

# Evolution Toward the Next-Generation Core Optical Network

Adel A. M. Saleh, *Fellow, IEEE*, and Jane M. Simmons, *Senior Member, IEEE*

*Invited Tutorial*

**Abstract**—With high-bandwidth and on-demand applications continuing to emerge, next-generation core optical networks will require significant improvements in capacity, configurability, and resiliency. These advancements need to be achieved with architectures and technologies that are scalable with respect to network cost, size, and power requirements. To investigate the limitations of extending today's solutions to meet these goals, a North American backbone network with a tenfold growth in traffic is modeled. The results of this paper illuminate at least three areas that will potentially require innovative solutions, namely 1) transmission modulation formats, 2) switching granularity, and 3) edge traffic grooming. In addition to probing issues related to increased capacity, configurability is also examined, mainly in the context of switching architectures. Advanced network protection is discussed as well, at a high level. A central theme is how to harness the trend of optics scaling better than electronics. Throughout this paper, potential advancements in architecture and technology are enumerated to serve as a foundation for the research needed to attain the goals of next-generation core networks.

**Index Terms**—All-optical regeneration, all-optical switch, backbone networks, configurability, core networks, grooming, long-haul networks, multidegree optical add/drop multiplexer (OADM-MD), optical bypass, optical reach, shared mesh restoration, wavebands.

## I. INTRODUCTION

ADVANCES in long-haul<sup>1</sup> network technology over the past ten years, most notably in fiber capacity, optical switching, and optical reach [1], have shifted the bandwidth and operational bottlenecks from the core network to metro and access networks. However, as advances in wavelength-division multiplexing (WDM) technology propagate closer to the network edge, thereby enabling the proliferation of applications with very high bandwidth and/or stringent performance requirements, core networks eventually will become strained in both capacity and flexibility. Thus, while the capabilities of today's

backbone network are sufficient for current levels and types of traffic, the surge in capacity at the network edge, combined with the burgeoning number of on-demand applications, will fuel a rapid growth over a range of services that will require major evolution of the core network. To continue the cycle of networking advancements, it is important to understand the benefits and limitations of today's technologies and architectures and determine what advances will be needed to meet the requirements of next-generation core networks in the next five to ten years.

More specifically, next-generation core networks will require significant improvements in capacity, configurability, and resiliency. Aggregate network demand, as measured by summing the traffic at demand endpoints, could easily be on the order of 100 Tb/s, an order of magnitude increase over today's networks. At current backbone growth rates, e.g., [2], a tenfold increase will occur in five to ten years. While the optical layer of current networks is generally static, the emergence of on-demand services will require that the optical layer support a much greater degree of configurability. Rapid reconfiguration also will be needed for network restoration; as mission-critical applications become more tied to the transfer of large amounts of data, the ability to survive multiple concurrent failures will become essential.

It is important that all of these goals be met in a scalable fashion with respect to network cost, equipment size, and power requirements. It is desirable that solutions that are suitable for a range of commercially deployed networks as well as government owned networks be found in order to achieve economies of scale. Furthermore, note that the requirements of the network will be very heterogeneous. For example, while some applications may require format transparency, the bulk of the applications will not. Thus, the overriding goal of the next-generation network will not be to provide all things to all users but rather to ensure that the various segments of users are able to achieve a level of service commensurate with their particular needs.

One solution is to simply scale up today's technology to meet the evolving requirements. To investigate the efficacy of this solution, we model a North American backbone network with an aggregate demand of up to 100 Tb/s. The results of this paper illuminate at least three areas that will potentially require innovative solutions to enable this tenfold growth over today's traffic, namely: 1) transmission (e.g., line rate and modulation format), 2) switching granularity, and 3) edge traffic grooming. These

Manuscript received December 1, 2005; revised June 19, 2006.

A. A. M. Saleh is with the Defense Advanced Research Projects Agency (DARPA), Arlington, VA 22203-1714 USA

J. M. Simmons is with Monarch Network Architects, Holmdel, NJ 07733 USA

Digital Object Identifier 10.1109/JLT.2006.880608

<sup>1</sup>The terms "core," "backbone," and "long-haul" are used interchangeably in this paper. These terms refer to, for example, U.S.-wide or pan-European networks.

areas are clearly interrelated and pose many tradeoffs. For example, the higher the transport rate of an individual wavelength, the fewer the number of wavelengths that need to be switched, but the more grooming needed to pack the wavelengths efficiently. Various solutions in these three areas are proposed and analyzed from the point of view of overall network cost and technological feasibility.

The requirement of rapid reconfigurability poses challenges especially in the area of switching architecture. We discuss several core optical switch architectures, along with their performance and cost tradeoffs. We also consider simplification of the switch technology through the use of more flexible transmit/receive cards. The ability to rapidly reconfigure a network typically requires the predeployment of some amount of networking equipment that is utilized on an as-needed basis (e.g., in response to a shift in demands). We propose architectures that can minimize the cost impact of predeployed equipment while maintaining a high degree of flexibility.

In looking at the challenges of providing very high capacity and configurability, the trend of optics to scale better than electronics suggests that optics should play a greater role as networks evolve. For example, optical aggregation of substrate traffic at the edge of the network may prove to be a more scalable means of packing wavelengths as compared to conventional electronic multiplexing or grooming. Furthermore, all-optical regeneration may become more cost effective and consume less power than its electronic counterpart when the line rate increases beyond a certain threshold. The role of optics will be discussed more in-depth throughout this paper. Advances in optoelectronic and photonic integration [3], though also important, will not be discussed in detail.

To gain insight into the requirements of next-generation core networks, the next section examines various types of applications that can be expected to evolve over the next several years. This is followed by a section that summarizes the state-of-the-art technology being deployed in today's backbone networks. We extrapolate this technology to model a core network that grows to an aggregate demand of 100 Tb/s; details are presented in Section IV. The implications of this growth in capacity are analyzed in Section V. Configurability, especially in terms of how it impacts the architecture of network elements, is explored in Section VI. Advanced network protection is discussed in Section VII.

As we identify limitations of today's networks in providing the desired levels of bandwidth and configurability needed for the future, we enumerate various advances in architecture and technology that will be needed. Numerous potential solutions are proposed at a high level to serve as a foundation for the research needed to attain the goals of next-generation core networks.

## II. APPLICATIONS

The applications that are currently emerging and that will continue to mature over the next several years are the driving force behind the need for advances in the long-haul network. Clearly, it would be impossible to predict the full range of future applications. Rather, the goal of this section is to enumerate

a number of these applications, both commercial and military, that have very diverse requirements.

Growth in capacity requirements will come from both a surge in the number of users with high-speed access and a proliferation of bandwidth-intensive applications. Assuming that service providers follow through with their plans to deploy optical fiber directly to homes or neighborhoods, in the next few years, access speeds of up to 100 Mb/s will be available to tens of millions of users. Some carriers are even planning for up to 1-Gb/s access speeds. This is substantially faster than current digital subscriber line (DSL) and cable modem speeds, which are typically less than 10 Mb/s. The growth of "triple-play services," i.e., the convergence of high-speed data, video, and telephony over a single pipe, will significantly increase network capacity demands, especially due to video traffic. On-demand video is already a rapidly growing application. In addition, it is expected that a large number of video "narrowcasters" will spring up, offering a variety of specialized content over the Internet.

To complement access infrastructure deployments, advanced protocols are being developed to provide higher quality high-bandwidth services. For example, to provide an enterprise Ethernet local area network (LAN)-like environment over a backbone network, the virtual private LAN service (VPLS) standard has been proposed [4]. It combines Ethernet access with multiprotocol label switching (MPLS) core technology to deliver end-to-end quality of service. Virtual all-to-all private networks can be established much more easily than is currently possible and at much lower cost, which will encourage carriers to expand their markets and businesses to subscribe to more advanced services. Furthermore, 100-Gb/s Ethernet is likely to emerge in the next five to ten years as both a bandwidth driver at the network edge as well as a transport mechanism in the core.

Protocols that depend on the optical layer being reconfigurable are also being developed. For example, the Optical Internetworking Forum user-network interface (UNI) [5] and the generalized MPLS UNI [6] provide a means for higher layer "clients," e.g., the Internet protocol (IP) layer, to request via the control plane the establishment and tear-down of connections in the optical layer. These protocols are designed to enable more optimal resource utilization, greater network resiliency, and advanced services such as end-user-initiated provisioning through automated network configuration.

Another growing application is grid computing, which is used as a means of sharing distributed processing and data resources that are not under centralized control in order to achieve very high performance. There are already dozens of grid networks in existence, some with requirements of petabyte data sets and tens of teraflops of computational power [7], [8]. To support these massive requirements, large pipes are required to connect the major sites, essentially forming a national-scale optical backplane. For example, the TeraGrid network, which is supported by the National Science Foundation for scientific research, has a 40-Gb/s bit rate between its major sites [9]. While grid computing for the most part has been limited to the academic arena, there has been growing interest from the commercial sector to take advantage of the synergies that can

be attained. As this application expands to businesses, it will be accompanied by a surge in demand for high-bandwidth pipes.

The demands of grid computing in supporting research in “e-science” areas such as high-energy physics, genomics, and astrophysics are expected to grow to exabyte data sets and petaflop computation over the next decade, requiring terabit link capacity [10]. For example, in some high-energy physics experiments, multiterabyte data files need to be disseminated to multiple locations in a very short period of time [11]. Such applications require on the order of terabit per second capacity, but for relatively short periods of time (minutes to hours). The outgrowth of this expected demand for huge capacity is the “Lambda Grid” concept, where the network wavelengths themselves are resources that need to be scheduled. In this model, rather than simply providing a static optical layer for sharing computing resources, lightpaths will be dynamically established as needed. Several projects are currently investigating supporting data-intensive applications through advanced dynamic optical networking [11]–[14].

The grid model also has been extended to military applications. Two critical components of the “network-centric warfare” concept are the “sensor grid” and the “global information grid” (GIG) [15], [16]. The sensor grid comprises both active and passive sensors that are deployed in the air, undersea, and on the ground to provide battlespace awareness. For example, sensors will be used to monitor troop position, the environment, etc. Sensor data are collected and securely transmitted by the GIG, which is a collection of wireless, satellite, and wired networks that span the globe. The GIG must be capable of rapidly delivering large amounts of data to generate an integrated view of the battlespace through real-time data fusion, synchronization, and visualization. An important goal of the GIG is to provide war fighters and planners with on-demand access to this information from any operating point in the world.

### III. TODAY’S CORE NETWORK ARCHITECTURE

While there is no single canonical core network architecture, the deployments of several new commercial and government long-haul networks over the past few years have several common aspects. Before describing the architecture of these networks, we briefly summarize the architecture of legacy networks, as a point of comparison.

Legacy backbone networks are typically optical–electrical–optical (O–E–O) based, with all traffic routed through a node being converted to the electronic domain, regardless of whether or not the traffic is destined for that node. The line rate of these legacy networks is generally 2.5 Gb/s, or in some cases, 10 Gb/s, with a total capacity per fiber of 50–200 Gb/s. The transponders are fixed; i.e., a given transponder transmits and receives one particular wavelength. (A transponder is a combination transmitter/receiver card that has a short reach interface on the client side and a WDM-compatible signal on the network side.) Almost all of the traffic is substrate, i.e., the traffic rate is lower than the line rate, and synchronous optical network/synchronous digital hierarchy (SONET/SDH)-based O–E–O grooming switches are used to

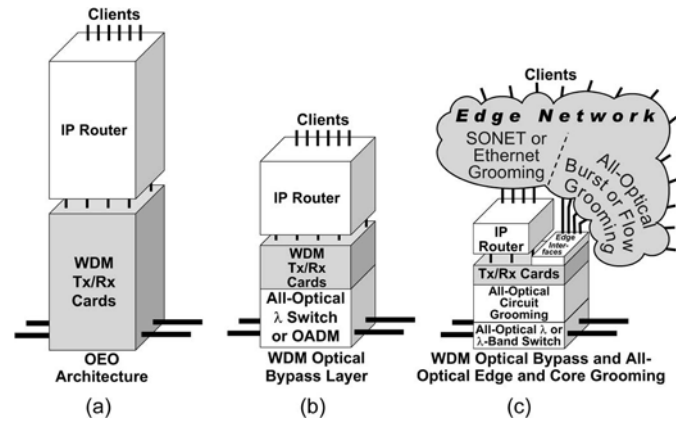


Fig. 1. Network node evolution. Architecture (a) represents a legacy node, architecture (b) represents the current state of nodal architecture, and architecture (c) is a possible evolution of the node, as detailed in the text. Note that the amount of electronics (i.e., the IP router size and the number of Tx/Rx cards) continues to decrease as the node evolves.

pack the wavelengths. (Grooming will be further described later in this paper.) The grooming switch may also be used as a core switch operating on all wavelengths entering the node and thus may be very large. A typical legacy network node is depicted in Fig. 1(a), with an IP router used as the grooming switch.

