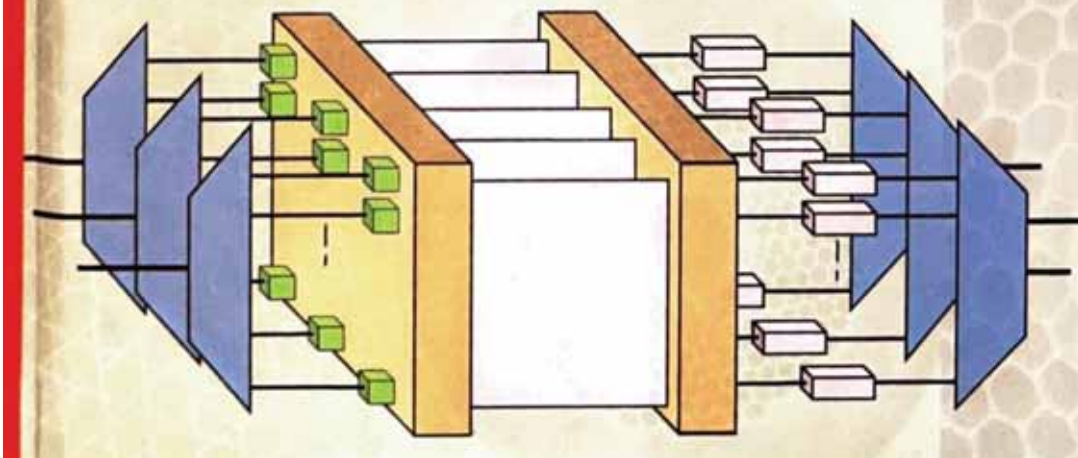
Optical Fiber Telecommunications V

B: Systems and Networks

Ivan Kaminow
Tingye Li
Alan E. Willner



INCLUDES CD-ROM

AP

Exhibit 1017
IPR2023-00581
U.S. Patent 8,886,772

Optical Fiber Telecommunications V B

Optical Fiber Telecommunications V B

Systems and Networks

Edited by

Ivan P. Kaminow
Tingye Li
Alan E. Willner



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
84 Theobald's Road, London WC1X 8RR, UK

This book is printed on acid-free paper.

Copyright © 2008, Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone: (+44) 1865 843830, fax: (+44) 1865 853333, E-mail: permissions@elsevier.com. You may also complete your request on-line via the Elsevier homepage (http://elsevier.com), by selecting "Support & Contact" then "Copyright and Permission" and then "Obtaining Permissions."

Library of Congress Cataloging-in-Publication Data
Application submitted

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ISBN: 978-0-12-374172-1

For information on all Academic Press publications
visit our Web site at www.books.elsevier.com

Printed in the United States of America
08 09 10 11 12 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries
www.elsevier.com | www.bookaid.org | www.sabre.org
ELSEVIER BOOK AID International Sabre Foundation

Contents

Contributors ix

Chapter 1 **Overview of OFT V volumes A & B** 1
Ivan P. Kaminow, Tingye Li, and Alan E. Willner

Chapter 2 **Advanced optical modulation formats** 23
Peter J. Winzer and René-Jean Essiambre

Chapter 3 **Coherent optical communication systems** 95
Kazuro Kikuchi

Chapter 4 **Self-coherent optical transport systems** 131
Xiang Liu, Sethumadhavan Chandrasekhar, and Andreas Leven

Chapter 5 **High-bit-rate ETDM transmission systems** 179
Karsten Schuh and Eugen Lach

Chapter 6 **Ultra-high-speed OTDM transmission technology** 201
Hans-Georg Weber and Reinhold Ludwig

Chapter 7 **Optical performance monitoring** 233
Alan E. Willner, Zhongqi Pan, and Changyuan Yu

Chapter 8 **ROADMs and their system applications** 293
Mark D. Feuer, Daniel C. Kilper, and Sheryl L. Woodward

Chapter 9 **Optical Ethernet: Protocols, management, and 1–100 G technologies** 345
Cedric F. Lam and Winston I. Way

Chapter 10 **Fiber-based broadband access technology and deployment** 401
Richard E. Wagner

Chapter 11 **Global landscape in broadband: Politics, economics, and applications** 437
Richard Mack

Chapter 12 **Metro networks: Services and technologies** 477
Loukas Paraschis, Ori Gerstel, and Michael Y. Frankel

Chapter 13 **Commercial optical networks, overlay networks, and services** 511
Robert Doverspike and Peter Magill

Chapter 14 **Technologies for global telecommunications using undersea cables** 561
Sébastien Bigo

Chapter 15 **Future optical networks** 611
Michael O'Mahony

Chapter 16 **Optical burst and packet switching** 641
S. J. Ben Yoo

Chapter 17 **Optical and electronic technologies for packet switching** 695
Rodney S. Tucker

Chapter 18 **Microwave-over-fiber systems** 739
Alwyn J. Seeds

Chapter 19 **Optical interconnection networks in advanced computing systems** 765
Keren Bergman

Chapter 20 **Simulation tools for devices, systems, and networks** 803
Robert Scarmozzino

Index to Volumes VA and VB 865

9

Optical Ethernet: Protocols, management, and 1–100 G technologies

Cedric F. Lam and Winston I. Way

OpVista, Milpitas, CA, USA

9.1 INTRODUCTION

After years of harsh winter in the telecom industry, which started from the burst of the technology bubble in the beginning of this century, telecom service providers are again hard working with vendors to deploy the next-generation equipment to prepare for the growing bandwidth demands, which are propelled by a slew of new broadband applications such as internet protocol television (IPTV), network gaming, peer-to-peer networking, video/web conferencing, telecommuting, and voice over IP (VOIP).

As the vehicle that interconnects billions of users and devices on the Internet, Ethernet has become the most successful networking technology in history. Even during the years of harsh winter, Ethernet development has never slowed down. The work of 10GBASE-T (IEEE 802.3an, 10-Gigabit Ethernet over twisted pair) was started in November 2002. In the following year (November 2003), IEEE 802.3 launched the project of 10GBASE-LRM (IEEE 802.3aq, 10-Gigabit Ethernet over 300 m of multimode fiber, MMF). The standard for Ethernet in the First Mile (EFM, IEEE 802.3ah) was finished in June 2004. Ethernet continues to evolve rapidly as the human society marches further into the information era.

Originally developed as an unmanaged technology for connecting desktops in local area networks (LANs) [1], nowadays Ethernet has also become a technology for metro and backbone networks. The success of Ethernet is attributed to its simplicity, low cost, standard implementation, and interoperability guarantee [2]. These attributes helped Ethernet and the data networking community it serves to prosper [3], hence producing the economy of scale.

