,000-km cable (∼$20M) therefore costs more than the WDM system operating over it. These basic commercial considerations of fiber re-use being the only real value proposition for ultra-wideband systems call into question proposals for entirely new (e.g., photonic crystal) fiber to operate over wide bandwidths outside the telecom wavelength range. Not only would such systems be incompatible with existing systems and fiber spans and hence would not be able to leverage the installed base, the main value proposition of broadband systems, but they would also require new component and subsystem technologies outside mature and high-volume C-band and possibly L-band solutions.

### D. Scaling System Capacity in Both Wavelength and Space

The above arguments exclude the sustainable scaling of system capacities in any of the discussed physical dimensions (time, quadrature, polarization, and frequency), leaving the *spatial dimension* as a last resort for capacity scaling [87], [94], with the added benefit that the component parallelism inherent to spatial multiplexing will bring down costs through both *volume* and *integration*, in analogy to the multicore microprocessor industry. Future systems will therefore populate a matrix of wavelength and space (WDM x SDM), cf. Fig. 8, performing space-division multiplexing (SDM) in addition to WDM.

The *bandwidth* of the elementary constituent of the WDM x SDM matrix, the 'unit cell' shown in Fig. 8, will be driven by the bandwidth accessible by electronics and opto-electronics. With a view on the commercial symbol rate evolution extrapolated to 2024, as shown in Fig. 4, we anticipate the commercial viability of modulators and detectors as well as fully ASIC-integrated

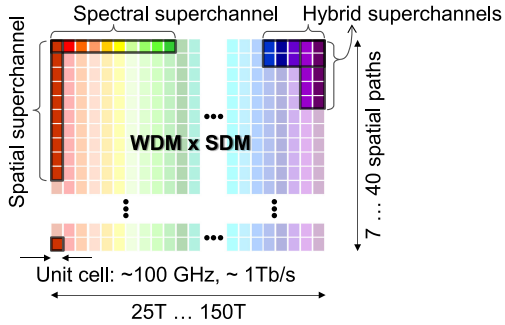


Fig. 8. Future systems will occupy a matrix of wavelength and space.

TABLE III
ELECTRONICALLY GENERATED SINGLE-CARRIER RESEARCH RECORDS

| Technique | Symbol rate | Format | Line rate | Ref |
|---|---|---|---|---|
| Multiplexer | 107 GBaud | 16-QAM | 856 Gb/s | [95] |
| | 120 GBaud | 16-QAM | 960 Gb/s | [96] |
| | 138 GBaud | QPSK | 554 Gb/s | [97] |
| Integrated DAC | 72 GBaud | 64-QAM | 864 Gb/s | [98] |
| | 90 GBaud | 64-QAM | 1.08 Tb/s | [99] |
| Dig. band interleaving | 180 GBaud | QPSK | 360 Gb/s (SP) | [100] |
| | 195 GBaud | PAM-4 | 390 Gb/s (EL) | [100] |

SP: Single-polarization; EL: Electrical only.

DACs and ADCs capable of ∼100-GBaud transmission. In research experiments, the highest-speed single-carrier signals that have been generated to-date use up to 138 GBaud and reach polarization-multiplexed single-carrier line rates up to 1.08 Tb/s, as summarized in Table III. Note that our '100-GBaud' assumption for 2024 only specifies an *addressable* bandwidth of ∼50 GHz (electrical) and 100 GHz (optical), irrespective of whether this bandwidth is used to digitally generate single-carrier or multi-subcarrier signals, which have recently been shown to have potentially improved nonlinear transmission performance [101], [102].

In order to determine the *bit rate* that can potentially be handled per unit cell, we again look at the trade-off between SE and transmission distance in Fig. 6. The right y-axis shows the unit cell bit rate, i.e., the achievable SE multiplied by a 100-GHz unit cell optical bandwidth. At terrestrial long-haul distances of a few thousand km, current research experiments have shown SEs of ∼10 b/s/Hz, which makes a 1-Tb/s unit cell bit rate a reasonable assumption for commercial systems in 2024. Ultra-long-haul systems may only be able to achieve a SE of ∼5 b/s/Hz, or a unit cell bit rate of 500 Gb/s. Short-reach systems operating over ∼100 km, as considered for datacenter interconnects (DCI), on the other hand, may be able to pack ∼1.5 Tb/s into a unit cell.

A typical C-band system consists of ∼50 unit cells, and a C+L-band system of ∼100 unit cells. Hence, the capacity per spatial path will be between ∼25 Tb/s (ultra-long-haul C-band) and ∼150 Tb/s (∼100-km C+L-band). For the Petabit/s system that will be needed in 2024 (cf. Fig. 4), a total of 1000 unit cells are required, which asks for between 7 and 40 parallel spatial paths. These observations once again show that there is no way around SDM in future optical transport networks.

Depending on whether a 10-Tb/s interface as a logical system channel is constructed using a single spatial path across multiple frequency slots or a single wavelength across multiple spatial paths, we speak of *spectral* or *spatial superchannels*, or of any suitable hybrid combination thereof, as shown in Fig. 8.

### *E. Spectral Superchannels*

Spectral superchannels [103], [104] are well established in today's optical transport systems to increase interface rates through inverse multiplexing to a few hundred Gb/s and beyond. As such, evolving WDM x SDM systems through the use of spectral superchannels represents an attractive smooth path from current technology. In addition, spectral superchannels do not initially require parallel fibers, which favors their use in terrestrial systems, where new system deployments are usually decoupled from new fiber installations.

Assuming a sustained 40% annual increase in router blade capacities, a single transponder will likely carry 10 Tb/s of traffic by 2024 (cf. Fig. 4), which implies that spectral superchannel systems with between 7 and 20 unit cells per superchannel will still use between 3 and 15 spectral superchannels across the employed optical amplification bands. Hence, wavelength switching will likely remain an ingredient in optical networks at that growth rate up until ~2030, when ~100-Tb/s superchannels might be needed as single logical transponder interfaces, which would occupy the entire optical amplification bandwidth across C and L bands. At that point, an entire fiber would become a logical point-to-point interface, and wavelength switching would be entirely replaced by space switching.

In terms of *routing and wavelength assignment*, wavelength-switched networks suffer from blocking even if there is still capacity left on particular spans. This is because in the absence of wavelength conversion, an optical end-to-end path requires the *same* wavelength to be available on all spans along an end-to-end link [105], necessitating defragmentation and/or re-routing to accommodate new traffic demands [106], [107]. However, once networks move to a small number of high-capacity spectral superchannels per fiber, with a sizeable number of fibers per system, the blocking probability can be substantially reduced, since the role of wavelength conversion to open up more routing possibilities per span is then accomplished by practically much simpler space switching. In the extreme case of a single (~100-Tb/s) superchannel per fiber (i.e., pure space switching), blocking only occurs when the entire capacity on a link is exhausted. In some networks, though, wavelength-based switching might be abandoned even before a superchannel fills an entire optical amplification band, if this simplifies the node architecture, control, and switching at only a small penalty in terms of blocking.