The network deployments of the past few years represent a major departure from this legacy architecture. The most significant development is the advent of optical bypass, where traffic transiting a node can remain in the optical domain as opposed to undergoing costly O–E–O conversion. Optical bypass is enabled by the combination of long (or ultralong) optical reach and all-optical switching nodes. Optical reach is the distance that an optical signal can travel before the signal quality degrades to a level that necessitates regeneration. In recently deployed backbone networks, the optical reach is on the order of 1500–4000 km, which has been achieved through the use of Raman amplification, advanced modulation techniques, and powerful forward error correcting codes. (Legacy networks use only erbium-doped fiber amplifier (EDFA)-based amplification and have an optical reach of about 500 km.)

The core all-optical network elements being deployed are optical add/drop multiplexers (OADMs) at degree-two nodes (i.e., nodes with two incident links) and all-optical switches or multidegree OADMs (OADM-MDs) at higher degree nodes. An OADM-MD of degree  $D$  is functionally similar to  $D \cdot (D - 1)/2$  OADMs, where optical bypass is supported in all directions through the node [17], [18].

In an all-optical network element, transponders are needed only for the add/drop traffic at a node, including regenerations, with the transiting traffic remaining in the optical domain. In current core networks, these elements usually operate on a per-wavelength granularity; they can selectively drop any combination of wavelengths and can be remotely reconfigured through software without affecting any connections already established. We will address the architecture of these network elements in more detail in Section VI-A.

Optical bypass has resulted in up to 90% of the required O–E–O regenerations being eliminated as compared to legacy

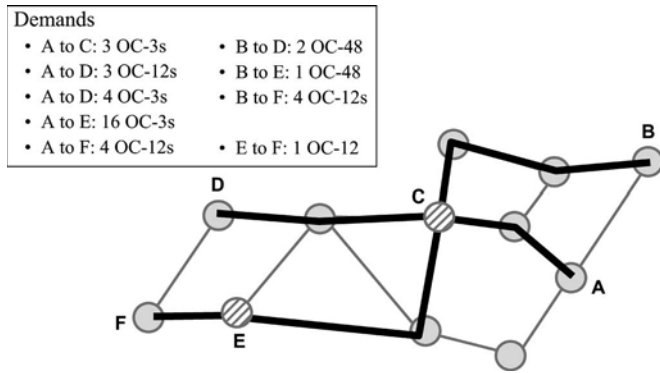


Fig. 2. Assume that the line rate is OC-192, and the demands are listed above. If end-to-end grooming is used to carry these demands, eight OC-192s are required, which correspond to the eight source/destination pairs. If intermediate grooming is used, with nodes C and E serving as intermediate grooming sites, only five OC-192s are required, as shown in the figure by the thick lines.

networks [19]. When a signal does need to be regenerated, either back-to-back transponders or regenerator cards are used; regeneration architectures are discussed in more detail in Section VI-C.

The line rate of recently deployed backbone networks is 10 Gb/s, with support for future upgrades to 40-Gb/s wavelengths. The capacity per fiber is on the order of 0.8–3.2 Tb/s (the capacity varies depending on wavelength spacing and whether one or both of the *C*-band and *L*-band are used). These networks typically employ tunable transponders. As the line rate has increased, so has the need for dispersion compensation. In recently deployed systems, this has been accomplished by installing dispersion compensating fiber (DCF), with inverse dispersion relative to the transmission fiber, at various sites along each link. DCF is generally expensive, has high loss, and provides only a static means of compensation. Depending on the transmission fiber, the DCF may not match the dispersion slope across the full band of channels, leading to imprecise compensation; this becomes more of a problem with higher line rates. As networks evolve, electronic dispersion compensation (EDC) is likely to be used as an enhancement to, or a replacement of, the DCF strategy. EDC can be deployed on a per-channel basis (e.g., as part of the transponder) and can be dynamically tuned over a range of dispersion levels to match the characteristics of a connection.

Much of the traffic that is delivered to today's core networks still does not fill an entire wavelength. As in the legacy architecture, some or all of the network nodes are equipped with SONET/SDH grooming switches and/or IP routers to efficiently pack the wavelengths. Grooming switches provide a degree of configurability, as wavelengths can be repacked in response to changing demands. Since all-optical network elements provide the core switching function, the grooming switch is not needed for this purpose, and can be appreciably smaller than in the legacy architecture [20]. Note that small nodes may not have a grooming switch, in which case their substrate traffic is backhauled to a remote grooming switch.

If grooming is done only at the network ingress and egress points (i.e., edge grooming), then any given wavelength can carry only traffic with the same endpoints; this can be very inefficient, depending on the level of traffic. Thus, intermediate

grooming is also typically implemented to produce more efficient packing, as illustrated in Fig. 2. In the figure, wavelengths are assumed to carry an OC-192; the traffic set is shown in the figure. If grooming were only performed at the demand endpoints, then eight OC-192 connections would need to be established, one corresponding to each of the eight source/destination pairs. If intermediate grooming is permitted, then one OC-192 is used to carry all of node A's traffic to node C, and one OC-192 carries all of node B's traffic to node C. At node C, further grooming occurs, such that one OC-192 is produced carrying all of the A–D and B–D traffic, and one OC-192 carries all of the traffic destined for nodes E and F. Node E serves as another intermediate grooming site. The traffic from A and B that is destined for F is groomed with the traffic from E to F, and an OC-192 is sent from E to F. This is a total of five OC-192 connections, as opposed to eight with no intermediate grooming. Furthermore, these five OC-192s traverse fewer hops than the eight, so there is an even larger bandwidth savings.

Another option for packing wavelengths is a multiplexer, or mux, card. For example, in a typical "quad card," four OC-48s on the client side are packaged into a single OC-192 on the network side. Multiplexers are generally used to backhaul substrate traffic from a small node to a node equipped with a grooming switch, or to carry substrate traffic end-to-end. Multiplexers are less costly than grooming switches, however, they provide much less flexibility. The groupings of circuits cannot be changed without manual intervention (unless there is some type of adjunct edge switch) and over time can lead to inefficiencies as circuits are brought up and down. Furthermore, as discussed above, when used to carry substrate traffic end-to-end, the lack of any intermediate wavelength packing may be very inefficient.

The level of IP traffic is growing in current networks. Core routers statistically multiplex the IP packets and generate 10- or 40-Gb/s pipes into the backbone. It is generally not efficient to create direct pipes between every pair of core routers, thus, IP traffic may be processed by multiple core IP routers. This is similar to the idea of intermediate grooming.

Fig. 1(b) represents the nodal architecture in today's backbone networks. Note that the relative amount of transponders (Tx/Rx cards) and the relative size of the grooming switch can be significantly smaller than in the legacy architecture. Optical bypass has played a major role in providing scalability in today's networks.

#### IV. NETWORK STUDY

One question going forward is how well the network architecture of today will scale with a tenfold growth in capacity. To investigate this, we performed a network study where the technologies and architectures enumerated in the previous section are extrapolated over a range of network capacities. The ramifications of this capacity growth are evaluated in terms of optimal line rate, switching granularity, and traffic grooming.

The network used in this study is representative of a typical North American backbone network. It consists of 55 nodes and 70 links, and the average nodal degree is 2.55; 60% of the nodes

have degree two, 25% degree three, and 15% degree four. The average link length is 470 km, with a minimum link length of 25 km and a maximum link length of 1110 km.

#### A. Demand Set Assumptions

Six demand scenarios were included in the network study. The aggregate demand totals in each of the scenarios were 1, 5, 10, 20, 50, and 100 Tb/s. (The aggregate demand is calculated by summing the total bidirectional traffic sourced in the network.) Today's networks are being designed for an aggregate demand around 10 Tb/s, thus, a tenfold growth was considered. The lower capacity networks were included in the study to better illustrate technology trends. It is expected that as core networks evolve, a growing and significant portion of the traffic will be IP; this was implicitly assumed in the study, as will be highlighted later.

A threefold process was used to scale up the traffic. First, some demands were assumed to scale up in traffic rate; e.g., upgrade a 2.5-Gb/s connection between a pair of endpoints to a 10-Gb/s connection. Second, new demands were added where the endpoints for the new traffic were selected from a distribution based on the existing traffic. Third, in each scenario, 5% of the demands were modified to have arbitrarily chosen endpoints. The main point is that the traffic from one scenario to the next was not a simple multiplication of the existing traffic.

Roughly 40% of the traffic was selected to require 1 + 1 dedicated client protection, with the work and protect paths fully link and node disjoint; the remaining 60% of the traffic was unprotected. In tabulating the aggregate traffic demand, protected traffic contributed twice, i.e., both the work and protect connections were counted. (Note that studies were also done where the protected traffic percentage ranged from 10% to 70%; the results were not significantly different from what is presented here.)

#### B. Architectural Assumptions

The network designs were based on optical bypass technology, where OADMs and OADM-MDs were used wherever they were justified based on cost. If a particular node did not have enough transit traffic in a scenario to justify an optical bypass element, then back-to-back optical terminals were used at that node instead.

The system line rate considered in the designs ranged from 2.5 to 640 Gb/s as indicated in Table I. (These are the rates carried by a single wavelength. The 160- and 640-Gb/s rates are not currently viable technologies and pose many technological challenges; one of the purposes of the study is to determine when, or if, such technologies might be warranted.) Higher rate signals are more susceptible to physical impairments such as chromatic dispersion, polarization mode dispersion (PMD), and nonlinear effects, resulting in shorter optical reach as the line rate increases. The optical reach assumed in the study for each line rate is shown in Table I.

The number of wavelengths per fiber also affects optical reach. For example, as wavelengths are packed more closely together, the launch power may be decreased to avoid nonlinear effects, which could lead to shorter optical reach. Nevertheless,

TABLE I  
OPTICAL REACH ASSUMED FOR EACH LINE RATE

Line Rate	Optical Reach
2.5 Gb/s	4,000 km
10 Gb/s	3,000 km
40 Gb/s	2,000 km
160 Gb/s	1,500 km
640 Gb/s	1,000 km

the reach assumptions shown in Table I were used over the range of fiber capacities. This implicitly assumes that technology will improve in the future such that the same reach can be maintained even with denser wavelength packing.

The optical reach assumptions shown in Table I warrant further discussion. First, the optical reach assumed for 2.5 and 10 Gb/s is in line with the parameters of current commercial systems (e.g., [21]). While longer optical reach at these rates is technologically possible, it has been shown that continuing to increase the reach actually results in a more costly network; i.e., the relatively small number of regenerations that are eliminated with further reach do not offset the extra cost of the more advanced technology [19]. Second, the optical reach that will ultimately be attained for 40 and 160 Gb/s is still an area of ongoing research. Much of the research has focused on using carrier-suppressed return-to-zero differential phase-shift keying (CSRZ-DPSK) to attain long reach at high bit rates [22], [23]. While Table I indicates the baseline assumptions that were made in the study, the sensitivity of the results for 40 and 160 Gb/s line rates over a range of 1000 to 3000 km optical reach is considered in Section V-A3. Finally, the optical reach assumed for 640 Gb/s is extremely optimistic. However, as will be shown, even with this aggressive assumption, 640 Gb/s never proves economical in any of the scenarios of the study; thus, considering a range of shorter reaches is not necessary.

Even with ultralong reach, regeneration was needed for some of the traffic. When needed, regeneration sites were selected on a per-demand basis. (In the 640-Gb/s line rate case, where the optical reach was shorter than five of the link lengths in the network, dedicated regeneration sites with back-to-back optical terminals were added along these links.) Routing was performed to minimize the regeneration in the network. O-E-O regenerator cards were used as opposed to back-to-back transponder cards. (Regenerator cards are similar to back-to-back transponder cards, except the two short reach interfaces are eliminated to reduce the cost.)

The average end-to-end path length of a demand ranged from 1300 km in the 1-Tb/s scenario to 1900 km in the 100-Tb/s scenario. The increase in path length reflects the growing amount of IP traffic, which tends to be more distant independent. For protected demands, the average length of the protect path was on the order of 3000 km. (In the 1-Tb/s scenario, the maximum end-to-end working path length is 7200 km, and the maximum end-to-end protect path length is 9250 km; these numbers are 8500 and 11 000 km, respectively, for the 100-Tb/s scenario.)

All of the traffic was assumed to undergo some type of grooming at the edge of the network. This could be, for

example, IP packets being processed by a router, or SONET traffic passing through a grooming switch. All nodes were equipped with grooming devices, so that backhauling to a grooming site was not required. The network design algorithm in part selects intermediate grooming sites for a demand based on where that demand needs to be regenerated. Regeneration is essentially obtained for free in these instances.

### C. Cost Assumptions

The cost of a transponder card was assumed to increase by a factor of 2.5 for every quadrupling of the line rate, except for a factor of 2 going from 2.5 to 10 Gb/s line rate (10 Gb/s is already a mature technology). Thus, the effective card cost per bit per second decreases as the line rate increases. Note that the factor of 2.5 is an aggressive assumption. While this is the cost target for 40-Gb/s cards, it still has not been achieved [24], [25]. It will be even more difficult to attain at higher line rates. We assume that the cost of the EDC is included in the transponder cost. The regenerator card cost was assumed to be 1.5 times that of a single transponder card.

In each scenario, the costs of the in-line amplifiers and the nodal pre- and postamplifiers were a function of the required maximum fiber capacity in the network in total bits per second; for every doubling of the maximum fiber capacity, the amplifier cost was increased by 25%. The cost increase factors used for amplifiers versus transponders implicitly capture the observation that optics scales better than electronics.