Optical Fiber Telecommunications V B: Systems and Networks
Copyright © 2008, Elsevier Inc. All rights reserved.
ISBN: 978-0-12-374172-1

The Internet, which was initially devised for data connectivity, is now being transformed into a converged platform to deliver voice, data, and video services (the so-called triple-play) through the universal Ethernet interface. Such convergence is made possible by several factors: (1) new mpeg compression technologies which tremendously reduced the bandwidth and storage required for both standard and high-definition broadcast quality videos to reasonable values, (2) advances in electronic memory, storage, and processing technologies, which allows thousands of movies to be stored and switched in practical size video servers, (3) abundance of bandwidth made available by low-cost wavelength-division multiplexing (WDM) and high-speed Ethernet technologies, (4) improvements in the availability and quality of service (QoS) offered by data networks, which made it possible to support always-on and delay-sensitive services such as voice. As an example, with mpeg2 compression [4], a Gigabit Ethernet link is capable of carrying 240 streams of standard-resolution video signals, each of which requires 3.75 Mb/s bandwidth.

Traditional data services offered on internet protocol (IP) and Ethernet networks are best-effort services. Such no-frills approach helps the Internet and Ethernet to penetrate with low initial cost at the beginning [5]. However, as the network grows and the information society becomes more and more network-dependent, best-effort services will no longer be sufficient. This requires the network infrastructure and its underlying technology to evolve in order to satisfy the growing needs as well as increasing levels of service quality expectations.

To cope with this trend, Ethernet itself has gone through many changes and is now taking many forms very different from its initial design. Yet, such changes have been carefully introduced in a controlled manner to allow the existing broad deployment base to grow smoothly. In this chapter, we review some of the evolutions in Ethernet technology development.

9.2 STANDARDS ACTIVITIES

Ethernet is developed within the IEEE 802 LAN/MAN Standard Committee (LMSC) [6]. The LMSC is responsible for developing standards for equipment used in LANs and metropolitan area networks (MANs). Some of the well-known works in LMSC include Ethernet (802.3), wireless LAN (802.11), token ring (802.6), resilient packet ring (802.17), and bridging (802.1). (Figure 9.1).

The IEEE 802.3 Ethernet standard covers the physical layer (PHY) and the medium access control (MAC) function of the data link layer in the OSI (open system interface) seven-layer reference model. Ethernet frame (packet) forwarding and switching is defined in various 802.1 bridging standards.

In addition, there are many other consortiums and standards organizations working on various Ethernet-related issues. For example, MSA (multisource agreement) consortiums such as Gigabit interface convertor (GBIC), small form factor pluggable (SFP), XENPAK, and XFP [7–10] have been formed by component manufacturers to specify transceiver modules (the so-called PMD or physical

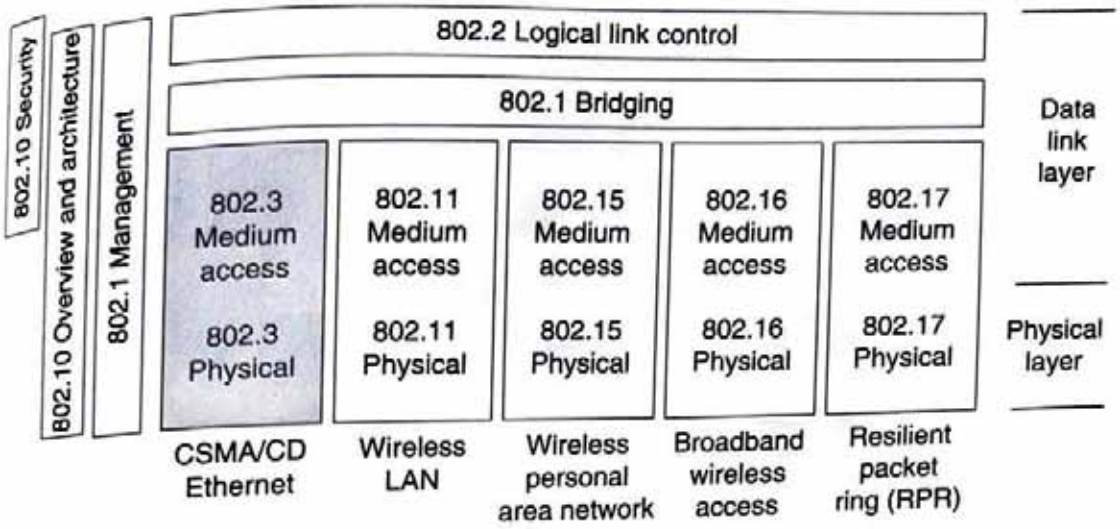


Figure 9.1 The IEEE 802 LMSC organization overview (this figure may be seen in color on the included CD-ROM).

medium dependent in Ethernet terminology) with common form factors and common electrical interfaces, which can be used interchangeably with different systems. Metro Ethernet Forum (MEF) [11], another industry consortium, is defining Ethernet service types, operation, administration, and maintenance (OAM) functions, and service level agreements (SLAs).

Within the international telecommunication union (ITU), standards have been published on carrying Ethernet over time-division multiplexing (TDM) circuits. These include the generic framing procedure (GFP) defined in ITU-T G.7041 [12], virtual concatenation (VCAT) defined in ITU-T G.7043 [13], and link capacity adjustment scheme (LCAS) defined in ITU-T G.7042 [14]. ITU-T G.8031 [15] is concerned with Ethernet protection switching. ITU-T Y.1731 [16] deals with OAM functions and mechanisms for Ethernet-based networks.

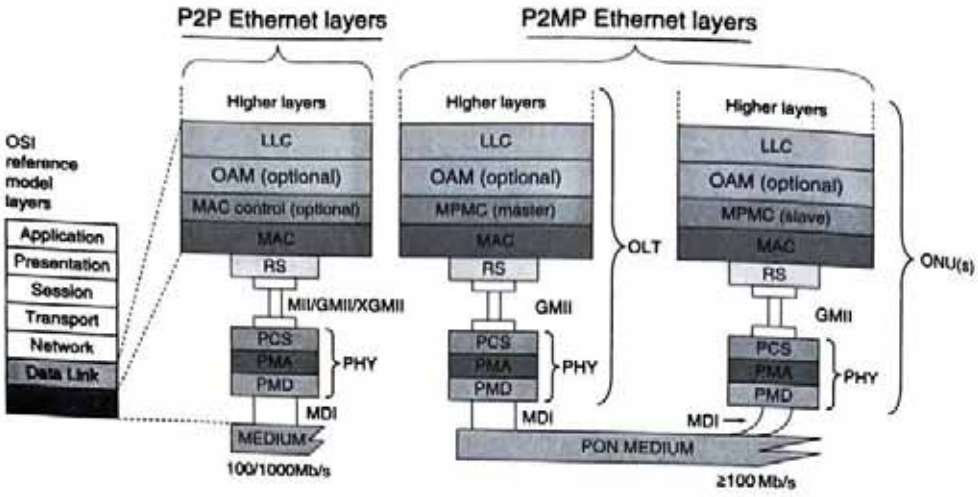
Optical Internet Forum (OIF) has defined user network interface (UNI) for signaling Ethernet connections in a generalized multiprotocol label switching (GMPLS) enabled optical networks [17].