In terms of *subcarrier access*, spectral superchannels would in principle allow for adding and dropping individual optical subcarriers, although that use case is not likely to weigh heavily in determining the superchannel architecture. Sub-wavelength switching and grooming, which used to be one of the key functions of circuit switched networks in times where multiple router interfaces could be mapped into a single wavelength channel, is increasingly being replaced by point-to-point router

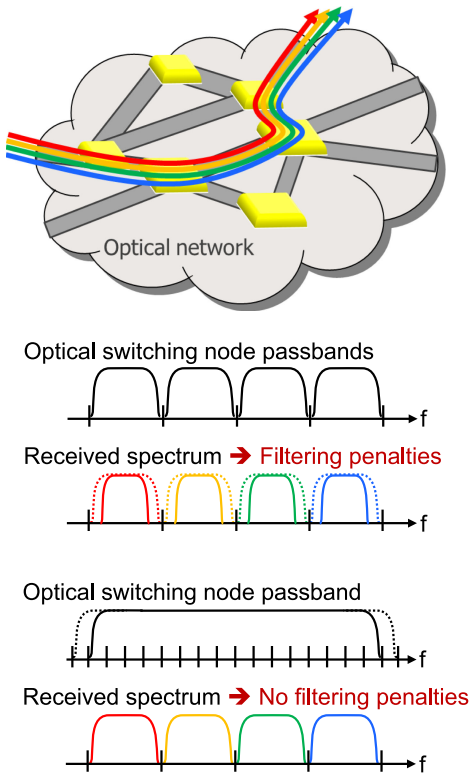


Fig. 9. Spectral superchannels avoid filtering penalties from optical routing node cascades through joint channel routing and flexible-grid passbands.

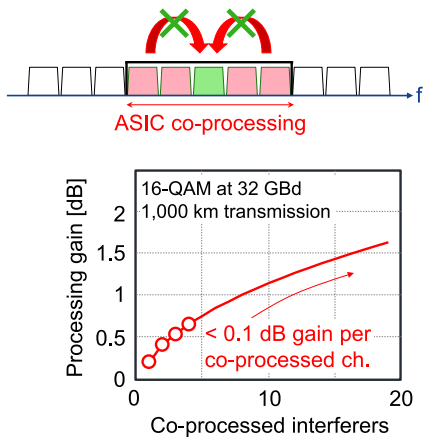


Fig. 10. Advantage from co-processing multiple superchannel subcarriers to reduce cross-channel nonlinear interference leads to rapidly diminishing returns [112].

connections with packet switching on a sub-wavelength basis. What spectral superchannels would offer, though, is a co-existence of interfaces of different spectral widths on the same optical line system, which together with dynamic modem-like transponder adaptation can be used to better exploit available capacity [108], [109].

Spectral superchannels have a clear advantage in *optically routed networks*, as the use of superchannel routing together with flexible-grid wavelength plans reduces penalties from filter concatenation if the need for subcarrier add/drop is abandoned [110], [111], cf. Fig. 9.

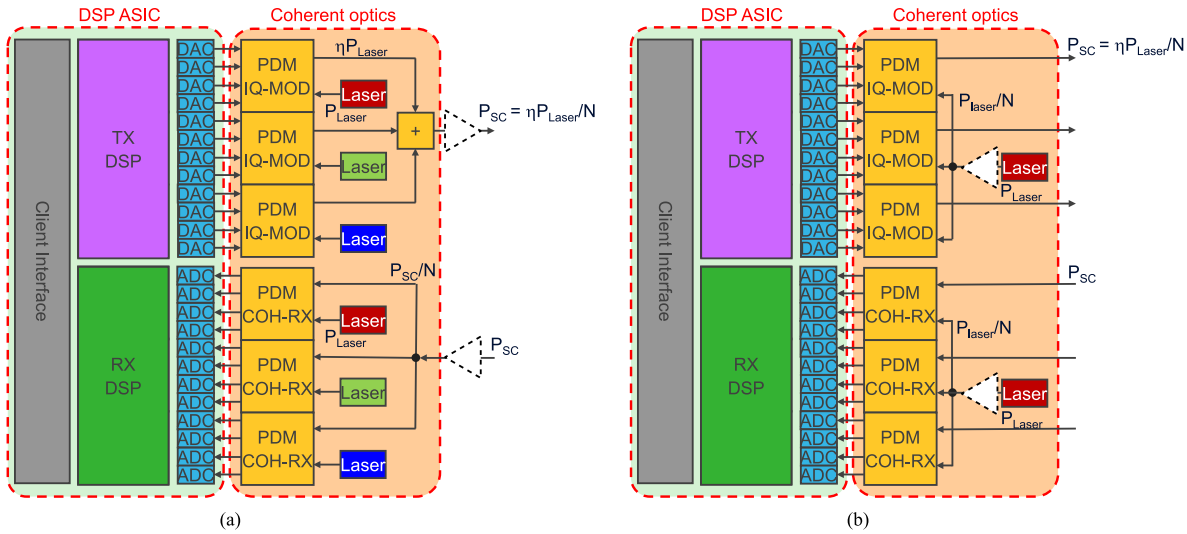


Fig. 11. Example for subsystem savings in terms of the required number of (equal-power) lasers for (a) spectral versus (b) spatial superchannels.

Another interesting aspect of spectral superchannels is their ability to compensate for NLIN between superchannel subcarriers, as shown in Fig. 10. However, these gains are very small yet require the complex co-processing of multiple subcarriers within the same ASIC. For the example shown in Fig. 10 [112], co-processing of one interfering subcarrier yields about 0.2 dB of SNR gain (at more than twice the required processing complexity), co-processing of two subcarriers yields 0.4 dB, and co-processing of more than 4 subcarriers yields as little as 0.1 dB per additionally co-processed subcarrier. These observations make nonlinear processing a weak argument for the use of spectral superchannels and let the smooth upgradability from existing networks as well as the possibility for swift optical routing in terrestrial optical networks be their main benefits.

In terms of *component integration*, spectral superchannels allow for a variety of options, including multi-channel electrical driver amplifiers, multi-channel optical modulators, multi-channel coherent receiver front-ends, and multi-core DSP ASICs, all of which are being commercialized today. To a limited extent, spectral superchannels also allow for the compensation of integration induced crosstalk between subcarriers [113], but such cross-wavelength compensation is much more limited than the single-wavelength digital compensation enabled by spatial superchannels [114], [115].