Currently, IP router costs tend to be two to three times higher than SONET switches [26]. It is not clear how much of this difference is due to immature large-router technology or other costs, such as software licensing; the premium for IP routers may very well shrink over time. Thus, the costs assumed for the switch fabric and grooming line cards were more in line with current SONET costs. Using current-level IP costs increases the network cost, although the cost comparisons relative to line rate do not change significantly. This point will be revisited in Section V-C.

The cost of the core switching elements, i.e., the switch fabrics of the OADMs and OADM-MDs, was a small percentage of the overall network cost, especially as the aggregate demand increased. Any cost scaling assumptions were immaterial to the final results.

## V. NETWORK STUDY RESULTS

This section includes numerous results from the network study. The discussion is broken into sections on transport line rate, switching granularity, and grooming.

### A. Transport Line Rate

As indicated above, the line rates considered in the study were 2.5, 10, 40, 160, and 640 Gb/s. Some rates were not tested for a particular demand scenario because they would produce too costly a design or require too many wavelengths per fiber. In selecting the optimal transport line rate for a particular demand scenario, we need to consider cost, technological feasibility,

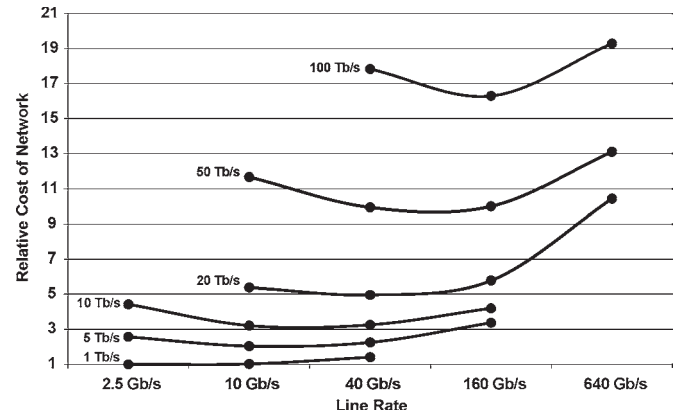


Fig. 3. Relative network cost for each of the six demand scenarios as a function of line rate. As the aggregate demand increases, the lowest-cost line rate increases.

TABLE II  
COST BREAKDOWN BY EQUIPMENT TYPE FOR THE  
1-AND 100-Tb/s SCENARIOS

	1 Tb/s Demand (10 Gb/s Line Rate)	100 Tb/s Demand (40 Gb/s Line Rate)
Optical Amplifiers	40%	9%
Nodal Optical Network Elements	17%	1%
Grooming/Routing	17%	40%
Transponders	23%	35%
Regenerator Cards	3%	15%

optical reach, and impact on switching and grooming. The first three points are discussed in this section, and the latter two points in Sections V-B and V-C, respectively.

1) *Network Cost*: Fig. 3 plots the relative total cost of the network as a function of line rate for each of the six demand scenarios. The network costs include amplifiers, nodal optical network elements, grooming switches and routers, and transponder and regenerator cards. The breakdown of the costs for the major network components is provided in Table II for the 1-Tb/s demand scenario with 10-Gb/s line rate and the 100-Tb/s demand scenario with 40-Gb/s line rate. As shown in the table, the cost of the electronics dominates as the network grows in size. This will be explored further in Section V-C.

The line rate that produced the lowest cost network for each demand scenario is summarized in Table III. (In some scenarios, two different line rates resulted in approximately the same lowest cost.) From a cost perspective, the optimal line rate increases with larger aggregate demand. As discussed previously, higher speed transponders are cheaper on a per bit per second basis. With low aggregate demand, this benefit is offset by inefficiencies in packing the wavelengths. As the aggregate demand increases, the wavelengths can be packed more efficiently, and the lower relative transponder costs dominate.

Clearly, the results are dependent on the cost factors assumed. One important assumption was the cost ratio of a 160-Gb/s transponder to a 40-Gb/s transponder, which was assumed in the study to be 2.5. Fig. 4 provides a sensitivity analysis with respect to this ratio for the 100-Tb/s demand scenario. For the assumptions made in this particular study, if this ratio is greater than  $\sim 2.8$ , then the 40-Gb/s line rate, as opposed to 160 Gb/s, would produce the lowest cost design

TABLE III  
OPTIMAL LINE RATE FROM A COST PERSPECTIVE FOR  
EACH DEMAND SCENARIO

Aggregate Demand	Optimal Line Rate
1 Tb/s	2.5 Gb/s, 10 Gb/s
5 Tb/s	10 Gb/s
10 Tb/s	10 Gb/s, 40 Gb/s
20 Tb/s	40 Gb/s
50 Tb/s	40 Gb/s, 160 Gb/s
100 Tb/s	160 Gb/s

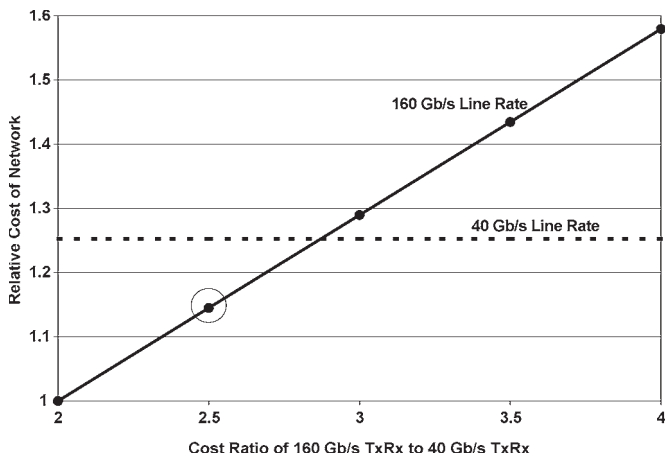


Fig. 4. Relative network cost for the 100-Tb/s demand scenario as a function of the transponder cost ratio for 160 Gb/s relative to 40 Gb/s. The circled point indicates the ratio assumed in the network study, which resulted in 160 Gb/s being lower cost. However, if the ratio is greater than  $\sim 2.8$ , then the 40-Gb/s line rate provides a lower cost network.

for the 100-Tb/s demand scenario. This illustrates the cost targets that are required for the 160-Gb/s line rate to provide an economic benefit at such traffic levels.

The costs plotted in Figs. 3 and 4 include only capital costs. Another important cost component is operational cost, which is dependent on many factors, including power consumption, space requirements, and amount of deployed equipment. In addition to lower cost per bit, higher line rate cards are expected to require less power and space per bit. For example, the targets for 40-Gb/s technology are 50% smaller size and 20–40% less power, as compared to four 10-Gb/s cards [27]. Higher bit rate also means fewer wavelengths and less transponder cards to deploy and manage. Consider the 100-Tb/s demand scenario: Although the reach was 25% shorter in the 160-Gb/s system as compared to the 40-Gb/s system, the 160-Gb/s system required roughly half the number of regenerator cards. While it is difficult to fully capture operational costs, in general, it is expected that they will decrease with higher line rate.

Overall, then, for the assumptions used in the study, 160 Gb/s is the preferred line rate for the 100-Tb/s scenario from a cost perspective. (Note that even with the aggressive assumptions used for the 640-Gb/s cost and reach, such systems were never found to be cost-effective.) However, this is predicated on the supposition that technology will continue to scale aggressively in both cost and performance. In reality, this will be very

challenging, as will be discussed in later sections. If these aggressive targets are not met for 160 Gb/s, then, in fact, the 40-Gb/s line rate is the “sweet spot” for networks carrying traffic ranging from 10 to 100 Tb/s.

2) *Capacity Requirements and Modulation Formats*: Innovations will be needed to meet the capacity requirements of the next-generation core network. Table IV indicates the number of wavelengths and the corresponding capacity required on the maximally loaded link in each demand scenario; the average link load was generally about 50%–60% of the maximum link load. (In almost all of the scenarios, the number of required wavelengths in the system equaled the number of routed wavelengths on the maximally loaded link. In a small number of scenarios, the number of required wavelengths was slightly greater than the maximal load because of wavelength contention.)

Clearly, the required link capacity increases with demand levels. Increased capacity can be attained by increasing the available optical spectrum (e.g., also using the *L*-band and even the *S*-band) and/or increasing the spectral efficiency of the modulation scheme. Spectral efficiency is defined as the ratio of the information bit rate to the total bandwidth consumed. To see the relationship between link capacity and spectral efficiency (assuming a single fiber-pair per link), refer to the first two rows of Table V. At an aggregate demand of 100 Tb/s, if just the *C*-band is used, the spectral efficiency required to meet the capacity needs is roughly 3.5–4.5 bits/s/Hz; it is half of this if both the *C*-band and *L*-band are used.

We first consider the theoretical limits on fiber spectral efficiency. The analysis in [28] showed that for a nonlinear system that is limited by cross-phase modulation (XPM), a lower bound on the maximum spectral efficiency is about 4 bits/s/Hz (the exact bound will depend on the specific system assumptions regarding dispersion, number of channels, etc.). In [29], it is shown that if the nonlinear system is limited by four-wave mixing (FWM) as opposed to XPM, the spectral efficiency limit is roughly 6 bits/s/Hz. (In both analyses, a single polarization system is assumed.) Both of these results are for a system with coherent detection. Today, most systems use direct detection, which is less complex to implement, but which yields significantly lower spectral efficiency (e.g., less than 0.5 bits/s/Hz). (Note that for very high rate systems, factors other than XPM and FWM need to be considered in calculating spectral efficiency limits. More research is needed in this area.)

While the required capacities are theoretically attainable as shown by these results, multilevel modulation formats will be required. See Table V for the capacities that can be attained in the *C*-band for a range of modulation formats and channel spacings. (This table provides examples of modulation techniques that can be used to attain a given spectral efficiency; it is not meant to be an exhaustive list.) Binary modulation has a spectral efficiency limit of 1 bit/s/Hz (assuming single polarization). Binary modulation in the *C*-band provides sufficient capacity only up to the 20-Tb/s aggregate demand scenario. To achieve the desired link capacities for 100 Tb/s aggregate demand (with a single fiber-pair per link), advanced techniques such as 16-level modulation [30], in both the *C*-band and *L*-band, will be needed.

TABLE IV  
LINK CAPACITY REQUIREMENTS FOR EACH DEMAND SCENARIO AND LINE RATE

Line Rate	Aggregate Demand (Sum of the traffic sourced in the network)											
	1 Tb/s		5 Tb/s		10 Tb/s		20 Tb/s		50 Tb/s		100 Tb/s	
	Max # WL	Max Req. Link Cap (Tb/s)	Max # WL	Max Req. Link Cap (Tb/s)	Max # WL	Max Req. Link Cap (Tb/s)	Max # WL	Max Req. Link Cap (Tb/s)	Max # WL	Max Req. Link Cap (Tb/s)	Max # WL	Max Req. Link Cap (Tb/s)
2.5 Gb/s	55	0.14	233	0.58	482	1.2						
10 Gb/s	24	0.24	67	0.67	124	1.2	251	2.5	666	6.7		
40 Gb/s	12	0.48	23	0.92	42	1.7	70	2.8	185	7.4	353	14.1
160 Gb/s			13	2.08	19	3.0	27	4.3	58	9.3	111	17.8
640 Gb/s							22	14.1	24	15.4	36	23.0

TABLE V  
POSSIBLE EVOLUTION OF OPTICAL FIBER TRANSMISSION CAPACITY

C-Band Capacity (Tb/s)	0.8	1.6	3.2	6.4	12.8				
Spectral Efficiency (b/s/Hz)	0.2	0.4	0.8	1.6	3.2				
Bit-Rate per $\lambda$ (Gb/s)	10	10	40	10	40	40	160	40	160
Inter- $\lambda$ Spacing (GHz)	50	25	100	12.5	50	25	100	12.5	50
No. of $\lambda$ 's in C-Band	80	160	40	320	80	160	40	320	80
Example of Modulation Format	Binary OOK		DPSK or Duobinary		4-Level (DQPSK)		16-Level		

Achieving long optical reach with high spectral efficiency is an area of ongoing research. Reference [31] summarizes numerous experiments that have been performed in this area and uses the results to plot spectral efficiency versus attained reach for various modulation formats. We highlight some of the results here. Reference [32] reported achieving 2.0 bits/s/Hz, using return-to-zero differential quadrature phase-shift keying (RZ-DQPSK) combined with polarization-division multiplexing (PDM); the optical reach, however, was only 400 km. PDM may be challenging to implement in a network with optical bypass because it may be difficult to insert an orthogonally polarized signal (i.e., orthogonal to the surrounding channels) in a node with through channels. Reference [33] achieved 1.14 bits/s/Hz with a 300-km reach but without the complexities of PDM. Reference [22] used CSRZ-DPSK, without PDM, to transmit  $160 \times 40$  Gb/s channels with 50-GHz spacing in the C- and L-bands, for a spectral efficiency of 0.8 bits/s/Hz; the demonstrated optical reach was 3200 km. Another promising technique to mitigate fiber transmission impairments and achieve long optical reach at high spectral efficiencies is based on optical phase conjugation [34], [35].