The overwhelming standard work around Ethernet implies that it is impossible to cover everything in this chapter. Therefore, our goal is to offer a direction to those interested readers to explore in-depth the rest of this rich subject.

9.3 POINT-TO-POINT ETHERNET DEVELOPMENT

9.3.1 Modern Ethernet Layering Architecture

Figure 9.2 shows the layering architecture of modern Ethernet as defined in the IEEE 802.3 standard [18]. In this figure, the MAC layer and the PHY layer are connected with an media-independent interface (MII) for 100 Mb/s Ethernet, GMII



MAC: media access control
MDI: medium dependent interface
MII: media independent interface
XGMII: 10 Gigabit media-independent interface
PHY: physical layer device
RS: reconciliation sublayer
MPMC: multipoint media access control
PCS: physical coding sublayer
GMII: Gigabit media-independent interface
PMA: physical medium attachment
PMD: physical medium dependent

Figure 9.2 Modern Ethernet layering architecture (this figure may be seen in color on the included CD-ROM).

(Gigabit media-independent interface) for Gigabit Ethernet and XGMII (10 G MII) for 10 Gb/s Ethernet.

This idea of separating the MAC layer from the physical layer started from the very beginning of the Ethernet history to allow the reuse of the same MAC design with different physical layer technologies and transmission media for Ethernet. Within the PHY layer, the physical coding sublayer (PCS) generates the line coding suitable for the channel characteristics of the transmission medium. The physical medium attachment (PMA) layer performs transmission, reception, collision detection, clock recovery, and skew alignment functions within the physical layer. The physical medium dependent (PMD) layer defines the optoelectronic characteristics of the actual physical transceiver. The term MDI (medium-dependent interface) is simply a fancy way to describe a connector. More detailed discussions of Ethernet layering functions can be found in Ref. [18].

9.3.2 Physical Layer Development

All the modern Ethernet systems are formed with full-duplex links, which do not have the speed and distance limitations imposed by the original CSMA/CD (carrier sense multiple access with collision detection) protocol [19]. Full-duplex Ethernets adopt a star-shaped hub-and-spoke architecture with point-to-point (P2P) connections between the hosts and a hub bridge. The distances between the bridge and hosts are only limited by physical transmission impairments. As mentioned before,

Ethernet embraces different physical layer technologies with a standard interface between the MAC layer and the physical layer. The MAC layer for P2P Ethernet has not changed much for a considerable period of time. Most of the developments in Ethernet happened in the physical layer in the last 10 years.

Gigabit Ethernet Physical Layer

10/100 Mbps Ethernets are mostly deployed on copper medium (coaxial cable or unshielded twisted pair, i.e., UTP). Gigabit Ethernet was first standardized on optical fiber in 1998. Two designs were ratified in IEEE 802.3z to transmit Gigabit Ethernet signals: the 1000BASE-SX uses short-wavelength lasers (850 nm) on MMFs, and the 1000BASE-LX uses long-wavelength laser (1310 nm) on the standard single-mode fiber. At that time, transmitting 1000 Mbps signals on the widely deployed Category 5 UTP was a significant challenge for silicon-chip designers. It requires tremendous signal processing to mitigate the channel impairments in copper wires such as ISI (intersymbol interference) introduced by limited channel bandwidth and signal crosstalks between pairs of copper wires.[1] It was not until a year later that the 1000BASE-T standard (IEEE 802.3ab) was finished.

Although Gigabit Ethernet is now mainly deployed with UTP interfaces, early Gigabit Ethernet was mostly deployed with optical interfaces. Fiber has the advantage of little signal impairments and wide bandwidth. It is suitable for backbone transmission which is the major application for early Gigabit Ethernet. To keep the cost of Gigabit Ethernet low, the IEEE 802.3z committee very conservatively defined the transmission distance limit of 1000BASE-SX as 300m, and that of 1000BASE-LX as 5 km.

Both 1000BASE-SX and 1000BASE-LX share the 8B10B 1000BASE-X PCS line coding [18, Clause 36]. Besides the transmission media, the only difference between 1000BASE-SX and 1000BASE-LX lies in the PMD layer which defines the laser transmitter and photodetector. The interface between the PMA and PMD layer is simply a serial interface. This made it easy to reuse all the designs between 1000BASE-SX and 1000BASE-LX except the PMD transceiver, which cannot interoperate with each other.

Although the IEEE 802.3z standard committee has made the PMD specification extremely conservative, it still represented a significant portion of the Gigabit Ethernet cost.[2] The cost of optical transceivers would explode in Gigabit Ethernet switches and routers containing high port counts. Luckily, the well-thought layered design of Ethernet allows the optical transceiver modules to be separated from the rest of system.

The IEEE 802.3z standard did not specify an exposed interface between the PMA and PMD. Nevertheless, transceiver manufactures formed MSA consortiums [20] that defined optical transceiver modules (i.e., PMDs) with a common electrical

[1] 1000 BASE-T uses four pairs of unshielded Category 5 cables simultaneously for signal transmission and reception.

[2] The cost of optical transceivers dominated the cost of Gigabit Ethernet. It is also well-known that the cost of silicon is always difficult to compete with.