### F. Spatial Superchannels

Given the above established need for a WDM x SDM matrix, *spatial superchannels* (or *hybrid superchannels* involving subcarriers in the spatial domain in addition to the wavelength domain, cf. Fig. 8) allow for better subsystem integration and digital array impairment compensation opportunities than spectral superchannels. One example for subsystems integration where spatial superchannels yield distinct savings over spectral superchannels is shown in Fig. 11 [116], depicting a spectral (a) and spatial (b) coherent superchannel transponder architecture that addresses 3 subcarriers (unit cells), with its multi-channel DSP ASIC and its multi-channel coherent transmit and receive optics. Apart from possible differences in the DSP, the main difference between the two transponder architectures is the number of lasers required: A *spectral superchannel* transponder, Fig. 11(a), typically uses one tunable laser per subcarrier, potentially shared between the transmitter and the local oscillator if wavelength translation within the transponder from receive to transmit path is not required. (As the number of subcarriers increases, multi-carrier generators [117] and comb sources [118] are economically increasingly attractive over individual lasers.) In order to combine the subcarriers of a spectral superchannel onto a common transmission fiber, *colorless* (and hence inherently lossy) spectral combination is required. With $N$ subcarriers per spectral superchannel, a low-loss subcarrier combiner would require a 1:$N$ wavelength-selective switch (WSS) to be integrated onto the linecard to retain superchannel wavelength tunability while avoiding a dedicated add/drop port per subcarrier on an external optical switch. In addition, WSS architectures with overlapping passbands would have to be used to avoid individual subcarrier filtering. In the absence of such a per-linecard WSS, as shown in Fig. 11, each subcarrier laser of power $P_{\mathrm{Laser}}$ in the transmit path produces a per-carrier combined superchannel power of $P_{\mathrm{SC}} = \eta P_{\mathrm{Laser}}/N$, where $\eta$ is the modulator insertion loss. A *spatial superchannel* on the other hand, Fig. 11(b), uses a single laser that is split into individual spatial paths prior to modulation, reducing its power to $P_{\mathrm{Laser}}/N$. The power per modulated spatial superchannel subcarrier is then $P_{\mathrm{SC}} = \eta P_{\mathrm{Laser}}/N$, which is identical to the case of a spectral superchannel, but is achieved with a *single laser* as opposed to with $N$ lasers of the same individual power. A similar situation is found in the receive path: the coherently detected electrical signal power is proportional to $P_{\mathrm{SC}} \cdot P_{\mathrm{Laser}}/N$ in both cases. Another advantage of spatial superchannels pertains to optional optical amplifiers as part of the transponders, as it is easier to build single-wavelength amplifiers without being concerned about power balancing and gain flatness of multi-channel optical amplifiers, especially as the number of subcarriers in a spectral superchannel grows. Other examples for cost or energy savings specific to spatial superchannels include *optical ampli-*

*fier arrays* [119] and *optical switch arrays* that can be built in a simpler way for parallel switching of multiple spatial paths [120]–[123].

It is important to stress that in view of the need for a smooth upgrade path, SDM will be expected to work over any hybrid fiber infrastructure, including already deployed parallel fiber strands. In fact, most SDM systems should be expected to initially operate across multiple parallel single mode fibers, which can be bundled into a single cable. This approach currently allows scalability to over 3000 fibers in a single cable suitable for duct installation [34], [124], and fibers with reduced cladding thickness may even increase the per-cable fiber count [125]. This cable scaling would satisfy growth from a single fiber pair with a 60% CAGR for more than 15 years. Yet, we can assume further savings from SDM-specific waveguides: Capital expenditure (CAPEX) savings can be expected on the long run, as fiber supporting $N$ spatial paths in the form of multi-core or multi-mode fiber will eventually be cheaper than $N$ individual fibers. (An existence proof is today's multi-mode fiber, which supports ~100 parallel spatial paths (modes) but is only slightly more expensive than single-mode fiber.) In terms of operational expenditure (OPEX), the benefits of SDM fiber lie in interconnecting and splicing as well as in its interfacing to integrated opto-electronic arrays and packaging. Interconnecting $N$-path fiber can be expected to be in general faster, occupy a smaller form factor, and be less prone to connection mistakes than making $N$ individual connections or splices, even when compared to ribbon interconnection technologies. Note in this context that splicing is an important aspect of new fiber deployments: Typical terrestrial fiber spans are deployed (and hence spliced) in few-km sections, which leads to long-haul terrestrial links having ~1000 splices per spatial path. This results in considerable splicing efforts for cables containing tens to thousands of fibers. An example for connection benefits of SDM fiber is the dual-core fiber proposed by Corning to increase the faceplate density of short-reach interfaces [126]. In terms of interfacing to integrated opto-electronic arrays and packaging of array components, benefits of multi-core fiber-to-chip coupling have been shown for both short-reach [127] and long-haul [128] applications.

Another area where savings from spatial parallelism can be found is in the DSP, where it has been shown that certain DSP functions can be consolidated across multiple spatial paths [115]. Most importantly, though, any integration induced crosstalk between spatial paths, be it caused within multi-path transponder arrays, multi-path optical amplifiers, multi-path switches, or within a possibly SDM-specific fiber itself, can be efficiently compensated through multiple-input multiple-output (MIMO) DSP [114], as long as all subcarriers are considered as a cohesive entity in the spirit of an optical superchannel and no subcarrier add/drop operation is performed [129]. The required MIMO-DSP complexity has been found to be manageable [130] and has recently been implemented in real-time on a field programmable gate array (FPGA) platform [131].

The biggest drawback of spatial superchannels is their need for multiple parallel spatial paths from the beginning, making it less likely that this option will be initially chosen for terrestrial long-haul networks where pay-as-you-grow and smooth upgradability are cornerstone aspects and multiple parallel fiber strands may not be available. Systems that are either deployed together with the fiber (such as submarine links), or systems that are installed on an already available or an easily deployable massively parallel fiber infrastructure (such as ~100-km DCIs with as many as several 10,000 parallel fibers already installed today between pairs of datacenters [34]), will be able to naturally leverage the benefits of SDM first. Recent studies on submarine systems have pointed to the advantages of SDM from an energy efficiency point of view [132].

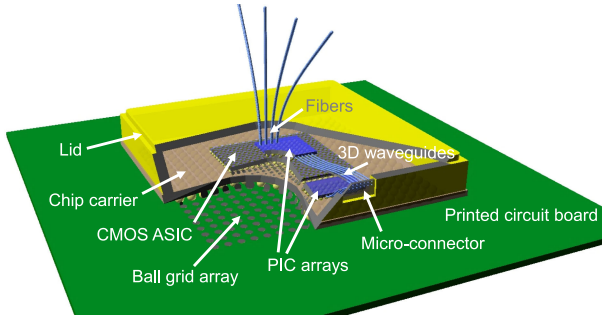


Fig. 12. Vision of a coherent 10-Tb/s optical interface solution.

### G. Transponder Integration

Irrespective of whether spatial or spectral superchannels will be chosen for a particular network application, component integration into massive arrays will be of key importance to reduce cost, energy consumption, and footprint of the resulting transponders. Given the fact that today's coherent DSP ASICs already handle 500 Gb/s worth of traffic, and considering the ~45% CAGR of CMOS processing capabilities, Fig. 3, it is not unreasonable to assume that the DSP underlying a 10-Tb/s interface will be able to fit within a single (or at most within a small number of) CMOS ASICs by 2024. In fact, scaling DSP chips to higher interface rates is limited today to a significant extent by interfacing problems to a correspondingly large number of high-speed opto-electronic components. As discussed above, a 10-Tb/s interface of 2024 will have to implement about 10 unit cells of 1 Tb/s each, which seems only doable by means of *integrated arrays*, directly coupled to the DSP ASIC. If the required integration asks for relaxed bit rates per array component, the size of the array has to be increased accordingly, as 10 Tb/s remains the target commercial interface rate. For example, if photonic integrated circuit (PIC) components are only able to achieve 200 Gb/s instead of 1 Tb/s, one would need integrated arrays of 50 instead of 10 parallel components. Such massive arrays, even when integrated on Silicon photonics or InP platforms, occupy a significant footprint compared to the powerful CMOS DSP chips they need to be interfaced to, which requires ultra-dense opto-electronic array integration. A vision of a future 10-Tb/s coherent optical interface with integrated CMOS DSP and photonics is shown in Fig. 12. Such integration poses many open challenges, including massive electronic interfacing, co-integration of an optical source (or provisioning of a remote 'optical power supply'), thermal management, and

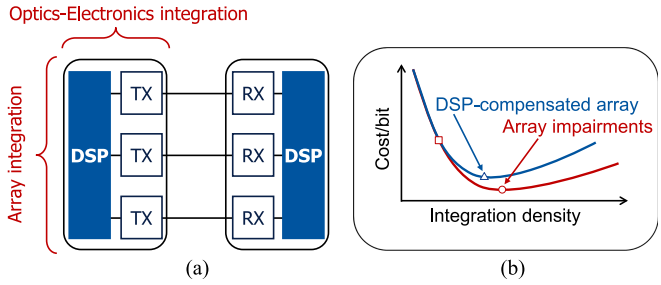


Fig. 13. Transponder array integration implies arrays of optoelectronic components as well as the close coupling between the array and the CMOS digital processing electronics (a). Codesign of integrated arrays and DSP allows for array-specific impairment mitigation and shifts the optimum integration density (b).

the interfacing to multiple parallel fiber strands, multi-core, or few-mode fiber through miniature connectors able to withstand conventional chip soldering processes.