In addition to questions regarding modulation format, more research is needed to determine whether 40 or 160 Gb/s is technically more feasible at high capacities (e.g., How does  $400 \times 40$  Gb/s compare to  $100 \times 160$  Gb/s?); 160 Gb/s is more susceptible to impairments such as PMD. Furthermore, the more rapid pulse broadening of the 160-Gb/s signal leads

to greater susceptibility to distortions such as intra-cross-phase modulation (IXPM) and intra-four-wave mixing (IFWM) [36]. In addition to these impairments, implementing 160 Gb/s requires more complex timing in the transmitter and receiver, more complex signal monitoring, etc.

Furthermore, as shown in Table IV, the maximum required fiber capacity is higher with a 160-Gb/s line rate as compared to a 40-Gb/s line rate. This is partially due to greater inefficiencies in filling the wavelengths at a higher line rate. The need for more wavelength packing with a 160-Gb/s line rate also results in less routing freedom. In addition, the shorter optical reach of 160 Gb/s provides fewer routing options that meet the minimum number of regenerations, thereby resulting in less opportunity to load balance. (The network study used lowest cost routing, with load balancing. The load on the maximally loaded link could be lowered somewhat, but with increased cost.)

Overall, from an implementation standpoint, the 160-Gb/s line rate will likely prove challenging.

3) *Optical Reach and All-Optical Regeneration*: The optical reach is a key assumption in the network study. Attaining long reach will be extremely challenging at high bit rates, especially when the number of wavelengths per fiber is large. Our network study postulated a 2000-km reach for the 40-Gb/s line rate and a 1500-km reach for 160 Gb/s. (With these assumptions, in the 100-Tb/s demand scenario, the average percentage of nodal all-optical bypass is 68% at the 40-Gb/s line rate and 59% at 160 Gb/s.) Fig. 5 illustrates how the 100-Tb/s network

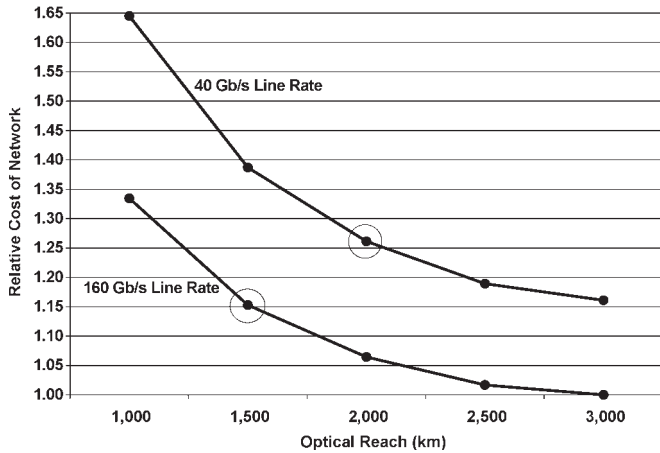


Fig. 5. Relative cost of the network for the 100-Tb/s demand scenario as a function of optical reach for 40- and 160-Gb/s line rates. The circled points indicate the optical reach values assumed in the network study.

cost changes for different reach assumptions for these two line rates, assuming all of the component costs remain unchanged. For example, if the optical reach for 160 Gb/s is 1000 km as opposed to 1500 km, then 40 Gb/s produces a lower cost solution than 160 Gb/s.

With the optical reach assumptions used in the study, regeneration accounts for less than 5% of the overall network cost in the 1-Tb/s demand scenario; the percentage gradually grows to 15%–20% in the 100-Tb/s demand scenario. In the latter scenario, if the reach of the 160-Gb/s line rate system is only 1000 km, the regeneration cost percentage jumps to 30%.

O–E–O regenerators were assumed in the study, however, all-optical regenerators that operate on individual wavelengths are close to being commercially viable. (It would be highly desirable to have multiwavelength all-optical regenerators; however, this is still a research goal [37], [38].) Various techniques for all-optical regeneration are described in [39] and [40]. It is expected that the cost of such regenerators will scale well with line rate, making them relatively more cost effective at higher line rates (up to a point), even more so than transponders (i.e., the cost increase factor in going from a 10- to a 40-Gb/s all-optical regenerator is expected to be less than the 2.5 factor assumed for O–E–O regenerators.) Replacing O–E–O-based regenerators with their all-optical counterparts will perhaps save on the order of 5%–10% of the cost of future networks. Furthermore, all-optical regenerators should have relatively low power consumption, which will also make such a solution more scalable. It is desirable that they be integratable on a chip, so that they will scale well in size also. These potential benefits in cost, power, and size still need to be proven in practice.

Being able to change wavelengths when regenerating is an important factor in achieving high network efficiency [41]. In the study, the regenerator cards were assumed to be capable of full any-to-any wavelength conversion. It is important that all-optical regenerators provide this same wavelength conversion capability as well.

Note that the first commercial all-optical regenerators are likely to be compatible only with modulation formats such as binary on–off keying. However, as was discussed in the previ-

ous section, more advanced modulation formats will be needed to meet the capacity requirements. If all-optical regenerators are to be a part of the next-generation core network, they will need to be compatible with these advanced formats. Research is already underway in this direction [42]–[44], although the solutions thus far are fiber-based, which have the drawback of being difficult to integrate.

Some all-optical regenerators provide only 2R regeneration, i.e., reamplification and reshaping, as opposed to 3R, which includes retiming as well. If 2R regenerators prove to be significantly cheaper than 3R, then they can be deployed at some points along a path, with 3R regenerators used intermittently to clean up the timing jitter.

4) *Multiple-Fiber-Pair Solution*: Thus far, we have been implicitly assuming single-fiber-pair solutions, where all of the bidirectional traffic of a link is carried on one fiber-pair, possibly in both the *C*- and *L*-bands. An alternative solution is to use multiple fiber-pairs, so that each individual fiber has lower capacity requirements, thus easing the challenging spectral efficiency goals. Using multiple fibers may provide lower up-front costs as fibers can be lit when needed [45]. In addition, less dense wavelength spacing is likely to result in longer reach. However, there are some disadvantages to this approach. First, multiple fiber-pairs may not be available, especially in areas that tend to be heavily loaded in many carrier networks. This approach requires another set of amplifiers for each fiber-pair, which is inefficient from a cost and maintenance perspective (although the increased cost could be offset by the use of simpler technology and the potential longer reach). If it is desired that there be full all-optical configurability and connectivity in the network, this solution would lead to switches with a larger number of fiber ports (e.g., an OADM at a degree-two node would have four fiber ports if there were two fiber-pairs per link). Finally, the use of multiple fiber-pairs does not alleviate the wavelength port count on a core switch, which is the subject of the next section. Overall, a single-fiber-pair solution is preferred, especially from an operational standpoint, however, a two-fiber-pair system may need to be deployed if the single-fiber-pair capacity requirements prove too challenging.

## B. Core Optical Switch Granularity

1) *Wavelength Switching*: The choice of transmission line rate greatly impacts the core optical switching technology. Table VI indicates the number of bidirectional wavelength ports required on the largest core optical switch fabric at any of the nodes, for each scenario in the network study. (The core optical switch in the study is the OADM or OADM-MD, not the edge grooming switch or IP router.) Note that we are referring to the size of the switch fabric, not the number of external switch ports. For example, some switch architectures have  $N$  external fiber ports, with each fiber internally demultiplexed into  $W$  wavelengths, which leads to a switch fabric of size  $N \cdot W$ .

The counts in Table VI include ports for all wavelengths on the network fibers entering the node, as well as ports for the *actual* number of add/drop wavelengths at the node, including regeneration. If the switch is designed to accommodate 100%

TABLE VI  
SIZE OF LARGEST OPTICAL SWITCH FABRIC REQUIRED IN EACH SCENARIO (IN TERMS OF WAVELENGTH PORTS)

Line Rate	Demand Scenario					
	1 Tb/s	5 Tb/s	10 Tb/s	20 Tb/s	50 Tb/s	100 Tb/s
2.5 Gb/s	294	1228	2568			
10 Gb/s	133	399	729	1492	3879	
40 Gb/s	74	149	251	424	1063	2056
160 Gb/s		79	114	169	333	643
640 Gb/s				134	144	214

drop from each node, then, for this study, the largest required switch sizes would be about 50% larger than what is shown in the table. (Furthermore, the largest nodal degree in the study was four; higher degrees would lead to larger required switches.)

Let us focus the discussion on the 100-Tb/s demand scenario, with 40- and 160-Gb/s line rates. With the 40-Gb/s line rate, the total number of wavelengths needing to be switched in the largest switch fabric is over 2000. Given that today's all-optical switches have switch fabrics capable of switching several hundred wavelengths in practice, it is not clear whether a three- to fourfold increase in size is economically feasible. Furthermore, even if it can be built, the ramifications for size and power requirements need to be considered.

At a line rate of 160 Gb/s, the maximum switch fabric operates on roughly 650 wavelengths. This size switch should be readily achievable. Thus, from the perspective of switch fabric size, and assuming the switching is done on a per-wavelength basis, 160-Gb/s wavelengths are preferred over 40-Gb/s wavelengths. However, as noted above, there are technical challenges in achieving 160-Gb/s transmission with a reasonable optical reach.

2) *Waveband Switching*: Another option is to employ waveband switching rather than per-wavelength switching. For example, in the 100-Tb/s scenario, 40-Gb/s wavelengths could be used in combination with core switches that operate on the granularity of 160-Gb/s wavebands. Bands of four 40-Gb/s wavelengths would be switched as a group such that the size of the switch fabric is essentially the same as in the case of 160-Gb/s wavelengths; the difference is that the switch operates on wavebands rather than wavelengths. This solution takes advantage of the less complex 40-Gb/s technology while maintaining a moderate switch size.

Looking at Table VI, if we assume that current core networks are being designed for a total demand of 10 Tb/s with both 10- and 40-Gb/s wavelengths, then wavebands are not necessitated from a switch-size feasibility standpoint. However, waveband switching has been implemented in some networks for purposes of reducing the switch cost, although this is more common in the metro area, which is very cost sensitive. The viability of waveband switching is discussed in [46]–[49].

Wavebands lead to some loss of network flexibility because switching and add/drop is performed on a coarser granularity. However, studies have shown that through the use of intelligent algorithms, the inefficiencies resulting from wavebands are

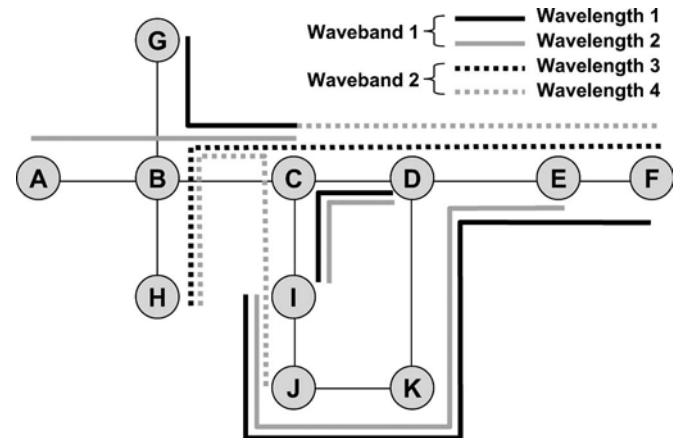


Fig. 6. In this example, there are two wavebands per fiber, with each waveband composed of two wavelengths. Various types of waveband grooming are illustrated: The wavelengths carrying G–F traffic and A–C traffic merge into the same waveband at node B; the wavelength carrying G–F traffic undergoes wavelength conversion at node C, which shifts this traffic from waveband 1 to waveband 2; at node E, the waveband drops the I–E traffic but continues on to node F with the I–F traffic.

small [50]. To efficiently pack the wavebands, it is important that intermediate waveband grooming be implemented, especially if there is a lot of network churn. As traffic is brought up and down in the network, wavebands can become more fragmented; waveband grooming is needed to keep them well packed.

Several modes of waveband grooming are illustrated in Fig. 6. In this simple example, it is assumed there are two wavebands, each composed of two wavelengths. Waveband “merging” is illustrated at node B: Wavelength 1 enters node B from link G–B, whereas wavelength 2 enters node B from link A–B; the two wavelengths exit node B in a single waveband on their way to node C. Waveband grooming with wavelength conversion is illustrated at node C (for the traffic between nodes G and F): Wavelength 1 enters node C from link B–C, where it is converted to wavelength 4; this shifts this traffic from the first waveband to the second waveband. Waveband “drop and continue” is illustrated at node E; waveband 1, which is composed of wavelengths 1 and 2, enters node E; wavelength 2 is dropped at node E, whereas wavelength 1 continues on to node F. (Wavelength 2 may or may not be reusable on link E–F, depending on the switch architecture at node E.)

The waveband-grooming functions can be performed concurrently with regeneration. For example, a regenerator card

can be used to change the wavelength of the incoming signal, so that it goes out in a different band. If the optical reach of the system is longer than what is compatible with the need for waveband grooming, however, then it would be desirable to develop banded switches that cost effectively incorporate a waveband-grooming function.

One implementation of a waveband-grooming switch is to use a two-level switch composed of both a waveband crossconnect and a wavelength crossconnect (similar to the hierarchical switch of [46]). The output ports of the wavelength crossconnect can be equipped with wavelength converters in order to exchange wavelengths between wavebands. One enhancement to a hierarchical switch is to use cyclic waveband mux/demuxes, which provides flexibility in selecting which wavebands are dropped to the wavelength switch for grooming [51].