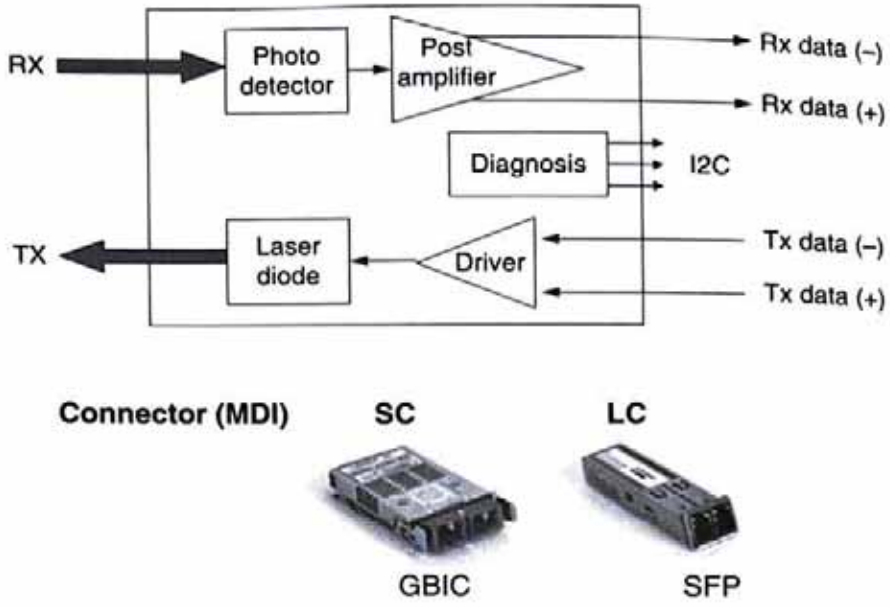


Figure 9.3 GBIC and SFP MSA modules: block diagram (top) and picture (bottom) (this figure may be seen in color on the included CD-ROM).

interface and uniform mechanical dimensions. The most commonly seen Gigabit Ethernet MSA PMD modules are GBIC [7] and SFP [8] (Figure 9.3). SFP modules are much smaller in size and became the most popular Gigabit PMD. To improve system density, SFPs use the compact-form LC connector not specified in the IEEE 802.3 standard. Both GBIC and SFP modules are hot swappable so that a router/switch does not need to be populated with expensive optical modules when they are manufactured. Instead, optical transceivers can be inserted when a port needs to be connected. In addition, one does not need to decide ahead of the time which type of optical PMD to be populated at the time of purchasing a piece of Ethernet equipment.

As shown in Figure 9.3, the GBIC and SFP MSA modules contain no data-rate[3] and protocol-specific processing blocks. Therefore, such modules can also be used for other applications such as Fiber Channel and Synchronous Optical NETwork/Synchronous Digital Hierarchy (SONET/SDH). Therefore, the MSA concept not only created a pay-as-you-grow upgrade scenario, but also the economy of scale for optical transceivers which helps to reduce their costs through mass production.

Besides the basic necessary optical–electrical (OE) and electrical–optical (EO) conversion functions, MSA modules also offer a digital diagnostic I2C (Inter-IC bus) interface, which provides information such as PMD type, laser wavelength, input, and output optical power to the host system. This interface can be used for optical link trouble shooting and performance monitoring.

Another advantage offered by MSA is the ease to incorporate new improved PMD capabilities when they are available. As mentioned before, the IEEE 802.3z committee selected an extremely conservative optical reach of 5 km for the

[3] Clock and data recovery is performed in the PMA layer.

1000BASE-LX PMD. Fueled by the explosion of WDM transmission systems, optical transceiver capabilities have been advancing rapidly. For example, an industry accepted 1000BASE-ZX (not in the IEEE standard) MSA specification exists which enables link transmission distances to be extended to 70 km using APD (avalanche photodiode) receivers. SFP modules with dense wavelength-division multiplexing (DWDM) lasers also exist, which allow users to easily improve fiber utilizations by using parallel wavelengths to multiply the link capacity. All these improvements only involve changes confined to MSA modules.

10 Gigabit Ethernet Physical Layer

10 Gigabit Ethernet Layering Architecture The technology of 10 Gigabit Ethernet was significantly more challenging than that of Gigabit Ethernet. When it was being standardized, 10 Gb/s transmission was still the state-of-the art technology. Many different schemes had been proposed to realize 10 Gigabit Ethernet. As expected, 10 Gigabit Ethernet was also first standardized on the optical fiber medium. At 10 Gb/s data rate, the 8B10B PCS code with 25% overhead (used in 1000BASE-X standard for optical fiber media) would lead to a physical symbol rate of 12.5 Gbaud/s, which is much higher than the conventional OC192/STM64 transmission rate. In order to contain the symbol rate and minimize the cost and technical challenge for 10 Gigabit Ethernet transceivers, 10 Gigabit Ethernet uses a new PCS code (64B66B) with only 3% of the coding overhead. Figure 9.4 shows the summary of 10 Gigabit Ethernet architectures. Two 10GBASE standards using the 64B66B PCS coding [18, Clause 49] were initially produced: the 10GBASE-R LAN standard carrying native Ethernet frames in the physical layer, and the 10GBASE-W WAN standard using SONET/SDH compliant frames in the physical layer.

The 10GBASE-W PHY contains a WAN interface sublayer (WIS) (Figure 9.4), which encapsulates Ethernet MAC frames within a SONET/SDH compliant frame [18, Clause 50]. The WIS layer also performs rate adaptation function by stretching the gaps between adjacent Ethernet frames so that the output data rate generated by the WAN interface matches the SONET/SDH OC-192 data rate of 9.953Gb/s.[4]

The 10GBASE-W PHY was created because most of the 10 Gb/s transport system existed in SONET/SDH forms at that time. At the time, 10 Gb/s Ethernet was envisioned as an aggregation technology for backbone applications. So it seemed logical to create a WAN standard which was compatible with the existing deployment base of 10 Gb/s transport systems. Nevertheless, the data communication world never liked the WAN standard and most of the 10 Gigabit Ethernet equipment deployed today uses the 10BASE-R standard.

Parallel to 10GBASE-R and 10GBASE-W, a 10GBASE-X standard was created. Similar to 1000BASE-X, the 10GBASE-X standard uses the 8B10B encoding scheme. Instead of transmitting on a single serial interface, the 10GBASE-X PHY transmits signals on a four-lane parallel interface, using four coarsely spaced wavelengths (4×2.5 Gbps) around the 1300 nm spectral region to form the so-called 10GBASE-LX4. It was the first time that the WDM technology was used in Ethernet standard. Even though the LX-4 interface has better dispersion tolerance and was easier to design than 10 Gb/s serial interfaces from a transmission viewpoint, it requires four sets of lasers and photoreceivers, which increase the packaging size, complexity, and cost. Within only a few years, 10 Gb/s serial PHYs have advanced so rapidly that they rendered the LX4 interface obsolete. Three types of 10 Gb/s serial optical PHY standards were initially created: 10GBASE-S, 10GBASE-L, and 10GBASE-E, which are summarized in Table 9.1. The 10GBASE-E interface uses the minimum loss wavelength region of 1550 nm minimum in the silica fiber (the first time in 802.3 standard) to support transmission distances up to 40 km.

The 10GBASE-LRM standard was not finished until 2006, 4 years after the first 10 Gigabit Ethernet Standard IEEE 802.3ae was finished. It enables the use of low-cost Fabry–Perot (FP) lasers to transmit up to 220 m on legacy MMF which has

Table 9.1
Summary of 10GBASE optical standards.