Transponder integration in this context has three aspects, as visualized in Fig. 13(a). *Opto-electronic array integration* pertains to the integration of multiple parallel high-speed optical and opto-electronic components into arrays, such as modulators or coherent receivers [133]–[137]. In order to interface these arrays to a CMOS ASIC performing the coherent DSP, *optics-electronic integration* is needed [138], [139], be it monolithic [140]–[143], hybrid [144], [145], or by chip stacking and 3D interconnection techniques [146], [147]. Tight optics-electronics integration opens up new opportunities such as non-50Ω electronics-optics interfaces or efficient uses of segmented modulator structures [148], potentially with built-in DAC functionality [149]–[151], and will eventually result in co-packaged optical-digital solutions, similar to the path that is being taken by short-reach optical interfaces [152].

The close integration of dense arrays with digital electronics also opens up new holistic approaches in the joint design of integrated components and impairment compensating DSP, as visualized in Fig. 13(b): For example, integration might be optimum in terms of device processes, packaging, or interfacing considerations at the integration density indicated by the red circle. However, array impairments (such as array crosstalk [135], [136]) might prevent operation at that point and force less dense array integration at the point indicated by the red square marker. In a holistic co-design of integrated arrays and DSP, some array-specific impairments can be mitigated, as demonstrated for modulator crosstalk in [153]. These new possibilities open up the design space for integrated opto-electronic arrays and may result in new array integration optima (blue triangle marker). In the context of superchannels, tightly integrated arrays and their co-design with impairment mitigating DSP is significantly facilitated by the spatial superchannel concept.

## IV. Conclusion

We have examined a variety of long-term network traffic trends across different geographies and applications, confirming sustained exponential network traffic growth. While a compound annual traffic growth rate of 60% is corroborated from different angles, knowledge of the exact growth rate is found to be of secondary importance in a systems context, leading to only a few years of difference in terms of the need for networks to scale well beyond today's technological capabilities. Traffic growth is supported by technologies that generate, process, and store information, as well by access technologies between the network and the end user and/or the end machine. However, increasingly visible scaling disparities between those technologies and technologies used to transport data are starting to manifest in a network capacity crunch, with the transport network falling behind its required capabilities by a factor of about four every five years. By 2024, we predict the need for 10-Tb/s optical interfaces working in 1-Pb/s optical transport systems. To satisfy these needs, multiplexing in both wavelength and space in the form of a WDM x SDM matrix will be required. This matrix will likely consist of 100-GHz wide unit cells carrying 1 Tb/s of traffic over ∼1000 km, using on the order of 10 parallel spatial paths. Whether a logical 10-Tb/s interface will group unit cells in wavelength (spectral superchannel) or in space (spatial superchannel), or in a hybrid mixture thereof, will depend on the needs of the underlying network. In either case, dense array integration as well as close interfacing between arrayed opto-electronics and DSP ASICs will be a key ingredient to superchannels of all sorts. The need for smooth upgradability may make spectral superchannels the preferred solution for terrestrial networks, with space switching superseding wavelength switching on the long run, but submarine or datacenter interconnection systems may benefit from the better array integration options offered by spatial superchannels. If these trends were extended another 5 years to ∼2030 and 2035, it would, respectively, require systems with 100 and 1000 parallel spatial paths, so efficiently achieving spatial parallelism is an unavoidable requirement. Speculating much beyond a 2030 to 2035 time horizon, and assuming unabated network traffic growth even until then, research might have to venture to carrier frequencies much beyond the infrared (e.g. into the soft x-ray regime); such systems, whose fundamental building blocks are not even in sight today, would represent as radical an advance as fiber was over copper in the 1970's. In today's terms, in order for a new waveguide at such higher carrier frequencies to be as revolutionary as fiber was over copper (and to therefore warrant a complete network infrastructure re-build), that waveguide would have to [129]: immediately support 10 Petabit/s; be scalable to 100 Exabit/s in the future; allow for a repeater spacing of 2500 km (without any active elements between transmitter and receiver); and yield a 10-fold reduction in cable diameter and weight compared to fiber. Much more immediately, though, using infrared carriers and SDMxWDM systems, a decathlon of very significant research progress will have to take place across multiple areas, from system and network architectures to digital signal processing to integrated arrayed device designs in order to avoid an otherwise imminent optical networks capacity crunch.

## Acknowledgment

The authors gratefully acknowledge contributions to Section I–B from A. Chraplyvy as well as insightful discussions with and comments on the manuscript from M. Berger,

S. Chandrasekhar, X. Chen, J. Cho, A. Chraplyvy, R. Dar, R. Essiambre, N. Fontaine, A. Gnauck, S. Randel, G. Raybon, R. Ryf, and R. Tkach (Nokia Bell Labs), R. Alferness (UCSB), Richard Mack (CRU Consulting), H. Steenman (AMS-IX), B. Sanghani (Euro-IX), A. Lord (British Telecom), G. Wellbrock (Verizon), and Werner Klaus (NICT).

## REFERENCES


[1] E. Harstead and R. Sharpe, "Forecasting of access network bandwidth demands for aggregated subscribers using Monte Carlo methods," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 199–207, Mar. 2015.

[2] K. Koch *et al.*, "How much the eye tells the brain," *Current Biol.*, vol. 16, pp. 1428–1434, 2006.

[3] A. Lord, A. Soppera, and A. Jacquet, "The impact of capacity growth in national telecommunications networks," *Philosoph. Trans. Roy. Soc. A*, vol. 374, 2016, Art. no. 20140431.

[4] L. G. Roberts and B. D. Wessler, "Computer network development to achieve resource sharing," in *Proc. AFIPS Spring Joint Comput. Conf.*, 1970, pp. 543–549.

[5] R. W. Tkach, "Scaling optical communications for the next decade and beyond," *Bell Labs Tech. J.*, vol. 14, no. 4, pp. 3–9, 2010.

[6] M. Z. Shafiq *et al.*, "A first look at cellular machine-to-machine traffic: Large scale measurement and characterization," in *Proc. SIGMETRICS/PERFORMANCE Joint Int. Conf. Meas. Model. Comput. Syst.*, 2012, pp. 65–76.

[7] *Industry Connections Ethernet Bandwidth Assessment Ad Hoc*, IEEE Standard 802.3, 2012.