The goal is to develop technology that improves the scalability of the network as a whole. For example, if passing through a waveband-grooming switch significantly reduces the reach of the signal, then the benefits of more efficiently packed wavebands will be negated. In addition, note that technology needs to be developed in concert with efficient network design algorithms.

While waveband switching alleviates the demands on the switching fabric, it does not reduce the number of transponders in the network. In the 100-Tb/s scenario with 40-Gb/s wavelengths, the largest amount of drop at any node (including regeneration) was about 600 wavelengths, or 24 Tb/s. The need to improve the space and power requirements, as well as the economics, will likely mandate solutions such as using one integrated multiwavelength Tx or Rx chip (i.e., a photonic integrated circuit) per waveband, e.g., [3].

### C. Substrate Traffic Grooming

As the aggregate network demand increases, the percentage of the network cost that is attributable to electronic wavelength grooming steadily grows. (Grooming in this section refers to packing of individual wavelengths, not packing wavelengths into wavebands. The grooming could be done, for example, in IP routers or SONET switches.) In the 1-Tb/s scenario, about 15% of the network cost is due to grooming; at 100 Tb/s, this percentage grows to 40% (all demands are assumed to undergo grooming at the edge; thus, the percentage grooming cost was approximately the same for a given demand scenario regardless of the line rate used). The escalating relative cost of grooming is indicative of the fact that optics scale better than electronics, which was implicitly assumed in the costs that were used. Recall that the grooming costs used in the study are more in line with current SONET grooming switches as opposed to current IP routers, with the expectation that the relative IP costs will decrease in the future. If IP router costs remain relatively high, however, then the proportion of the network cost due to grooming will be even greater. For example, if the ratio of IP costs to SONET costs remains as it is today [26] and the 100-Tb/s demand scenario was totally composed of IP traffic, then roughly 70% of the network costs would be due to IP processing in core routers.

The size of the required grooming switch clearly increases as the traffic grows. With an aggregate network demand of 1 Tb/s,

the largest required grooming switch capacity at a node is about 0.3 Tb/s. At a network demand of 100 Tb/s, assuming the traffic is largely IP, then IP routers on the order of 30-Tb/s capacity are needed. At least one commercial IP router scales to this size, although, as indicated earlier in this paper, the cost may be prohibitive. In addition, there could be scalability issues related to physical size, heat dissipation, and power requirements.

One approach for dealing with the growing size of electronic routers and switches and their associated cost and scalability issues is to implement at least some of the grooming functionality in the optical domain. One active area of research is optical packet switching (OPS), where data packets are processed all-optically (although the packet header may be processed with electronics), e.g., [52]–[55]. The need for optical buffers must be addressed for optical routers to be a viable technology [56], [57].

Rather than replacing large electronic routers with optical routers, another approach is to employ switching paradigms that reduce the amount of required electronic grooming. Examples of such techniques are listed as follows: optical burst switching (OBS) [58], [59]; light trails [60]; time-domain wavelength interleaved networking (TWIN) [61]; and randomized load-balancing [62], [63]. These schemes address the grooming of substrate traffic without requiring optical buffers or huge electronic routers; they have the additional potential benefit of being able to handle dynamic traffic without long connection setup times. The performance of these schemes in real networks still needs to be demonstrated.

1) *Edge Grooming*: Schemes for optically aggregating traffic (e.g., OBS and TWIN) typically require a scheduling protocol or need to deal with contention in the network. It is likely that the scheduling complexities and conflicts will worsen as the network size increases. Implementing such schemes may make more sense in smaller scale edge networks as opposed to a core network (assuming that the edge network has enough traffic such that statistical multiplexing gains can be realized). The edge network can be, for example, the metro or regional network connecting the various customers to the core network. (Note that the edge network can share some fiber routes or even individual fibers with the core. In the latter case, some wavelengths or wavebands would be dedicated to the edge traffic.)

Furthermore, the network edge is where deploying alternatives to electronic grooming potentially will have the largest impact on reducing overall network cost. The network study showed that a significant portion of the grooming occurs at the traffic source/destination (i.e., edge), not at intermediate nodes. For example, in the 1-Tb/s demand scenario, with 2.5-Gb/s line rate, 90% of the grooming occurs at the edge. In the 100-Tb/s demand scenario, with 40-Gb/s line rate, 95% of the grooming is at the traffic edge. Thus, developing more cost-effective and more scalable means of aggregating the traffic at the edge can have a large impact on the network cost.

Consider the architecture shown in Fig. 7, with edge networks interconnected by the core network. An edge network feeds into a small number of core network nodes for diversity. A hybrid provisioning approach can be used, where optical aggregation occurs within the edge network, with more traditional

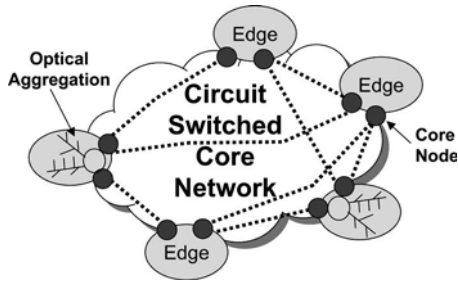


Fig. 7. In this two-tier architecture, edge networks are interconnected by the core network. Optical aggregation is implemented in the edge network, delivering traffic to the core nodes; circuit switching is implemented in the core. The edge/core interface could be all-optical or O–E–O.

circuit switching used in the core. [A similar hybrid multitiered approach was proposed by the All-Optical Network (AON) Consortium [64].] The edge network handles highly dynamic traffic, aggregating the traffic onto high-speed wavelengths, whereas connections are brought up and down in the core network on a slower time scale.

Due to the relatively smaller scale of the edge network, there are more options for optically aggregating the traffic. The aggregation scheme could be based on a switched architecture, requiring very fast switches; alternatively, a passive broadcast solution could potentially be implemented, as is advocated for efficient optical flow switching in [65]. (A flow is a consecutive sequence of packets between the same two routers.) In addition, ring architectures may be used within an edge network; thus, ring-based optical aggregation schemes, e.g., [66] and [67], can be used.

There are a couple of options for designing the interface between the edge and core networks. First, assume that optical signals pass transparently between the edge and core networks without any O–E–O conversion. The advantage of this approach is that a separate transponder is not needed at the edge/core interface. However, this means that a complex transceiver would be required at the customer premise in order to be compatible with the transmission system of the core network. This may prove to be too expensive for customer premises equipment. Another possible disadvantage of an all-optical interface is that any voids between the multiplexed data bursts will remain as signal voids in the core network, which could possibly cause amplifier transient effects.

A second option is to have an O–E–O interface separating the edge and core networks. This isolates the transmission systems of the two networks, so that lower cost equipment can be used within the edge network, where the capacity and reach requirements are much less stringent. In addition, the electronic interface can stuff bits if there is a void in the optically multiplexed signal coming from the edge. The disadvantage is the extra cost of the transponders needed at the edge/core interface (but note that these transponders would be shared among multiple customers).

The edge/core architecture affects how resource contention is handled. It is assumed that resource scheduling is done only within a given edge network, not across the whole network, to simplify the implementation of the optical aggregation scheme.

This can lead to conflicts at the destination edge network; e.g., two different regions could send traffic on the same wavelength that arrives at the same destination region at the same time. In the transparent architecture, all-optical rapidly tunable wavelength converters could be used at the destination edge network to shift one of the conflicting streams to another wavelength. Customers would need to be equipped with an array receiver (ideally integrated), so that multiple wavelengths can be received at one time. In the O–E–O interface architecture, wavelength conversion can be obtained simply by tuning the transmitter on the edge network side to a different wavelength. Again, an array receiver would be needed at the customer. Alternatively, once the signal is in the electronic domain, electronic buffering at the edge/core interface could be used to resolve conflicts, eliminating the need for array receivers.

2) *Intermediate Grooming*: Given that the bulk of grooming occurs at the network edge, a natural question is whether intermediate grooming of wavelengths within the core is needed at all. This, in general, will depend on the traffic matrix and network connectivity. To examine the efficacy of intermediate grooming in the study, we will focus on the 100-Tb/s demand scenario. If intermediate grooming is not permitted and 40-Gb/s line rate is used, the number of wavelengths on the most heavily utilized link almost doubles. If 160-Gb/s line rate is used, the number of required wavelengths increases by a factor of 4. The number of required wavelength ports on a switch increases accordingly. Clearly, in this study, intermediate grooming plays an important role in reducing the required capacities.

Given that the schemes for all-optical grooming discussed earlier may not scale to a full national-size network, some type of O–E–O intermediate grooming is needed. However, O–E–O intermediate grooming is not very effective if optical aggregation at the edge is used for all traffic. Traffic that is sourced at a node will not be able to be packed with traffic that is transiting the node due to the different grooming mechanisms. A more effective solution is to partition the traffic into the following two types: Type 1 traffic that is optically aggregated at the edge and does not undergo any intermediate wavelength grooming and type 2 traffic that uses O–E–O grooming at both the edge and at intermediate nodes (i.e., type 2 is traditional grooming). It is desirable that most traffic be type 1, so that significant cost benefits can be realized through the use of lower cost optical techniques, but to assign some traffic as type 2 to moderate the capacity and switch requirements.

Consider the following design that was performed for the 100-Tb/s, 40-Gb/s line rate scenario: 90% of the traffic was designated as type 1 and underwent only optical aggregation at the network edge; the remaining traffic used O–E–O grooming at the edge and at intermediate nodes. (Traffic partitioning was based on the fill rate of end-to-end multiplexed wavelengths.) Compared to the traditional all-type 2 grooming scenario, the network capacity requirements increased by just 3%. The big benefit is that the maximum required electronic IP router was only 3 Tb/s, which is a factor of 10 smaller than in the conventional scenario. In reality, not all traffic will be eligible for type 1 aggregation due to the necessary optical equipment not being installed or being too expensive for some customers.

Thus, the required IP router will likely be greater than 3 Tb/s. This example design is simply meant to demonstrate that a hybrid optical/ O–E–O grooming solution is feasible.

3) *Network Node Evolution*: Based on the discussion in this section, a core network node is envisioned as evolving as shown in Fig. 1(c). Note that the IP router represents a relatively smaller percentage of the network than that in Fig. 1(b) mainly due to optical aggregation occurring in the edge network as opposed to in the IP router. With an O–E–O edge/core interface, the “edge interfaces” box represents transponders and, possibly, electronic buffering and bit-stuffing electronics; with an all-optical interface, this box represents all-optical wavelength converters.

The all-optical grooming box shown in the figure above the optical switch represents circuit-based grooming (as opposed to the burst or flow grooming that occurs at the edge). For example, in a scenario with 40-Gb/s wavelengths and 160-Gb/s wavebands, this represents grooming of wavelengths into wavebands, or if the wavelength granularity was 160 Gb/s, this could represent SONET-like all-optical time-domain grooming of, for example, 40- or 10-Gb/s circuits into wavelengths (which is technologically challenging).

Handling functionality such as wavelength conversion, regeneration, and waveband grooming in the optical domain leads to a relative decrease in the number of required transponders.

## VI. CONFIGURABILITY

In the network study of the previous section, a fixed set of demands was routed on the network in each scenario to probe scalability relative to increases in capacity. While most backbone traffic today is fairly static, with connection holding times on the order of months or longer, the growing on-demand nature of applications will be accompanied by an increase in dynamic traffic. One line of thought is that if the capacity provided by the network is large enough, then bandwidth can be dedicated to connections even if it is not needed most of the time. However, with high-bandwidth applications rapidly increasing, it is unlikely that the capacity will be so plentiful that it can be squandered on a large-scale basis. Some percentage of the network will need to respond to changing demands and deliver capacity to where it is needed at a particular time. In this section, we discuss some of the ramifications of this need for configurability, with respect to switching architectures and predeployed equipment. The reconfigurability times assumed in this section are on the order of a fraction of a second or more. (See [68] for an overview of various switch technologies and associated switching speeds.)

### A. Switch Architecture

1) *First-Generation OADM-MD*: As stated in Section III, one of the core all-optical network elements being deployed in today’s networks is the OADM-MD (the OADM can be considered a special case of an OADM-MD with degree equal to 2). The first generation of OADM-MD was based on liquid crystal wavelength blockers, which are also called “dynamic spectral equalizers” or DSEs [17], [18]. In this architecture, for

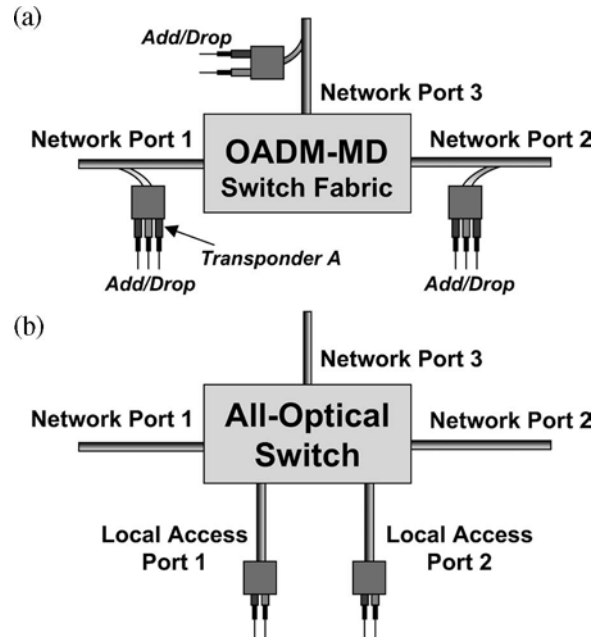


Fig. 8. (a) OADM-MD with three network ports. The add/drop traffic is tapped off of the network ports; each transponder can access only one network port. (b) All-optical switch with three network ports and two local access ports. The add/drop traffic passes through the switch fabric, allowing any transponder to access any network port. In both network elements, there are several architectures for handling the add/drop interface, as represented by the squares.

a degree- $D$  node, a total of  $D(D - 1)$  DSEs are needed; thus, the size and cost of this switch fabric scale quadratically with  $D$ . In practice, this architecture is limited to a maximum degree of about 4.