PHY standard	Wavelength (nm)	Serial/parallel	Link distance	Medium
10GBASE-SR/W	850	Serial	300/33 m	50 μm/62.5 μm MMF
10GBASE-LRM	1310	Serial	220 m	50 μm/62.5 μm MMF
10GBASE-LX4	1310	WDM (parallel)	300 m	50 μm/62.5 μm MMF
10GBASE-LR/W	1310	Serial	10 km	Single-mode fiber
10GBASE-ER/W	1550	Serial	10 km	Single-mode fiber
			40 km	Single-mode fiber

[4] 10GBASE Ethernet has a MAC throughout of 10 Gb/s.

been widely deployed in the early 1990s for FDDI and Fast Ethernet applications. To realize 10GBASE-LRM requires advanced electronic dispersion compensation (EDC) techniques in the receiver [21]. Chapter 18 (Volume A) by Yu, Shanbhag, and Choma discusses electronic dispersion compensation (EDC) techniques in detail.

10GBASE-T Interface The 1000BASE-T UTP standard was ratified a year after the standardization of 1000BASE-X. It quickly became the dominating Gigabit Ethernet interface. UTP interfaces have proven to be popular for interconnecting servers, switches, and routers because of the ease in their cable termination and handling. However, it was not until 4 years after the standardization of 10GBASE optical Ethernet that the 10GBASE-T interface standard had been finished [22].

The 10GBASE-T interface uses an low-density parity check (LDPC) PCS. It employees a two-dimensional 16-level pulse amplitude modulation (PAM) encoding scheme on copper wire. The traditional ubiquitous Category 5 cables are no longer capable of supporting 10GBASE-T. 10GBASE-T allows transmission distances of up to 55 m on Category 6 cables. To reach the 100 m distance achieved by 10/100/1000BASE-T interfaces, 10GBASE-T requires a new Augmented Category 6 (or CAT-6A) cable, which has the frequency responses, crosstalk, and alien crosstalk[5] characteristics specified up to 500 MHz [23].

It can be expected that for a considerable period time, optical PHYs will still dominate in 10GBASE Ethernets.

The XAUI Interface 10GBASE PHY and 10GBASE MACs are interconnected with the XGMII. The XGMII interface uses a 32-bit wide data bus with a limited distance support of 7cm.

To facilitate module interconnect, an XGXS (10 Gigabit extender) interface was defined to extend the reaches of XGII. The XGXS interface reduces the 32-bit XGMII data path into a 4-bit 8B10B encoded XAUI (10 Gigabit attachment unit interface) interface as shown in Figure 9.5 [18, Clause 47]. The XAUI interface uses the exactly same coding scheme used in 10GBASE-LX4 standard. It also has a longer reach of 25 cm to facilitate the connection between a PHY device and the MAC layer. Even though the 10GBASE-LX4 PHY using the same coding scheme has never been popular, the XAUI interface has been used in many 10 Gb/s MSA modules.

10Gb/s MSA Modules 10 Gb/s MSA modules are divided into two major categories, MSA transceivers and MSA transponders, which are shown in Figure 9.6. The main difference is that transceivers interface with the host system using a serial interface whereas transponders using a parallel interface. Therefore an electrical MUX/DMUX (multiplexer/demultiplexer) (also called SERDES—serializer/deserializer) is included in a transponder.

[5] Alien crosstalk refers to the crosstalk between neighboring UTP cables in a bundle.

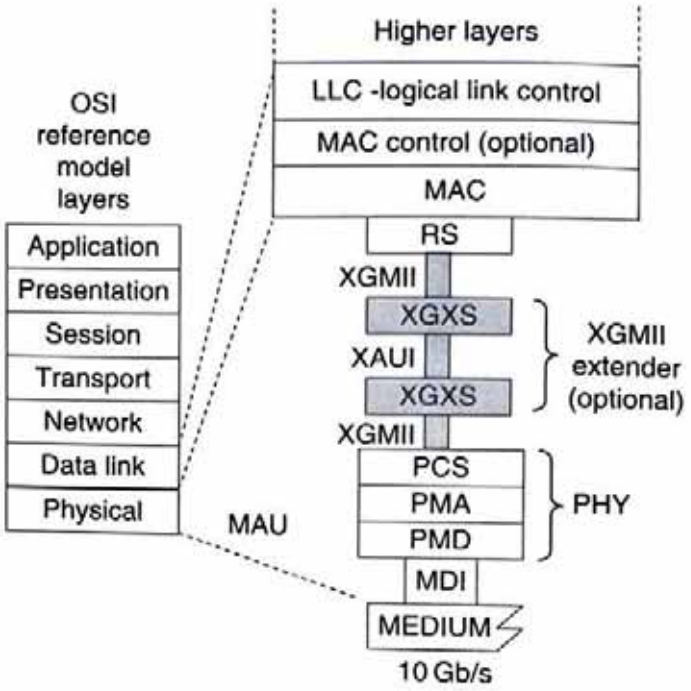


Figure 9.5 The XAUI interface (this figure may be seen in color on the included CD-ROM).

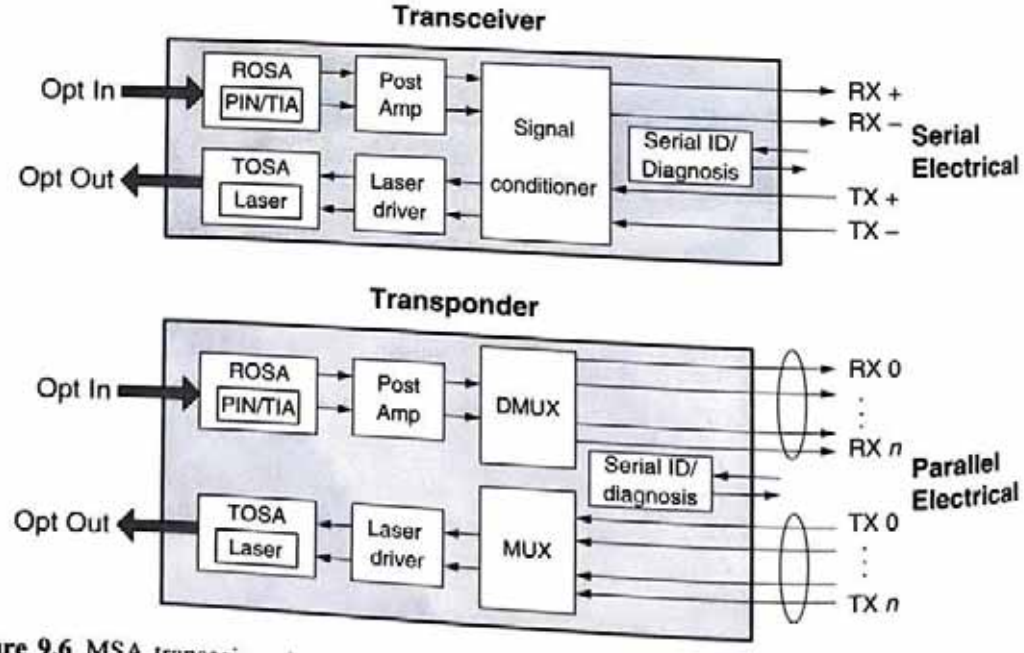


Figure 9.6 MSA transceiver (top) vs transponder (bottom) (this figure may be seen in color on the included CD-ROM).