[8] "Metro network traffic growth: An architecture impact study," Bell Labs Strategic White Paper. 2013. [Online]. Available: http://www.tmcnet.com/tmc/whitepapers/documents/whitepapers/2013/9378-bell-labs-metro-network-traffic-growth-an-architecture.pdf

[9] M. Nowell, Cisco Visual Networking Index (VNI) Global IP Traffic Forecast Update; 2010–2015, 2011. [Online]. Available: www.ieee802.org/3/ad_hoc/bwa/public/sep11/nowell_01_0911.pdf

[10] CISCO Visual Networking Index: Forecast and Methodology, 2013–2018, 2016. [Online]. Available: www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf

[11] Amsterdam Internet Exchange "Historical monthly traffic volume" 2001–2016. [Online]. Available: https://ams-ix.net/technical/statistics/historical-traffic-data

[12] European IXP Reports. 2006–2016. [Online]. Available: https://www.euro-ix.net/tools/euro-ix-reports/

[13] V. K. Adhikari *et al.*, "Measurement study of Netflix, Hulu, and a tale of three CDNs," *IEEE/ACM Trans. Netw.*, vol. 23, no. 6, pp. 1984–1997, Dec. 2015.

[14] Netflix Open Connect. 2016. [Online]. Available: https://openconnect.netflix.com/

[15] D. Temkin, "Netflix: Open Connect: Starting from a Greenfield (a mostly Layer 0 talk)," 2015. [Onlilne]. Available: https://www.nanog.org/sites/default/files/meetings/NANOG64/1014/20150601_Temkin_Netflix_Open_Connect__v5.pdf

[16] W. B. Norton, "The 21st century peering ecosystem," DrPeering International, 2014. [Online]. Available: http://drpeering.net/core/chl 0.2-The-21 st-Century-Internet-Peering-Ecosystem.html

[17] A. Lodhi *et al.*, "Using peeringDB to understand the peering ecosystem," in *Proc. SIGCOMM Comput. Commun. Rev.*, 2014, pp. 20–27.

[18] PeeringDB: A database of current peering connections. [Online]. Available: https://www.peeringdb.com/

[19] P. Richter *et al.*, "Peering at peerings: On the role of IXP route servers," in *Proc. Internet Meas. Conf.*, 2014, pp. 31–44.

[20] Seattle Internet Exchange Statistics. [Online]. Available: https://www.seattleix.net/statistics/

[21] Ministry of Internal Affairs and Communications, Telecommunications Division Data Communication Section, Jul. 22, 2016. [Online]. Available: http://www.soumu.go.jp/main_content/000430359.pdf

[22] Ericsson Traffic and Market Report 2012, 2012. [Online]. Available: https://www.ericsson.com/res/docs/2012/traffic_and_market_report_june_2012.pdf

[23] Erisson Mobility Report 2016, 2016. [Online]. Available: https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf

[24] Data of Youtube views collected from: http://usatoday30.usatoday.com/tech/news/2006-07-16-youtube-views_x.htm https://youtube.googleblog.com/2012/01/holy-nyans-60-hours-per-minute-and-4.html https://2.bp.blogspot.com/_nitktFIV4LU/TNsaE3PuRQI/AAAAAAAAAEM/tlGgq1B2q_A/s1600/35.png https://youtube.googleblog.com/2012/05/its-youtubes-7th-birthday-and-youve.html https://youtube.googleblog.com/2013/05/heres-to-eight-great-years.html https://youtube.googleblog.com/2015/05/youtube-to-z-happybirthdayyoutube.html http://www.forbes.com/sites/stevenrosenbaum/2015/07/24/fans-selfies-and-the-future-of-tv/

[25] R. Hastings, Netflix CES 2016 Keynote, 2016. [Online]. Available: https://www.youtube.com/watch?v=5TR-NRpkW9I

[26] D. Fullagar, "Netflix open connect," 2014. [Online]. Available: http://conferences.infotoday.com/documents/197/2014CDNSummit-Netflix.pdf and https://johannesbergsummit.com/wp-content/uploads/sites/6/2014/11/Monday_8_David-Fullagar.pdf

[27] D. Temkin, "Help! My big expensive router is really expensive!" NANOG 60 (2014), 2014. [Online]. Available: https://www.nanog.org/sites/default/files/wednesday.general.temkin.panel.pdf

[28] A. R. Chraplyvy, personal communication, 2016.

[29] A. R. Chraplyvy, "The coming capacity crunch," plenary talk, ECOC, 2009.

[30] List of the 500 most powerful computer systems. 2016. [Online]. Available: https://www.top500.org/statistics/perfdevel/

[31] 500th most powerful computer system on 06/2016 list, 2016. [Online]. Available: https://www.top500.org/system/178858

[32] J. Hamilton, "AWS at scale," Nov. 12, 2014. [Online]. Available: http://mvdirona.com/jrh/talksandpapers/JamesHamiltonReInvent2014.pdf

[33] "An introduction to high performance computing on AWS," 2015. [Online]. Available: https://d0.awsstatic.com/whitepapers/Intro_to_HPC_on_AWS.pdf

[34] J. Hamilton, "Tuesday Night Live with James Hamilton," AWS re: Invent 2016, Las Vegas, NV, USA, 2016. [Online]. Available: http://mvdirona.com/jrh/talksandpapers/ReInvent2016_James%20Hamilton.pdf

[35] S. H. Fuller and L. I. Millett, Eds., *The Future of Computing Performance: Game Over or Next Level.?* Washington, DC, USA: Nat. Academies Press, 2011.

[36] G. Bell, "Laws of predictions," 1997. [Online]. Available: http://research.microsoft.com/en-us/um/siliconvalley/events/acm97/gb/sld086.htm

[37] J. Preshing, "A look back at single-threaded CPU performance," Feb. 8, 2012. [Online]. Available: http://preshing.com/20120208/a-look-back-at-single-threaded-cpu-performance/

[38] W. Gibbs, "A split at the core," *Sci. Amer.*, vol. 291, Nov. 2004, pp. 96–101.

[39] K. M. Bresniker, S. Singhal, and R. S. Williams, "Adapting to thrive in a new economy of memory abundance," *Computer*, vol. 48, pp. 44–53, 2015.

[40] K. Rupp, "40 years of microprocessor trend data," Jun. 25, 2015. [Online]. Available: http://www.karlrupp.net/2015/06/40-years-of-microprocessor-trend-data/

[41] W. D. Nordhaus, The Progress of Computing. 2001. [Online]. Available: http://www.econ.yale.edu/∼nordhaus/homepage/prog_083001a.pdf

[42] K. Grace and J. Salvatier, "Trends in the cost of computing," 2015. [Online]. Available: http://aiimpacts.org/trends-in-the-cost-of-computing/

[43] D. R. Karger *et al.*, "Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the world wide web," in *Proc. ACM Symp. Theory Comput.*, 1997, pp. 654–663.

[44] D. Karger *et al.*, "Web caching with consistent hashing," in *Proc. 8th World Wide Web Conf.*, 1999, pp. 1203–1213. [Online]. Available: http://www8.org

[45] V. Stocker *et al.*, "Content may be king, but (peering) location matters: A progress report on the evolution of content delivery in the internet," in *Proc. 27th Eur. Regional ITS Conf.*, 2016, vol. 7, Art. no. 148708.

[46] M. Cha *et al.*, "Analyzing the video popularity characteristics of large-scale user generated content systems," *IEEE/ACM Trans. Netw.*, vol. 17, no. 5, pp. 1357–1370, Oct. 2009.

[47] J. C. McCallum, "Price-performance of computer technology," in *The Computer Engineering Handbook*, V. G. Oklobdzija, Ed. Boca Raton, FL, USA: CRC Press, 2002, ch. 4.