A functional diagram of this OADM-MD is shown in Fig. 8(a). The OADM-MD can be remotely configured to add/drop or bypass any given wavelength without affecting any connections already established. Traffic can be rerouted through the node while remaining in the optical domain. One important restriction, however, is that the transponders are tied to a particular network port (and, hence, to a particular network direction). Referring to Fig. 8(a), note that the add/drop traffic does not enter the OADM-MD switching fabric, but rather is tapped off of the network ports. In the figure, transponder A can be used only to add/drop to/from network port 1. It does not have access to the other network ports. Thus, any client attached solely to this transponder would be forced to enter/leave the network through network port 1. To gain flexibility, the edge device (e.g., an IP router) would need to connect to transponders on all desired add/drop paths; switching at the edge would then provide the desired reconfigurability. (For further discussion on combined edge and core configurability, see [69].)

It is worth noting that there are several architectures for handling the add/drop interface, as represented by the squares in Fig. 8, which affect network configurability. For example, the transponders could plug into a splitter/combiner, providing a “broadcast-and-select” architecture. In this architecture, the transponder slots are “colorless,” so that a transponder of any wavelength can be placed in any slot. When used with tunable transponders, this architecture is advantageous with respect to wavelength assignment and efficient transponder

utilization under dynamic traffic conditions. Another option is to have the transponders plug into a WDM mux/demux with wavelength-specific slots. While this architecture provides lower loss, it cannot take full advantage of tunable transponders and can result in more transponders being required to handle dynamic traffic. (Hybrid solutions are also possible, where the mux/demux slots can be tuned over a limited range of wavelengths [70].)

2) *Second-Generation OADM-MD and All-Optical Switches:* Second-generation OADM-MDs make use of wavelength-selective switches (WSS) [71]. For a degree- $D$  OADM-MD, a total of  $D$  WSSs are needed; i.e., linear scaling. The add/drop paths can be tapped off of the network ports as in the first-generation architecture. However, with the improved scalability of this architecture, a second option is feasible: The add/drop traffic can be passed through the switch fabric. In this case, we refer to the add/drop paths as local access ports. In addition, when the add/drop traffic is switched rather than tapped, the network element should be referred to as an all-optical switch rather than an OADM-MD.

A high-level diagram of this all-optical switch is shown in Fig. 8(b). Any transponder can access any of the network ports, which provides greater flexibility in reconfiguring the network. As described in the previous section for the OADM-MD, colorless or wavelength-specific options exist for the add/drop interface.

This architecture supports additional services as well. It allows the output of a single transponder to be multicast to multiple network ports. It also supports cost-effective optical layer protection, where a single transponder is used for either the work or protect paths, with the all-optical switch toggling between paths at the time of a failure. (This form of optical layer protection should be combined with 1: $N$  protection of the transponders.)

With this architecture, an interesting question arises as to how many local access ports should there be relative to the number of network ports (where the network and local access ports are both WDM ports). Because transponders can access any network port, and because the percentage of drop at a node is typically less than 100%, it is possible to have fewer local access ports than network ports. This can lead to wavelength conflicts however. For example, with three network ports and just two local access ports, as shown in Fig. 8(b), the same wavelength cannot be dropped from all three network ports. This issue is more of a problem when the network is heavily loaded and there are few free wavelengths on a link. While advanced routing and wavelength assignment algorithms can be used to minimize wavelength conflicts, contention can still arise if there are not enough local access ports.

When sizing the number of local access ports, it is important to consider both of the following types of drop at the node: connection endpoint drops and regeneration drops. In the network study of Section IV, with the 100-Tb/s demand scenario and 40-Gb/s line rate, the average nodal drop percentage due to connection terminations (including intermediate grooming) was 20%; however, the average nodal drop percentage was 32% when regeneration is included. For the 160-Gb/s line rate, the corresponding numbers are 22% and 41%.

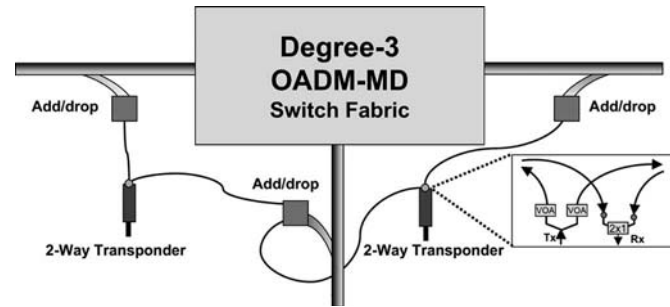


Fig. 9. Any one or any two optical paths can be accessed by a two-way transponder to provide unprotected, 1:1-protected, or 1 + 1-protected connections. The details of one possible implementation are shown in the inset. It is desirable to use an optical backplane to connect the transponders to their associated network ports.

3) *Large Port Count Microelectromechanical System (MEMS) Switches:* A third switch architecture is based on large port count MEMS switches. Both the network ports and the local access ports pass through the switch; however, the ports are single wavelength, i.e., the WDM network fiber is demultiplexed into its constituent wavelengths prior to entering the switch; a local access port is tied to just one transponder. This architecture provides the desired transponder flexibility and avoids the wavelength conflict issue. The drawback is that very large switches would be needed. Consider a degree-4 node with 50% drop, and assume that there are 256 wavelengths per fiber. This would require a  $1536 \times 1536$  switch, which is significantly larger than what is currently available.

In the network study, network elements constituted a very small percentage of the overall cost; thus, the type of switch architecture used was inconsequential.

### B. Smaller Switches Through More Flexible Transponders

The ability of a transponder to access any network port provides greater networking flexibility. Rather than gaining this ability through larger switches, as discussed in the previous section, it is also possible to use more flexible transponders. Assume that an OADM-MD architecture is used (i.e., the add/drop paths are simple taps off of the network ports). Consider an " $N$ -way transponder," where the WDM transmitter output is split into  $N$  paths, with each path feeding into a different add port, as shown in Fig. 9 for  $N$  equal to 2. Variable optical attenuators (VOAs) can be used to control which of the  $N$  feeds actually passes through to a network fiber. (VOAs are often used on the add path for power control anyway.) On the receive side, a switch selects one of the  $N$  corresponding drop feeds. This  $N$ -way transponder is then able to access  $N$  network ports, providing transponder reconfigurability, although if  $N$  is less than the OADM-MD degree, it does not support total flexibility.

This architecture supports multicasting and optical layer protection as well. For multicasting, multiple VOAs pass traffic in the add direction. For 1:1 or shared optical layer protection, the VOAs and switches open and close, depending on which path is active. For dedicated optical layer protection, multiple VOAs are on in the add direction; in the drop direction,

decision circuitry determines the best received signal and sets the receiver switch accordingly.

Clearly, one of the limits on  $N$  is the power loss incurred due to the  $N$ -way splitter. However, if the transponder does not need to access more than two ports simultaneously, then it is possible to architect the transponder to have roughly 3 dB loss even with  $N$  greater than 2; i.e., the transponder can access  $N$  ports, but no more than two at a time, which is sufficient for  $1 + 1$  optical layer protection.

Another limit on  $N$  is due to the number of transponders that can be supported by a single add/drop path, because of splitting loss, number of possible connections in the chassis, etc. A single  $N$ -way transponder occupies  $N$  port connections, even though only one is usually active at a time. For example, assume that a degree-4 OADM-MD is capable of supporting 80 drops from each of the four drop paths. If a conventional single-port transponder is used, 320 simultaneous drops can be supported. However, if two-way transponders are used, only 160 simultaneous drops can be supported. This suggests that only a portion of the transponders should have  $N$ -way capability. Automated reconfigurability is then limited to these transponders.

C. Predeployed Equipment

On-demand and time-of-day services will require that new connections be established on the order of a second or less. Being able to remotely reconfigure the OADMs, OADM-MDs and all-optical switches through software is one important aspect in meeting these requirements. However, it is also critical that the necessary transponders and regenerators already be deployed in the network.

Since the optical reach of the system is typically less than the maximum path length, setting up a new connection may require that there be free transponders or regenerators in the middle of the path for purposes of regeneration. First, consider using two back-to-back transponder cards to regenerate the signal at a particular node. One transponder at this node would need to drop from the incoming link of the path and be connected to a transponder that adds to the outgoing link. If the nodal degree is  $D$ , then there are  $D(D - 1)/2$  possible bidirectional regeneration paths through the node.

There are three options for deploying the back-to-back transponders such that any of the regenerative paths can be established on demand, as illustrated in Fig. 10: (a) Jumper together the short reach interfaces of two transponders and use an all-optical switch; by reconfiguring the switch, the two transponders can connect to any pair of incoming and outgoing links. (b) Use an OADM-MD (i.e., add/drop tap architecture), and pass the individual transponders through an adjunct edge switch that dynamically connects two transponders when regeneration is needed; this assumes that there are free transponders tied to the desired ports. This architecture also allows these transponders to be used for sourcing/terminating traffic, as opposed to just regeneration. (c) Deploy a large number of jumpered transponder pairs to cover each one of the possible regeneration paths through the node. Architecture (a) requires the least number of transponders but has the expense of the more complex all-optical switch, architecture

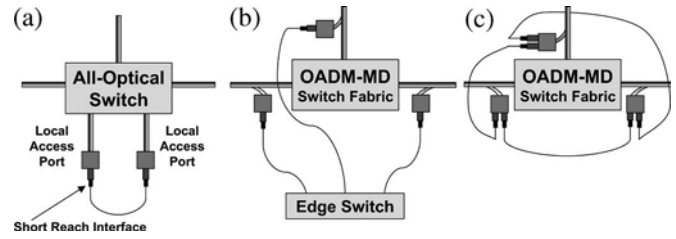


Fig. 10. Regeneration architectures. In architecture (a), an all-optical switch where the add/drop paths pass through the switch fabric is used with a pair of transponders that have their short reach interfaces jumpered together. In architecture (b), the add/drops are taps, and the transponders feed into an edge switch. In architecture (c), again, the add/drops are taps, and there are enough transponder pairs jumpered together to cover all possible regeneration directions.

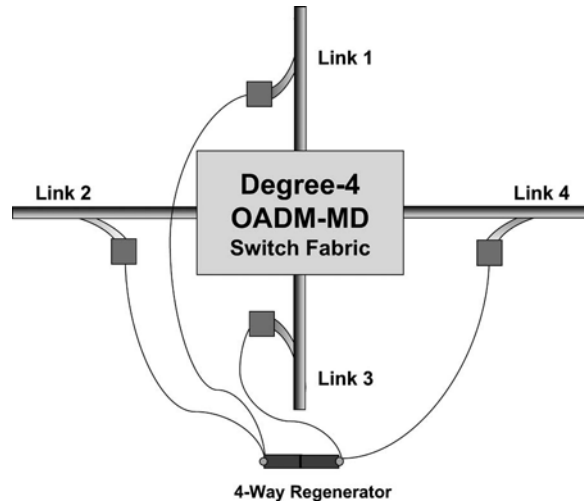


Fig. 11. Four-way regenerator that can provide regeneration between links 1 and 3, 1 and 4, 2 and 3, or 2 and 4.

(b) incurs the cost of an edge switch and slightly more transponders, whereas architecture (c) has the cost of many extra transponders.

Regenerator cards can also be used instead of back-to-back transponders. These cards are functionally the same as two jumpered transponder cards, except the two short reach interfaces are eliminated to reduce the cost. Regenerator cards can be used with options (a) and (c); the advantage is that they have lower costs. (Using regenerator cards with an adjunct edge switch is not efficient; four switch ports are needed per regeneration.)

With the OADM-MD, another possibility is to use  $N$ -way regenerator cards. For example, assume that each side of the regenerator card has a two-way splitter, as shown in Fig. 11. One side of the regenerator is attached to links 1 and 2, the other side is attached to links 3 and 4. This allows regeneration to occur between links 1 and 3, 1 and 4, 2 and 3, or 2 and 4. With one card, four of the six possible regeneration paths are covered. As discussed earlier with the  $N$ -way transponder, this design occupies additional add/drop slots on the OADM-MD; thus, only those regenerator cards that are predeployed for use with dynamic traffic should have this architecture.