Figure 9.7 shows the diagrams of three types of commonly seen 10 Gb/s MSA modules. The XENPAK and XFP modules are hot swappable modules while the 300-pin module is not. Both XENPAK and 300-pin MSA are transponders while XFP belongs to the transceiver family. Standard 300-pin module implements the

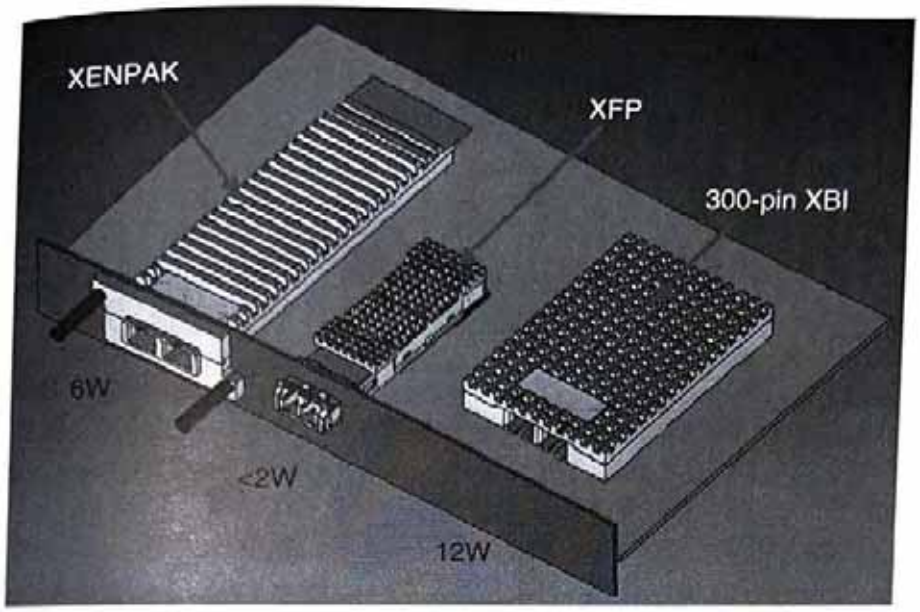


Figure 9.7 Commonly seen 10 Gb/s MSA modules. (this figure may be seen in color on the included CD-ROM)

16-bit wide OIF SFI-4 (SERDES framer interface, Release 4) electrical interface for SONET/10G-WAN/10G-LAN signals [76].

Transponders produce lower-speed parallel signals, which are easier to handle on electrical printed circuit boards (PCBs). In contrast, they also require bigger packages and complicated processing circuits. Moreover, transponders are also often format and bit-rate dependent, which limit them to a single application.

Despite the challenge in handling serial 10 Gb/s signals at the electrical interface, MSA transceiver modules are more compact and consume less power. Figure 9.8 compares the block diagrams and applications of 10 Gigabit Ethernet MSA transponders and transceivers. XENPAK and XFP are the most popular 10 Gb/s MSA transponder and transceiver, respectively. Besides maintaining the signal integrity, heat dissipation is a challenge for 10 Gigabit MSA modules, which limits the compactness of their sizes. XPAK and X2 are essentially more compact versions of XENPAK. Significant progresses have been made in the recent years to reduce MSA module power consumptions. A new MSA transceiver standard called SFP+ with form factor compatible with SFP is being standardized at the time of writing [4]. It provides even higher density and lower power than XFP transceivers.

Figure 9.8 illustrates that all three 10 Gigabit Ethernet transponders (XENPAK, XPAK, and X2) share the same design with embedded PCS and PMA sublayers and a XAUI interface to the host system. This allows the host system to use any type of the PHY device irrespective of the PCS line coding scheme (i.e., whether 10GBASE-R, 10GBASE-W, or 10GBASE-X PHY is required). For Layer-2/3 switch and router manufactures, this has the advantage of allowing their switching equipment to interface with any PHY devices. Nonetheless, as silicon design advances and the Ethernet community converges to the LAN interface, this

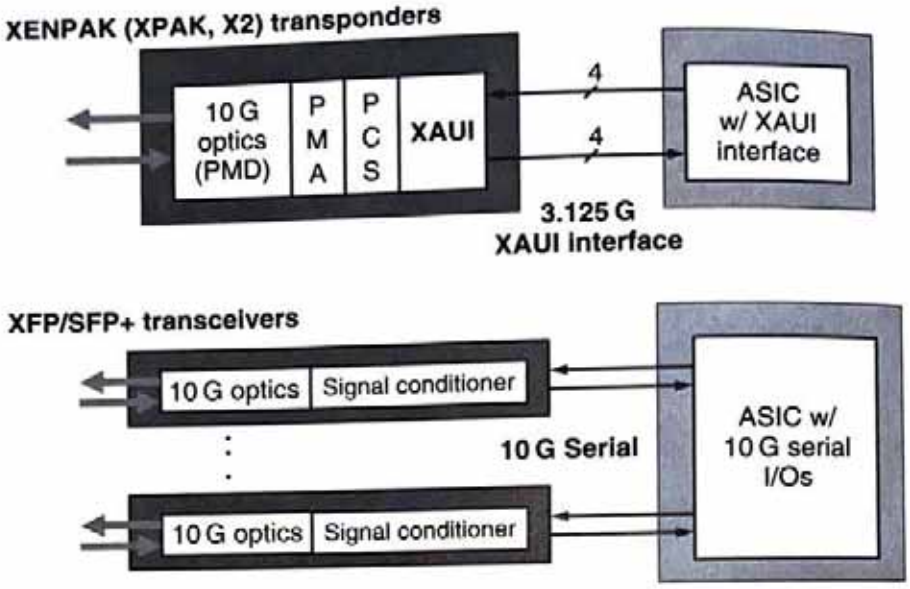


Figure 9.8 10 Gigabit Ethernet transponder (top) vs transceiver (bottom) (this figure may be seen in color on the included CD-ROM).

flexibility advantage gives way to the high port-count density and integration benefit offered by transceiver modules. There is a growing industry trend to converge to XFP- and SFP+-based systems. Furthermore, for operational and management efficiencies, the industry prefers only a small handful number of 10GBASE PHY interface types than having many different flavors.