[48] J. C. McCallum, "Disk Drive Prices (1955–2016), 2016. [Online]. Available: http://www.jcmit.com/diskprice.htm

[49] J. Gray and P. Shenoy, "Rules of thumb in data engineering," Microsoft Res., Redmond, Australia, Tech. Rep. MS-TR-99-100, 2000.

[50] Open Compute Vault Storage, 2013. [Online]. Available: http://www.opencompute.org/assets/download/Open_Compute_Project_Open_Vault_Storage_Specification_v0.7.pdf

[51] P. Casas *et al.*, "YouTube all around: Characterizing YouTube from mobile and fixed-line network vantage points," in *Proc. Eur. Conf. Netw. Commun.*, 2014, pp. 1–5.

[52] F. Larumbe and A. Mathur, "Under the hood: Broadcasting live video to millions," Dec. 3, 2015. [Online]. Available: https://code.facebook.com/posts/1653074404941839/under-the-hood-broadcasting-live-video-to-millions/

[53] "New DNA for the evolution of service routing: The Nokia FP3 400Gb/s network processor," Application Note, [Online]. Available: http://resources.alcatel-lucent.com/?cid=150427

[54] "Cisco ASR 9000 Series 12-Port 100-Gigabit Ethernet Line Cards Data Sheet," Aug. 23. 2016. [Online]. Available: http://www.cisco.com/c/en/us/products/collateral/routers/asr-9000-series-aggregation-services-routers/datasheet-c78-737623.html

[55] Nokia 7950 Extensible Routing System: 1 x 400G DWDM XMA LC fiber connector (dual 200G carriers in C-band grid), 2016. [Online]. Available: https://resources.alcatel-lucent.com/theStore/files/Nokia_7950_XRS_Media_Adapters_R14_Data_Sheet_EN.pdf

[56] J. S. Marcus, "The economic impact of Internet traffic growth on network operators," WIK-Consult, 2014. [Online]. Available: http://www.wik.org/uploads/media/Google_Two-Sided_Mkts.pdf

[57] G. P. Fettweis, "A 5G wireless communications vision," *Microw. J.*, vol. 55, Dec. 14, 2012, Art. no. 18751.

[58] H. Viswanathan and T. Sizer, "The future of wireless access," in *The Future X Network*. Boca Raton, FL, USA: CRC Press, 2016.

[59] China Mobile Research Institute, "C-RAN: The road towards green RAN," White Paper, 2011. [Online]. Available: http://labs.chinamobile.com/cran/wp-content/uploads/CRAN_white_paper_v2_5_EN.pdf

[60] F. Boccardi *et al.*, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[61] P. Vetter, "The future of broadband access," in *The Future X Network*. Boca Raton, FL, USA: CRC Press, 2016.

[62] J. Nielsen, "Nielsen's Law of Internet Bandwidth," Apr. 5, 1998. [Online]. Available: https://www.nngroup.com/articles/law-of-bandwidth/

[63] United States Personal Consumption Expenditures Data. [Online]. Available: http://www.bea.gov/

[64] P. S. Brogan, "The economic benefits of broadband and information technology," in *Proc. Media Law Policy, Broadband Policy Symp.*, Spring 2009, pp. 65–93.

[65] G. Kim, "Are households spending more, or the same, on communications, as a percentage of income?" Mar. 9, 2016. [Online]. Available: http://ipcarrier.blogspot.com/2016/03/are-households-spending-more-or-same-on.html

[66] Household communications expenditure, 2015. [Online]. Available: https://www.communications.gov.au/what-we-do/bureau-communications-research/data/household-communications-expenditure

[67] Nokia PSE-2 Super Coherent Technology, 2016. [Online]. Available: https://networks.nokia.com/products/pse-2-super-coherent-technology

[68] N. Eiselt *et al.*, "Real-time evaluation of 26-GBaud PAM-4 intensity modulation and direct detection systems for data-center interconnects," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2016, Paper Th1G.3.

[69] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 4, pp. 623–656, Oct. 1948.

[70] R.-J. Essiambre *et al.*, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.

[71] R. S. Tucker, "Green optical communications—Part I: Energy limitations in transport," *IEEE J. Sel. Topics Quantum Electron.*, vol. 17, no. 2, pp. 245–260, Mar./Apr. 2011.

[72] D. Qian *et al.*, "101.7-Tb/s (370x294-Gb/s) PDM-128QAM-OFDM transmission over 3x55-km SSMF using pilot-based phase noise mitigation," in *Proc. Opt. Fiber Commun. Conf.*, 2011, Paper PDPB5.

[73] A. Sano *et al.*, "102.3-Tb/s (224x548-Gb/s) C- and extended L-band all-Raman transmission over 240 km using PDM-64QAM single carrier FDM with digital pilot tone," in *Proc. Opt. Fiber Commun. Conf.*, 2012, Paper PDP5C.3.

[74] A. Splett, C. Kurzke, and K. Petermann, "Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities," in *Proc. Eur. Conf. Opt. Commun.*, 1993, Paper MoC2.4.

[75] P. Poggiolini *et al.*, "The GN model of fiber non-linear propagation and its applications," *J. Lightw. Technol.*, vol. 32, no. 4, pp. 694–721, Feb. 2014.

[76] R. Dar *et al.*, "Accumulation of nonlinear interference noise in fiber-optic systems," *Opt. Express*, vol. 22, pp. 14199–14211, 2014.

[77] Y. Yamamoto, Y. Kawaguchi, and M. Hirano, "Low-loss and low-nonlinearity pure-Silica-core fiber for C- and L-band broadband transmission," *J. Lightw. Technol.*, vol. 34, no. 2, pp. 321–326, Jan. 2016.

[78] Y. Yamamoto, M. Hirano, and T. Sasaki, "Low-loss and low-nonlinearity pure-Silica-core fiber for large capacity transmission," *SEI Tech. Rev.*, vol. 76, pp. 63–68, 2013.

[79] P. J. Roberts *et al.*, "Ultimate low loss of hollow-core photonic crystal fibres," *Opt. Express*, vol. 13, pp. 236–244, 2005.

[80] R.-J. Essiambre and R. W. Tkach, "Capacity trends and limits of optical communication networks," *Proc. IEEE*, vol. 100, no. 5, pp. 1035–1055, May 2012.

[81] R. Dar, M. Shtaif, and M. Feder, "New bounds on the capacity of the nonlinear fiber-optic channel," *Opt. Lett.*, vol. 39, pp. 398–401, 2014.

[82] S. Beppu *et al.*, "2048 QAM (66 Gbit/s) single-carrier coherent optical transmission over 150 km with a potential SE of 15.3 bit/s/Hz," *Opt. Express*, vol. 23, pp. 4960–4969, 2015.

[83] D. Rafique and A. D. Ellis, "Impact of signal-ASE four-wave mixing on the effectiveness of digital back-propagation in 112 Gb/s PM-QPSK systems," *Opt. Express*, vol. 19, no. 4, pp. 3449–3454, 2011.

[84] P. Serena, "Nonlinear signal-noise interaction in optical links with nonlinear equalization," *J. Lightw. Technol.*, vol. 34, no. 6, pp. 1476–1483, Mar. 2016.

[85] W. Shieh and K.-P. Ho, "Equalization-enhanced phase noise for coherent detection systems using electronic digital signal processing," *Opt. Express*, vol. 16, no. 20, pp. 15718–15727, 2008.