The discussion in this section thus far has focused on O-E-O regeneration. As mentioned in Section V-A3, all-optical regeneration is another option. The discussion regarding

O–E–O regenerator cards holds for the all-optical option as well, including the possibility of deploying  $N$ -way all-optical regenerators. The possible advantage of all-optical over O–E–O cards is that they should have relatively lower costs and that they will generate relatively less heat as the line rate increases. Thus, with 40- or 160-Gb/s transport, it is possible to more cost effectively predeploy a large number of such cards to better ensure that dynamic traffic demands can be satisfied.

Regardless of the regeneration option used, having the transponders and regenerators be tunable is essential in a rapidly reconfiguring network. This provides the ability to establish new connections with any wavelength that is free along the new path. Tunability is already supported in most new networks.

For a quantitative study of some of the issues regarding predeployment of equipment, see [72].

## VII. PROTECTION

High availability is very important to some, although typically not all, of the traffic. Protection can be provided by mechanisms at the optical layer, or at higher layers, e.g., IP/MPLS-based restoration. As networks increase in size, scalability of higher layer restoration will become an issue. Optical layer restoration, which operates on the granularity of a wavelength (or a waveband or a fiber), is more scalable, however, it is not able to recover from all failures (e.g., an IP router failure). Thus, it is likely that recovery mechanisms will be implemented in multiple layers [73]. Here, we discuss some important aspects of optical layer protection. There are a variety of optical layer protection/restoration schemes, as described in [74]–[76]. A combination of schemes is likely to be employed in a core network, depending on the availability requirements of the demands.

Backbone networks have traditionally provided protection against single link or node failures through the use of dedicated  $1 + 1$  resources (one work path, one hot standby). Dedicated protection is relatively simple to implement and provides very rapid recovery; the disadvantage is the large amount of network capacity that needs to be dedicated to protection. This has prompted an interest in shared protection, which can be significantly more bandwidth efficient.

Protection capacity requirements will become even more of an issue as the need arises for protection against multiple failures. While connections affected by an initial failure may be rapidly reestablished, the failure itself may not be repaired for several hours. In this repair interval, the network may be subjected to one or more additional failures. For demands with the most stringent availability requirements, it is necessary that the network provide protection against multiple failures. One means to provide this redundancy is to provision dedicated  $1 + 2$  protection (one work path; two active standbys). However, in most network topologies, including the network analyzed in this paper, it is not possible to find three totally disjoint paths for all demands. Furthermore, dedicating three active paths to a given demand is very expensive. Another option is to combine  $1 + 1$  dedicated protection with shared protection to protect against a second failure. Or, the protection mechanism

can be totally shared to provide the most bandwidth-efficient solution.

Shared protection does have its disadvantages: It is more complex to implement, and it tends to be slow because of the amount of communication and/or switching required. Several shared protection schemes have been proposed that potentially can deliver fast recovery (although not as fast as dedicated) while remaining capacity efficient; e.g., hierarchical restoration [77], p-cycles [78], and pre-cross-connected trails [79]. These schemes generally rely on preplanning the protection paths and minimizing the amount of switching that is needed at the time of failure. Note that in the case of protection against multiple failures, greater bandwidth efficiency can be obtained if the scheme dynamically recomputes protection paths after the first failure has occurred. More details on providing protection against dual failures are discussed in [80] and [81].

## VIII. CONTROL PLANE

The optical control plane will play an integral role in achieving the goals of the next-generation backbone network. Intelligent control is needed to enable features such as rapid provisioning and restoration, efficient resource allocation, support for differentiated services with end-to-end quality-of-service, end-user-initiated provisioning, and automated inventory management. With hundreds of wavelengths per link and thousands of connections to manage, scalability of the control plane will be a considerable challenge. The dynamic nature of the network will require that frequent updates of the network state be distributed, which potentially poses problems in terms of control overhead and algorithm convergence. This will likely mandate intelligent data abstraction, where a summarized view of the network is propagated that still allows for efficient resource utilization. It will also require control protocols to be robust with respect to stale information. The presence of optical bypass also adds to the challenge of network control: Connections remaining in the optical domain over several links makes performance monitoring [82] and fault isolation more difficult. In addition, relevant system physical impairments may need to be captured for routing and regeneration purposes. To enable the most efficient use of resources, horizontal integration of control planes, e.g., across metro, regional, and backbone networks, as well as vertical integration, e.g., the interplay of the IP and optical control planes, is needed. Determining how much state information should be shared across the various boundaries is still an area of active debate.

## IX. CONCLUSION

This paper has examined numerous aspects related to the requirements of next-generation core optical networks. The network modeled in this paper was representative of a North American backbone network, but the results are applicable to pan-European backbones as well. Modeling a core optical network with 100-Tb/s aggregate demand, which is an order of magnitude increase over today's networks, highlighted several areas where research is needed. First, the capacity requirements

on a link will be on the order of 16 Tb/s. To reach this goal on a single fiber-pair, the spectral efficiency will need to increase by a factor of 10 over today's networks; this needs to be achieved while maintaining an optical reach of 1500–2000 km. These targets are significantly beyond even current experimental results and will require the development of advanced multilevel modulation formats and detection schemes. The ramifications of this will affect the development of other technologies. For example, all-optical regeneration is a technology that will possibly improve the scalability of future networks; however, most current work in this area is compatible with only relatively simple binary modulation schemes.

From the point of view of cost and optical switch size, the network study showed that 160 Gb/s is the optimal line rate for the 100-Tb/s network, assuming reasonable economics and reach can be achieved. However, 160 Gb/s may prove too challenging at high fiber capacities and long reach, so that 40 Gb/s may be the preferred line rate. While technically easier to achieve, a 40-Gb/s line rate will result in switch fabrics with a very large number of wavelength ports. Waveband switching, possibly combined with an all-optical waveband-grooming functionality, was discussed as a means of moderating the switch size. Another possible solution that may be somewhat easier to implement than 160 Gb/s while still achieving reasonable sized switches is to use the 100-Gb/s Ethernet transport standard being discussed today.

One of the least scalable aspects of today's technology in terms of cost, size, and power consumption is electronic grooming. The network study showed that the bulk of the grooming is needed at the network edge. Thus, a two-tiered hybrid solution, with optical aggregation at the edge (e.g., metro or regional network) and circuit switching in the core, was proposed as a means of achieving greater scalability. The relatively small amount of traffic that requires intermediate grooming within the core network could still be processed with electronic grooming switches; these switches, however, will be appreciably smaller than if all grooming were done electronically.

With the amount of on-demand applications rapidly increasing, it is important that the optical layer incorporate a degree of configurability. With new WSS technology, it is possible to cost effectively design switches where the local access ports can add/drop from any of the network ports, providing greater flexibility at the edge of the network.  $N$ -way transponders were presented as another alternative to achieve configurability at the network edge. Several options for predeploying regeneration equipment in support of on-demand circuits were also discussed. The switch architectures and the  $N$ -way transponder architectures discussed are also suitable for optical multicast applications.

One desirable network attribute that was not discussed is protocol and format transparency, which is a lofty goal that has been bandied about for over a decade. While transparency to various digital signal formats is possible in principle, transparency to analog signals represents a significant challenge. However, this is likely to be desired by only a small number of users, e.g., those involved with special types of sensor networks. Thus, it may be possible to meet this requirement by designating a small number of wavelengths or wavebands

that are capable of carrying an analog signal all-optically end-to-end in the network. By selecting portions of the spectrum with minimal impairments, increasing the wavelength spacing in these spectral regions, and giving preferential signal-to-noise-ratio treatment to these wavelengths within the optical amplifiers and switches, true end-to-end transparency may be attainable. However, more research is needed to determine the feasibility of this approach.

## REFERENCES

- [1] A. A. M. Saleh, "Transparent optical networking in backbone networks," in *Proc. OFC*, Baltimore, MD, Mar. 7–10, 2000, pp. 62–64, Paper ThD7.
- [2] S. Elby, (2006, Mar. 5–10). "Beyond the Internet—Emerging bandwidth drivers," in *Proc. OFC/NFOEC*, Anaheim, CA. [Online]. Available: [http://www.ofcnoec.org/conference\\_program/Market\\_Watch\\_2006.aspx](http://www.ofcnoec.org/conference_program/Market_Watch_2006.aspx)
- [3] S. Melle *et al.*, "Network planning and economic analysis of an innovative new optical transport architecture: The digital optical network," presented at the Optical Fiber Communication Conf. Expo. (OFC)/Nat. Fiber Optic Engineers Conf. (NFOEC), Anaheim, CA, Mar. 6–11, 2005, Paper NTuA1.
- [4] M. Lasserre and V. Kompella, "Virtual private LAN services over MPLS," *IETF*, Feb. 2005. draft-ietf-12vpn-vpls-ldp-06, Work in Progress.
- [5] "User network interface (UNI) 1.0 signaling specification, release 2," *Optical Internetworking Forum*, Feb. 27, 2004.
- [6] G. Swallow *et al.*, "Generalized multiprotocol label switching (GMPLS) user-network interface (UNI): Resource reservation protocol-traffic engineering (RSVP-TE) support for the overlay model," *IETF*, Oct. 2004. draft-ietf-ccamp-gmpls-overlay-05, Work In Progress.
- [7] I. Foster, "The grid: A new infrastructure for 21st century science," *Phys. Today*, vol. 55, no. 2, pp. 42–47, Feb. 2002.
- [8] I. Foster and C. Kesselman, Eds. "The Grid 2: Blueprint for a New Computing Infrastructure," 2nd ed. San Francisco, CA: Morgan Kaufmann, 2004.
- [9] C. Catlett, "The philosophy of TeraGrid: Building an open, extensible, distributed TeraScale facility," in *Proc. IEEE Int. Symp. Cluster Computing and Grid*, Berlin, Germany, May 21–24, 2002, p. 8.
- [10] "About GLIF," *Global Lambda Integrated Facility Web Page*. [Online]. Available: [www.glif.is/about](http://www.glif.is/about)
- [11] J. Mambretti *et al.*, "The photonic TeraStream: Enabling next generation applications through intelligent optical networking at iGRID2002," *Future Gener. Comput. Syst.*, vol. 19, no. 6, pp. 897–908, Aug. 2003.
- [12] S. Figueira *et al.*, "DWDM-RAM: Enabling grid services with dynamic optical networks," in *Proc. IEEE Int. Symp. Cluster Computing and Grid*, Chicago, IL, Apr. 19–22, 2004, pp. 707–714.
- [13] L. Smarr *et al.*, "The OptIPuter, quartzite, and starlight projects: A campus to global-scale testbed for optical technologies enabling Lambda-Grid computing," presented at the Optical Fiber Communication Conf. Expo. (OFC)/Nat. Fiber Optic Engineers Conf. (NFOEC), Anaheim, CA, Mar. 6–11, 2005, Paper OWG7.
- [14] G. Karmous-Edwards, "Today's optical network research infrastructures for E-Science applications," presented at the Optical Fiber Communication Conf. Expo. (OFC)/Nat. Fiber Optic Engineers Conf. (NFOEC), Anaheim, CA, Mar. 5–10, 2006, Paper OWU3.
- [15] A. Cebrowski and J. Garstka, "Network-centric warfare: Its origin and future," *Proc. Naval Inst.*, vol. 124, no. 1, pp. 28–35, Jan. 1998.
- [16] F. Stein, "Observations on the emergence of network centric warfare," in *Proc. Command Control Res. and Technol. Symp.*, Monterey, CA, Jun. 1998, pp. 212–220.
- [17] I. Tomkos *et al.*, "Ultra-long-haul DWDM network with 320 × 320 wavelength-port 'Broadcast & Select' OXCs," in *Proc. ECOC*, Copenhagen, Denmark, Sep. 8–12, 2002, pp. 1–2.
- [18] A. Pratt *et al.*, "40 × 10.7 Gb/s DWDM transmission over a meshed ULH network with dynamically reconfigurable optical crossconnects," presented at the Optical Fiber Communication Conf. (OFC), Atlanta, GA, Mar. 23–28, 2003, Paper PD09.
- [19] J. M. Simmons, "On determining the optimal optical reach for a long-haul network," *IEEE J. Lightw. Technol.*, vol. 23, no. 3, pp. 1039–1048, Mar. 2005.
- [20] J. M. Simmons and A. A. M. Saleh, "The value of optical bypass in reducing router size in gigabit networks," in *Proc. ICC*, Vancouver, BC, Canada, Jun. 6–10, 1999, pp. 591–596.
- [21] P. Hofmann *et al.*, "DWDM long haul network deployment for the Verizon GNI nationwide network," presented at the Optical Fiber Communication