Like their Gigabit counterparts, 10 Gigabit MSA transceivers can be designed to operate at multiple data rates so that they can be used with other 10Gb/s transport systems such as SONET OC-192 and ITU-T OTU-2. Unlike GBIC and SFP transceivers, which normally only have a simple laser driver and postamplifier, to maintain high-speed signal quality and integrity, 10 Gb/s MSA transceivers are normally built with a signal conditioner which performs regeneration to clean up the distortions introduced by the electrical reshape, retime, and reamplify interface between the module and the host system. The signal conditioner can represent (3-R) clock data recovery (CDR) units in transmit and receive paths, or even electronic dispersion compensators. To improve integration and further reduce power consumptions, most of the SFP+ modules will not have built-in CDR to achieve less than 1 W power consumption. A transport equipment manufacturer would usually prefer transceiver-based MSA modules because (1) they can design transponders to work with different format signals and (2) they may not want to deal with the management and configuration complexity associated with the XAUI interface.

Nevertheless, 10 Gb/s transponders still represent the state-of-the-art commercial technology. New 10 Gb/s transmission techniques with higher performance continue to emerge. Transponder manufacturers are taking the advantages of the extra spaces available in 300-pin and XENPAK modules to embed new

transmission capabilities. For example, EDC [21], tunable laser [25], and duobinary [26] modulated transmitters have been incorporated in commercial 300-pin modules. These improved capabilities simplify the job of transport system integrators.

Link Diagnosis in 10 Gb/s Ethernet Traditionally, for cost and simplicity, Ethernet does not include much diagnosis capabilities besides CRC frame integrity check and PHY layer link-up/link-down verification. This was adequate when Ethernet was mainly used in LAN environments. 10-Gigabit Ethernet was intended for MAN applications. To improve network troubleshooting capabilities, for the first time, the IEEE 802.3 standard group introduced loopback and remote link fault diagnosis functions into 10-Gigabit Ethernet designs. These capabilities are shown in Figure 9.9.

The 10-Gigabit Ethernet standard includes optional loopback functions at various PHY sublayers as indicated in Figure 9.9. These loopback functions can be implemented in MSA modules and invoked through the digital diagnosis interfaces so that when a port is not functioning properly, the problem can be isolated and localized with various loopback tests.

Another capability introduced in 10-Gigabit Ethernet is the local fault (LF) and remote fault (RF) signals, which are conceptually similar to the loss of signal (LOS) and remote fault inductor (RDI) maintenance signals on a SONET link. When a link error is detected, if the local receiver receives a corrupted signal, it will generate the LF code words (called LF ordered set, or LFOS) to the reconciliation sublayer (RS) layer [18, Clause 46]. At the same time, the local RS layer inserts RF ordered set (RFOS) to the transmitter which will be received by the link partner. The LF/RF

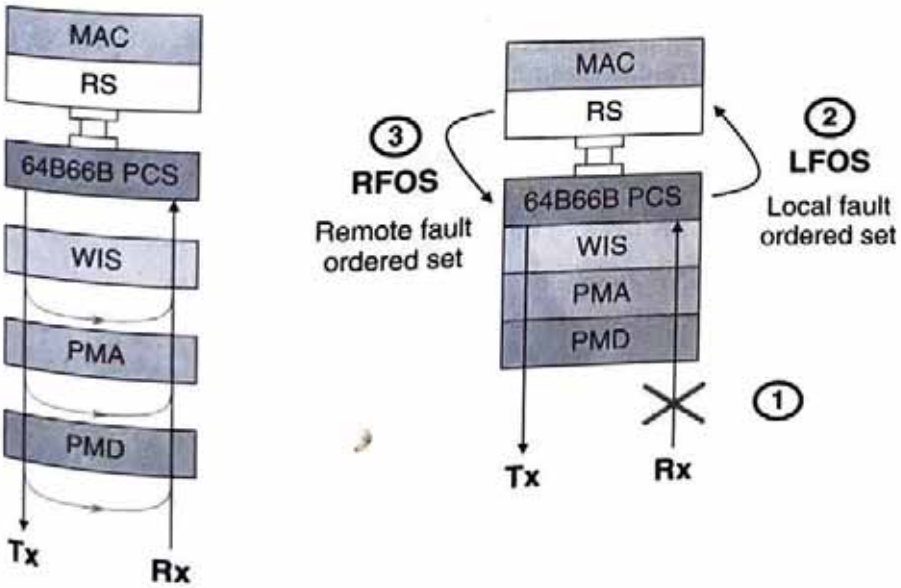


Figure 9.9 Loop-back modes (left) and link-fault signaling (right) in 10 Gigabit Ethernet (this figure may be seen in color on the included CD-ROM).

signals are represented using special 64B66B code words. Thus they are terminated in the physical layer and not passed up to the upper layers.

9.4 LAYER-2 FUNCTIONS IN ETHERNETS

Layer-2 functions include MAC and Ethernet frame switching, which is also called bridging. Unlike traditional circuit switched networks, Ethernet is a packet switched technology. Every Ethernet frame is labeled with a source address (SA) and a destination address (DA) which are used by Ethernet bridges to forward the frame to the proper destination. The IEEE 802.3 standard only covers the MAC portion. Ethernet bridging is covered by the IEEE 802.1 standards. The most important idea for Ethernet bridging is the IEEE 802.1D Spanning Tree Protocol (STP) [27].

9.4.1 Ethernet MAC Frames

To discuss the bridging operation, one needs to first understand the format of Ethernet frames. Figure 9.10 shows the basic Ethernet frame format. This basic format has remained invariant for a considerable period of time, despite the rapid development in Ethernet speed and different physical layer technologies.

Ethernet is a multimedia technology because it operates on different media with various speeds. Ethernet devices are designed with clearly defined interfaces between the MAC layer and the PHY layer. This layered approach allows the physical layer to evolve independent of the MAC layer. Ethernet frames represent the data format at the MAC layer. It is the common MAC layer specification and MAC frame formats that allows Ethernet devices of different speeds and PHY technologies to interoperate with one another. In fact, switches are often built with ports of different speeds and medium types.