[86] C. Xie, "Local oscillator phase noise induced penalties in optical coherent detection systems using electronic chromatic dispersion compensation," in *Proc. Opt. Fiber Commun. Conf.*, 2009, Paper OMT4.

[87] P. J. Winzer, "Energy-efficient optical transport capacity scaling through spatial multiplexing," *IEEE Photon. Technol. Lett.*, vol. 23, no. 13, pp. 851–853, Jul. 2011.

[88] A. H. Gnauck, R. M. Jopson, and P. J. Winzer, "Demonstration of counter-propagating Raman pump placed near signal-channel wavelengths," *IEEE Photon. Technol. Lett.*, vol. 21, no. 1, pp. 154–157, Jan. 2017.

[89] P. J. Winzer, A. H. Gnauck, A. Konczykowska, F. Jorge, and J. -Y. Dupuy, "Penalties from in-band crosstalk for advanced optical modulation formats," in *Proc. 37th Eur. Conf. Opt. Commun.*, 2011, Paper Tu.5.B.7.

[90] P. M. Krummrich, R. E. Neuhauser, and C. Glingener, "Bandwidth limitations of broadband distributed Raman fiber amplifiers for WDM systems," in *Proc. Opt. Fiber Commun. Conf.*, 2001, Paper MI3.

[91] T. Morioka, "New generation optical infrastructure technologies: EXAT initiative towards 2020 and beyond," in *Proc. 14th OptoElectron. Commun. Conf.*, 2009, Paper FT4.

[92] R. Mack, CRU International. [Online]. Available: www.crugroup.com

[93] P. J. Winzer, "From first fibers to mode-division multiplexing," *Chin. Opt. Lett.*, vol. 14, p. 120002, 2016.

[94] P. J. Winzer, "Modulation and multiplexing in optical communication systems," *IEEE/LEOS Newslett.*, Feb. 2009. [Online]. Available: http://photonicssociety.org/newsletters/feb09/modulation.pdf

[95] G. Raybon *et al.*, "All-ETDM 107-Gbaud PDM-16QAM (856-Gb/s) transmitter and coherent receiver," in *Proc. 39th Eur. Conf. Exhib. Opt. Commun.*, 2013, Paper PD2D.3.

[96] J. Zhang and J. Yu, "WDM transmission of twelve 960 Gb/s channels based on 120-GBaud ETDM PDM-16QAM over 1200-km TeraWaveTM fiber link," in *Proc. Opt. Fiber Commun. Conf.*, 2016, Paper Tu3A.2.

[97] J. Zhang and J. Yu, "Generation and transmission of high symbol rate single carrier electronically time-division multiplexing signals," *IEEE Photon. J.*, vol. 8, no. 2, Apr. 2016, Art. no. 7902506.

[98] S. Randel *et al.*, "All-electronic flexibly programmable 864-Gb/s single-carrier PDM-64-QAM," in *Proc. Opt. Fiber Commun. Conf.*, 2014, Paper Th5C.8.

[99] G. Raybon *et al.*, "Single-carrier all-ETDM 1.08-Terabit/s line rate PDM-64-QAM transmitter using a high-speed 3-Bit multiplexing DAC," in *Proc. IEEE Photon. Conf.*, 2015, pp. 1–2.

[100] X. Chen *et al.*, "180-GBaud Nyquist shaped optical QPSK generation based on a 240-GSa/s 100-GHz analog bandwidth DAC," in *Proc. Asia Commun. Photon. Conf.*, PDP, 2016, Paper AS4A.1

[101] W. Shieh and Y. Tang, "Ultrahigh-speed signal transmission over nonlinear and dispersive fiber optic channel: the multicarrier advantage," *IEEE Photon. J.*, vol. 2, no. 3, pp. 276–283, Jun. 2010.