- Conf. Expo. (OFC)/National Fiber Optic Engineers Conf. (NFOEC), Anaheim, CA, Mar. 6–11, 2005, Paper OTuP5.
- [22] B. Zhu *et al.*, "High spectral density long-haul 40-Gb/s transmission using CSRZ-DPSK format," *J. Lightw. Technol.*, vol. 22, no. 1, pp. 208–214, Jan. 2004.
- [23] Z. Xu, K. Rottwitt, and P. Jeppesen, "Evaluation of modulation formats for 160 Gb/s transmission systems using Raman amplifiers," in *Proc. LEOS Annu. Meeting*, Rio Grande, Puerto Rico, Nov. 7–11, 2004, pp. 613–614.
- [24] S. Hardy, "A long road ahead for 40G pioneers," *Lightwave*, vol. 22, no. 3, p. 39, Mar. 2005.
- [25] K. Grobe, L. Friedrich, and V. Lempert, "Assessment of 40 Gb/s techniques for metro/regional applications," presented at the Optical Fiber Communication Conf. Expo. (OFC)/National Fiber Optic Engineers Conf. (NFOEC), Anaheim, CA, Mar. 5–10, 2006, Paper NThA1.
- [26] S. Sengupta *et al.*, "Switched optical backbone for cost-effective scalable core IP networks," *IEEE Commun. Mag.*, vol. 41, no. 6, pp. 60–70, Jun. 2003.
- [27] T. Hausken, "The 40G optical network: why, how, when?," *FibreSystems Europe in Association With Lightwave Europe*, p. 16, vol. 8, no. 11, Nov. 2004.
- [28] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications," *Nature*, vol. 411, no. 6841, pp. 1027–1030, Jun. 28, 2001.
- [29] J. M. Kahn and K.-P. Ho, "Spectral efficiency limits and modulation/detection techniques for DWDM systems," *IEEE J. Sel. Topics Quantum Electron.*, vol. 10, no. 2, pp. 259–272, Mar./Apr. 2004.
- [30] E. Ip and J. Kahn, "Carrier synchronization for 3- and 4-bit-per-symbol optical transmission," *J. Lightw. Technol.*, vol. 23, no. 12, pp. 4110–4124, Dec. 2005.
- [31] M. Suzuki and I. Morita, "High spectral efficiency for large-capacity optical communication systems," in *Proc. Int. Conf. Transparent Optical Netw.*, Wroclaw, Poland, Jul. 4–8, 2004, pp. 171–176.
- [32] P. Cho *et al.*, "Investigation of 2-b/s/Hz 40-Gb/s DWDM transmission over  $4 \times 100$  km SMF-28 fiber using RZ-DQPSK and polarization multiplexing," *Photon. Technol. Lett.*, vol. 16, no. 2, pp. 656–658, Feb. 2004.
- [33] N. Yoshikane and I. Morita, "1.14 b/s/Hz spectrally efficient  $50 \times 85.4$ -Gb/s transmission over 300 km using copolarized RZ-DQPSK signals," *J. Lightw. Technol.*, vol. 23, no. 1, pp. 108–114, Jan. 2005.
- [34] A. Chowdhury *et al.*, "Compensation of intrachannel nonlinearities in 40-Gb/s pseudolinear systems using optical-phase conjugation," *J. Lightw. Technol.*, vol. 23, no. 1, pp. 172–177, Jan. 2005.
- [35] S. Jansen *et al.*, "Optical phase conjugation for ultra long-haul phase-shift-keyed transmission," *J. Lightw. Technol.*, vol. 24, no. 1, pp. 54–64, Jan. 2006.
- [36] P. Bayvel and R. Killey, "Nonlinear optical effects in WDM transmission," in *Optical Fiber Telecommunications IV B*, I. Kaminow and T. Li, Eds. New York: Elsevier, 2002.
- [37] N. Chi *et al.*, "All-optical wavelength conversion and multichannel 2R regeneration based on highly nonlinear dispersion-imbalanced loop mirror," *Photon. Technol. Lett.*, vol. 14, no. 11, pp. 1581–1583, Nov. 2002.
- [38] Y. K. Huang *et al.*, "Simultaneous all-optical 3R regeneration of multiple WDM channels," in *Proc. LEOS Annu. Meeting*, Sydney, Australia, Oct. 23–27, 2005, pp. 135–136.
- [39] O. Leclerc *et al.*, "Optical regeneration at 40 Gb/s and beyond," *J. Lightw. Technol.*, vol. 21, no. 11, pp. 2779–2790, Nov. 2003.
- [40] J. Leuthold, J. Jaques, and S. Cabot, "All-optical wavelength conversion and regeneration," presented at the Optical Fiber Communication Conf. (OFC), Atlanta, GA, Mar. 23–28, 2003, Paper WN1.
- [41] J. M. Simmons, "Analysis of wavelength conversion in all-optical express backbone networks," presented at the Optical Fiber Communication Conf. (OFC), Anaheim, CA, Mar. 17–22, 2002, Paper TuG2.
- [42] K. Croussore, C. Kim, and G. Li, "All-optical regeneration of differential phase-shift keying signals based on phase-sensitive amplification," *Opt. Lett.*, vol. 29, no. 20, pp. 2357–2359, Oct. 15, 2004.
- [43] A. Striegler *et al.*, "NOLM-based RZ-DPSK signal regeneration," *Photon. Technol. Lett.*, vol. 17, no. 3, pp. 639–641, Mar. 2005.
- [44] M. Matsumoto, "Regeneration of RZ-DPSK signals by fiber-based all-optical regenerators," *Photon. Technol. Lett.*, vol. 17, no. 5, pp. 1055–1057, May 2005.
- [45] O. Gerstel *et al.*, "Merits of low-density WDM line systems for long-haul networks," presented at the Optical Fiber Communication Conf. (OFC), Atlanta, GA, Mar. 23–28, 2003, Paper FA3.
- [46] A. A. M. Saleh and J. M. Simmons, "Architectural principles of optical regional and metropolitan access networks," *J. Lightw. Technol.*, vol. 17, no. 12, pp. 2431–2448, Dec. 1999.
- [47] L. Noire *et al.*, "Impact of intermediate traffic grouping on the dimensioning of multi-granularity optical networks," presented at the Optical Fiber Communication Conf. (OFC), Anaheim, CA, Mar. 19–22, 2001, Paper TuG3.
- [48] R. Lingampalli and P. Vengalam, "Effect of wavelength and waveband grooming on all-optical networks with single layer photonic switching," in *Proc. OFC*, Anaheim, CA, Mar. 17–22, 2002, pp. 501–502, Paper ThP4.
- [49] X. Cao *et al.*, "Waveband switching in optical networks," *IEEE Commun. Mag.*, vol. 41, no. 4, pp. 105–112, Apr. 2003.
- [50] P. Bullock *et al.*, "Optimizing wavelength grouping granularity for optical add-drop network architectures," presented at the Optical Fiber Communication Conf. (OFC), Atlanta, GA, Mar. 23–28, 2003, Paper WH2.
- [51] V. Kaman *et al.*, "A cyclic MUX-DMUX photonic cross-connect architecture for transparent waveband optical networks," *Photon. Technol. Lett.*, vol. 16, no. 2, pp. 638–640, Feb. 2004.
- [52] D. Blumenthal, "Optical packet switching," in *Proc. LEOS Annu. Meeting*, Rio Grande, Puerto Rico, Nov. 7–11, 2004, pp. 910–912.
- [53] H. Yang and S. J. B. Yoo, "All-optical variable buffering strategies and switch fabric architectures for future all-optical data routers," *J. Lightw. Technol.*, vol. 23, no. 10, pp. 3321–3330, Oct. 2005.
- [54] D. Blumenthal and M. Masanovic, "LASOR (label switched optical router): Architecture and underlying integration technologies," in *Proc. ECOC*, Glasgow, U.K., Sep. 25–29, 2005, p. 49.
- [55] M. Zirngibl, "IRIS: Optical switching technologies for scalable data networks," presented at the Optical Fiber Communication Conf. Expo. (OFC)/National Fiber Optic Engineers Conf. (NFOEC), Anaheim, CA, Mar. 5–10, 2006, Paper OWP1.
- [56] A. Chowdhury *et al.*, "DWDM reconfigurable optical delay buffer for optical packet switched networks," *IEEE Photon. Technol. Lett.*, vol. 18, no. 10, pp. 1176–1178, May 2006.
- [57] R. Tucker, "Petabit-per-second routers: Optical vs. electronic implementations," presented at the Optical Fiber Communication Conf. Expo. (OFC)/National Fiber Optic Engineers Conf. (NFOEC), Anaheim, CA, Mar. 5–10, 2006, Paper OFJ3.
- [58] C. Qiao and M. Yoo, "Optical burst switching (OBS)—A new paradigm for an optical internet," *J. High Speed Netw.*, vol. 8, no. 1, pp. 69–84, 1999.
- [59] Y. Chen, C. Qiao, and X. Yu, "Optical burst switching: A new area in optical networking research," *IEEE Netw.*, vol. 18, no. 3, pp. 16–23, May/June 2004.
- [60] A. Gumaste and I. Chlamtac, "Mesh implementation of light-trails: A solution to IP centric communication," in *Proc. Int. Conf. Comput. Communications and Netw.*, Dallas, TX, Oct. 20–22, 2003, pp. 178–183.
- [61] I. Widjaja *et al.*, "Light core and intelligent edge for a flexible, thin-layered, and cost-effective optical transport network," *IEEE Commun. Mag.*, vol. 41, no. 5, pp. S30–S36, May 2003.
- [62] H. Nagesh *et al.*, "Load-balanced architecture for dynamic traffic," presented at the Optical Fiber Communication Conf. Expo. (OFC)/National Fiber Optic Engineers Conf. (NFOEC), Anaheim, CA, Mar. 6–11, 2005, Paper OME67.
- [63] P. Winzer, "Robust network design and selective randomized load balancing," in *Proc. ECOC*, Glasgow, U.K., Sep. 25–29, 2005, pp. 23–24.
- [64] S. Alexander *et al.*, "A precompetitive consortium on wide-band all-optical networks," *J. Lightw. Technol.*, vol. 11, no. 5/6, pp. 714–735, May/June 1993.
- [65] V. Chan, S. Chan, and S. Mookherjee, "Optical distribution networks," *Opt. Netw. Mag.*, vol. 3, no. 1, pp. 25–33, Jan./Feb. 2002.
- [66] A. Fumagalli and P. Krishnamoorthy, "A low-latency and bandwidth-efficient distributed optical burst switching architecture for metro ring," in *Proc. ICC*, Anchorage, AL, May 11–15, 2003, pp. 1340–1344.
- [67] Y. Hsueh *et al.*, "Traffic grooming on WDM rings using optical burst transport," *J. Lightw. Technol.*, vol. 24, no. 1, pp. 44–53, Jan. 2006.
- [68] G. I. Papadimitriou *et al.*, "Optical switching: Switch fabrics, techniques, and architectures," *J. Lightw. Technol.*, vol. 21, no. 2, pp. 384–405, Feb. 2003.
- [69] O. Gerstel and H. Raza, "On the synergy between electrical and photonic switching," *IEEE Commun. Mag.*, vol. 41, no. 4, pp. 98–104, Apr. 2003.
- [70] A. Chiu, G. Li, and D. Hwang, "New problems on wavelength assignment in ULH networks," presented at the Optical Fiber Communication Conf. Expo. (OFC)/National Fiber Optic Engineers Conf. (NFOEC), Anaheim, CA, Mar. 5–10, 2006, Paper NThH2.
- [71] S. Mechels *et al.*, "1D MEMS-based wavelength switching subsystem," *IEEE Commun. Mag.*, vol. 41, no. 3, pp. 88–94, Mar. 2003.
- [72] O. Gerstel and H. Raza, "Predeployment of resources in agile photonic networks," *J. Lightw. Technol.*, vol. 22, no. 10, pp. 2236–2244, Oct. 2004.

- [73] M. Pickavet *et al.*, "Recovery in multilayer optical networks," *J. Lightw. Technol.*, vol. 24, no. 1, pp. 122–134, Jan. 2006.
- [74] G. Ellinas *et al.*, "Routing and restoration architectures in mesh optical networks," *Opt. Netw. Mag.*, vol. 4, no. 1, pp. 91–106, Jan./Feb. 2003.
- [75] J. Vasseur, M. Pickavet, and P. Demeester, *Network Recovery: Protection and Restoration of Optical, SONET-SDH, IP, and MPLS*. San Francisco, CA: Morgan Kaufmann, 2004.
- [76] W. Grover, *Mesh-Based Survivable Transport Networks: Options and Strategies for Optical, MPLS, SONET and ATM Networking*. Upper Saddle River, NJ: Prentice-Hall, 2003.
- [77] J. M. Simmons, "Hierarchical restoration in a backbone network," in *Proc. OFC*, San Diego, CA, Feb. 21–26, 1999, pp. 167–169.
- [78] W. Grover and D. Stamatelakis, "Cycle-oriented distributed preconfiguration: Ring-like speed with mesh-like capacity for self-planning network restoration," in *Proc. ICC*, Atlanta, GA, Jun. 7–11, 1998, pp. 537–543.
- [79] T. Chow *et al.*, "Fast optical layer mesh protection using pre-cross-connected trails," *IEEE/ACM Trans. Netw.*, vol. 12, no. 3, pp. 539–548, Jun. 2004.
- [80] M. Clouqueur and W. Grover, "Computational and design studies on the unavailability of mesh-restorable networks," in *Proc. Design Reliable Commun. Netw.*, Munich, Germany, Apr. 2000, pp. 181–186.
- [81] J. Zhang *et al.*, "A comprehensive study on backup provisioning to remedy the effect of multiple-link failures in WDM mesh networks," in *Proc. ICC*, Paris, France, Jun. 20–24, 2004, pp. 1654–1658.
- [82] A. Willner, "The optical network of the future: Can optical performance monitoring enable automated, intelligent and robust systems?," *Optics Photon. News*, vol. 17, no. 3, pp. 30–35, Mar. 2006.