[102] P. Poggiolini *et al.*, "Analytical and experimental results on system maximum reach increase through symbol rate optimization," *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1872–1885, Apr. 2016.
[103] S. Chandrasekhar *et al.*, "Transmission of a 1.2-Tb/s 24-carrier no-guard-interval coherent OFDM superchannel over 7200-km of ultra-large-area fiber," in *Proc. Eur. Conf. Opt. Commun.*, 2009, Paper PD2.6.
[104] G. Bosco *et al.*, "On the performance of Nyquist-WDM terabit superchannels based on PM-BPSK, PM-QPSK, PM-8QAM or PM-16QAM subcarriers," *J. Lightw. Technol.*, vol. 29, no. 1, pp. 53–61, Jan. 2011.
[105] R. A. Barry and P. A. Humblet, "Models of blocking probability in all-optical networks with and without wavelength changers," *IEEE J. Sel. Areas Commun.*, vol. 14, no. 5, pp. 858–867, Jun. 1996.
[106] K. Christodoulopoulos, I. Tomkos, and E. A. Varvarigos, "Elastic bandwidth allocation in flexible OFDM-based Optical networks," *J. Lightw. Technol.*, vol. 29, no. 9, pp. 1354–1366, May 2011.
[107] O. Gerstel *et al.*, "Elastic optical networking: A new dawn for the optical layer?" *IEEE Commun. Mag.*, vol. 50, no. 2, pp. s12–s20, Feb. 2012.
[108] G. A. Wellbrock and T. J. Xia, "True value of flexible networks," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2015, Paper M3A.1.
[109] X. Zhou, L. E. Nelson, and P. Magill, "Rate-adaptable optics for next generation long-haul transport networks," *IEEE Comm. Mag.*, vol. 51, no. 3, pp. 41–49, Mar. 2013.
[110] T. Zami, "Current and future flexible wavelength routing cross-connects," *Bell Labs Tech. J.*, vol. 18, no. 3, pp. 23–38, 2013.
[111] J. M. Fabrega *et al.*, "On the filter narrowing issues in elastic optical networks," *J. Opt. Netw.*, vol. 8, no. 7, pp. A23–A33, 2016.
[112] R. Dar and P. J. Winzer, "On the limits of digital back-propagation in fully loaded WDM systems," *IEEE Photon. Technol. Lett.*, vol. 28, no. 11, pp. 1253–1256, Jun. 2016.
[113] J. Pan, A. Stark, C. Liu, and S. E. Ralph, "Fractionally-spaced frequency domain linear crosstalk cancellation with spectral alignment techniques for coherent superchannel optical systems," in *Proc. Opt. Fiber Commun. Conf. Expo./Nat. Fiber Opt. Eng. Conf.*, 2013, Paper OW4B.6.
[114] S. Randel *et al.*, "6×56-Gb/s mode-division multiplexed transmission over 33-km few-mode fiber enabled by 6×6 MIMO equalization," *Opt. Express*, vol. 19, pp. 16697–16707, 2011.
[115] M. D. Feuer *et al.*, "Joint digital signal processing receivers for spatial superchannels," *IEEE Photon. Technol. Lett.*, vol. 24, no. 21, pp. 1957–1960, Nov. 2012.
[116] P. J. Winzer, "Spatial multiplexing in fiber optics: The 10x scaling of metro/core capacities," *Bell Labs Tech. J.*, vol. 19, pp. 22–30, 2014.
[117] S. Chandrasekhar and X. Liu, "Terabit superchannels for high spectral efficiency transmission," in *Proc. 36th Eur. Conf. Opt. Commun.*, 2010, Paper Tu.3.C.5.
[118] E. Myslivets *et al.*, "Generation of wideband frequency combs by continuous-wave seeding of multistage mixers with synthesized dispersion," *Opt. Express*, vol. 20, pp. 3331–3344, 2012.
[119] A. H. Gnauck *et al.*, "Efficient pumping scheme for amplifier arrays with shared pump laser," in *Proc. 42nd Eur. Conf. Opt. Commun.*, 2016, Paper M2.A.1.
[120] D. M. Marom and M. Blau, "Switching solutions for WDM-SDM optical networks," *IEEE Commun. Mag.*, vol. 53, no. 2, pp. 60–68, Feb. 2015.
[121] R. Ryf *et al.*, "Wavelength-selective switch for few- mode fibers transmission," in *Proc. 39th Eur. Conf. Opt. Commun.*, 2013, Paper PD1.C.2.
[122] J. A. Carpenter *et al.*, "1 x 11 few-mode fiber wave- length selective switch using photonic lanterns," *Opt. Express*, vol. 22, pp. 2216–2221, 2014.
[123] N. K. Fontaine *et al.*, "Few-mode fiber wavelength selective switch with spatial-diversity and reduced- steering angle," in *Proc. Opt. Fiber Commun. Conf.*, 2014, Paper Th4A.7.
[124] [Online]. Available: https://www.sumitomoelectric.com/cms/wp-content/uploads/2016/08/Slotted-Core-Ribbon-Cable-3456F.pdf
[125] D. J. Richardson, J. M. Fini, and L. E. Nelson, "Space-division multiplexing in optical fibres," *Nature Photon.*, vol. 7, pp. 354–362, 2013.
[126] Y. Geng *et al.*, "High-speed, bi-directional dual-core fiber transmission system for high-density, short-reach optical interconnects," *Proc. SPIE*, vol. 9390, 2015, Art. no. 939009.
[127] B. G. Lee *et al.*, "120-Gb/s 100-m transmission in a single multicore multimode fiber containing six cores interfaced with a matching VCSEL array," in *Proc. IEEE Summer Topical Meeting*, 2010, pp. 223–224.
[128] C. R. Doerr and T. F. Taunay, "Silicon photonics core-, wavelength-, and polarization-diversity receiver," *IEEE Photon. Technol. Lett.*, vol. 23, no. 9, pp. 597–599, May 2011.
[129] P. J. Winzer, "Making spatial multiplexing a reality," *Nature Photon.*, vol. 8, pp. 345–348, 2014.
[130] S. Randel *et al.*, "Complexity analysis of adaptive frequency-domain equalization for MIMO-SDM transmission," in *Proc. Eur. Conf. Exhib. Opt. Commun.*, 2013, Paper Th.2.C.4.
[131] S. Randel *et al.*, "First real-time coherent MIMO-DSP for six coupled mode transmission," in *Proc. IEEE Photon. Conf.*, 2015, pp. 1–2.
[132] H. Zhang *et al.*, "Power-efficient 100 Gb/s transmission over transoceanic system," *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1859–1863, Apr. 2016.
[133] F. A. Kish *et al.*, "Current status of large-scale InP photonic integrated circuits," *IEEE J. Sel. Topics Quantum Electron.*, vol. 17, no. 6, pp. 1470–1489, Nov./Dec. 2011.
[134] V. Lal *et al.*, "Terabit photonic integrated circuits in InP," in *Proc. Integr. Photon. Res., Silicon Nanophoton. Conf.*, 2013, Paper IW5A.1.
[135] W. Yao *et al.*, "Performance degradation of integrated modulator arrays due to electrical crosstalk," in *Proc. Adv. Photon. Commun.*, 2014, Paper IW3A.3.
[136] W. Yao *et al.*, "Experimental and numerical study of electrical crosstalk in photonic-integrated circuits," *J. Lightw. Technol.*, vol. 33, no. 4, pp. 934–942, Feb. 2015.
[137] W. Heni *et al.*, "High speed plasmonic modulator array enabling dense optical interconnect solutions," *Opt. Express*, vol. 23, pp. 29746–29757, 2015.
[138] R. Soref, "The past, present, and future of silicon photonics," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, no. 6, pp. 1678–1687, Nov./Dec. 2006.
[139] D. Thomson *et al.*, "Roadmap on silicon photonics," *J. Opt.*, vol. 18, 2016, Art. no. 073003.
[140] A. Huang *et al.*, "A 10 Gb/s photonic modulator and WDM MUX/DEMUX integrated with electronics in 0.13/spl mu/m SOI CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2006, pp. 922–929.
[141] S. Assefa *et al.*, "Monolithically integrated silicon nanophotonics receiver in 90 nm CMOS technology node," in *Proc. Opt. Fiber Commun. Conf. Expo./Nat. Fiber Opt. Eng. Conf.*, 2013, Paper OM2.H.4.
[142] L. Zimmermann *et al.*, "BiCMOS Silicon photonics platform," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2015, Paper Th4E.5.
[143] D. Knoll *et al.*, "High-Performance BiCMOS Si photonics platform," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting*, 2015, pp. 88–96.
[144] M. J. R. Heck *et al.*, "Hybrid silicon photonic integrated circuit technology," *IEEE J. Sel. Topics Quantum Electron.*, vol. 19, no. 4, Jul./Aug. 2013, Art. no. 6100117.
[145] J. J. G. M. van der Tol *et al.*, "Photonic Integration on an InP-membrane," in *Proc. Integr. Photon. Res., Silicon Nanophoton. Conf.*, 2016, Paper ITu2A.1.
[146] M. Rakovski *et al.*, "Low-power, low-penalty, flip-chip integrated, 10Gb/s ring-based 1V CMOS photonics transmitter," in *Proc. Opt. Fiber Commun. Conf. Expo./Nat. Fiber Opt. Eng. Conf.*, 2015, Paper OM2.H.5.
[147] K. Yashiki *et al.*, "25-Gbps error-free operation of chip-scale Si-photonics optical transmitter over 70oC with integrated quantum dot laser," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2016, Paper Th1.F.7.
[148] M. Papuchon, C. Puech, and A. Schnapper, "4-bits digitally driven integrated amplitude modulator for data processing," *Electron. Lett.*, vol. 16, no. 4, pp. 142–144, 1980.
[149] D. Patel *et al.*, "Silicon photonic segmented modulator-based electro-optic DAC for 100 Gb/s PAM-4 generation," *IEEE Photon. Technol. Lett.* vol. 27, no. 23, pp. 2433–2436, Dec. 2015.
[150] T. N. Huynh *et al.*, "Flexible transmitter employing Silicon-segmented Mach–Zehnder modulator with 32-nm CMOS distributed driver," *J. Lightw. Technol.*, vol. 34, no. 22, pp. 5129–5135, Nov. 2016.
[151] M. Vanhoecke *et al.*, "Segmented optical transmitter comprising a CMOS driver array and an InP IQ-MZM for advanced modulation formats," *J. Lightw. Technol.*, 2017, to be published.
[152] M. P. Li, "Overcome copper limits with optical interfaces," Altera White Paper, 2011.[Online]. Available: https://www.altera.com/content/dam/altera-www/global/en_US/pdfs/literature/wp/wp-01161-optical-fpga.pdf?WT.mc_id=ap_pr_al_ne_tx_j_132
[153] X. Chen *et al.*, "Characterization and digital pre-compensation of electro-optic crosstalk in silicon photonics I/Q modulators," in *Proc. 42nd Eur. Conf. Opt. Commun.*, 2016, Paper Tu3.A.5.

Authors' biographies not available at the time of publication.