Ethernet frames are variable length with a payload area between 46 and 1500 octets. An invariant MAC frame format allows each generation Ethernet to be backward compatible with early generations so that users do not need to upgrade

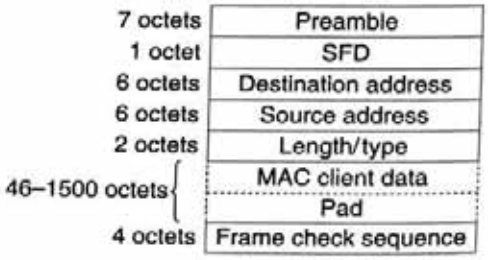


Figure 9.10 Basic Ethernet frame format (SFD: start frame delimiter).

upper layer software and applications when the network speed is increased. This played an important role to ensure the commercial successes of Ethernet. An Ethernet frame starts with a preamble field with alternating 0's and 1's, which is used in early days by burst-mode receivers at a destination node to recover the signal clock. When Ethernet connections became P2P, transmitter and receiver synchronization is always maintained by transmitting idle symbols when there is no data to send. This obviates the need of the preamble field, which nonetheless, is kept for backward compatibility.

The preamble is followed by an start frame delimiter (SFD) to signify the beginning of a frame. Following the SFD is the DA and SA. The first bit of DA determines if the frame is a unicast or broadcast frame. A unicast frame is represented by a 0 value in the first bit of the DA and a multicast packet by a 1 value. The all 1 address is reserved as the universal broadcast address. Normally, a bridge receiving a frame with a broadcast/multicast address will forward the frame to all other ports except the incoming port. A block of multicast addresses has also been reserved by IEEE for protocol implementations. Packets with these reserved addresses are interpreted as protocol data units (PDUs) with special meanings. A station receiving these special multicast packets will normally terminate such packets without forwarding them. In other words, if a bridge receives a multicast frame with its DA in the reserved address block and the switch does not understand the frame, it will simply drop the frame.

Ethernet frames also have a two-octet length/type field to represent the length of the payload field. Since the allowed maximum payload frame is only 1500 bytes, a length/type value above 1536 represents the type of the Ethernet frames. It is often used to represent the upper layer protocol or the type of management information contained in the payload. The frame check sequence (FCS) field uses a four-octet cyclic redundancy check value (CRC) to protect the frame.

Figure 9.10 shows that Ethernet frames contain minimum management information. Such simple frame structure helped to keep the network equipment simple and low-cost. However, as network infrastructures continue to grow and Ethernet service management becomes more and more important, the original frame format with minimal overhead designs is no longer sufficient. Expansions in Ethernet frames have been carefully introduced in the recent years to allow the growths of Ethernets while minimizing the impacts on legacy Ethernet devices [28].

9.4.2 Transparent Bridging

A CSMA/CD collision domain is a multipoint-to-multipoint mesh-connected broadcast domain. Any station in a broadcast domain can directly communicate with any other station in the same domain by broadcasting the frame in the domain. In fact, a station just sends a frame to its destination assuming that the frame will eventually arrive at the destination. As explained before, stations in a broadcast domain share the common medium and its capacity. As the number of hosts in a domain grows and network size increases, the network performance will be degrade.

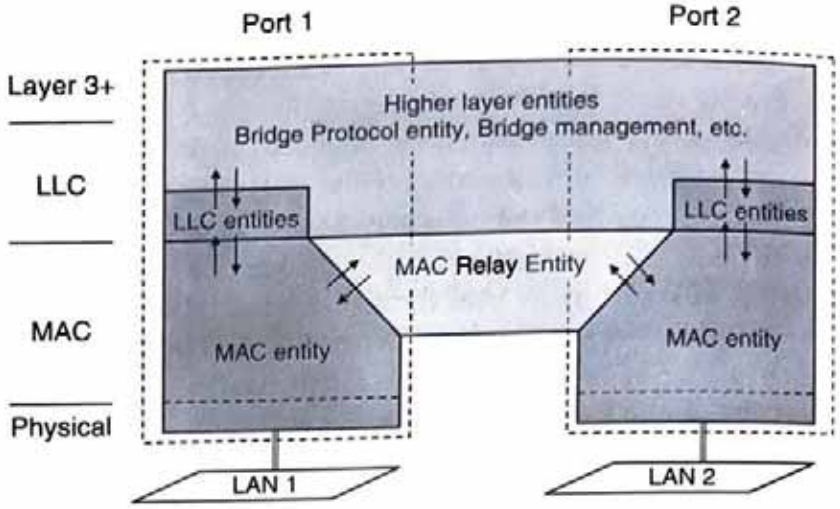


Figure 9.11 Architecture layering of a bridge (this figure may be seen in color on the included CD-ROM).

A bridge improves the network performance by limiting the size of a collision domain. The term switch and bridge are used interchangeably in the networking industry. A bridge is a multiport device with a layering architecture shown in Figure 9.11. Each bridge port is connected to a separate LAN (i.e., separate collision domain). A bridge contains a MAC relay entity to forward MAC frames from one port to another.

Normally, besides broadcast/multicast frames, an Ethernet port only accepts unicast frames with DA matching its own MAC address. A bridge port, in contrast, works in a promiscuous mode. It receives frames with any destination addresses and performs one of the three functions:

(1) Broadcast (flooding)
(2) Forwarding
(3) Filtering

Figure 9.12 shows the functional diagram of a bridge, which contains a source address table (SAT), a filter/forward lookup logic and a learning logic associated with port interfaces. When an Ethernet frame arrives at a port interface, the filter/forward lookup logic makes use of the DA and SAT to decide whether the frame needs to be broadcast, forwarded, or filtered.

The SAT is populated automatically through the learning logic or manually through management provision. Each entry of the SAT contains the association of a host address and the bridge port that the addressed host can be reached from. In automatic learning, when a new frame arrives at a bridge port, the learning logic examines its SA. If that address is not yet in the SAT, the learning logic will populate the SA and the port number in the SAT, so that next time, when a frame with the DA matching that address arrives at the bridge, the bridges knows how to forward the frame. Notice that the bridge operation assumes bidirectional links.

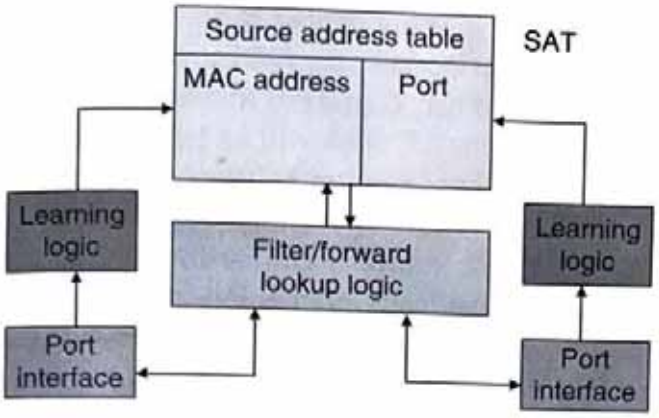
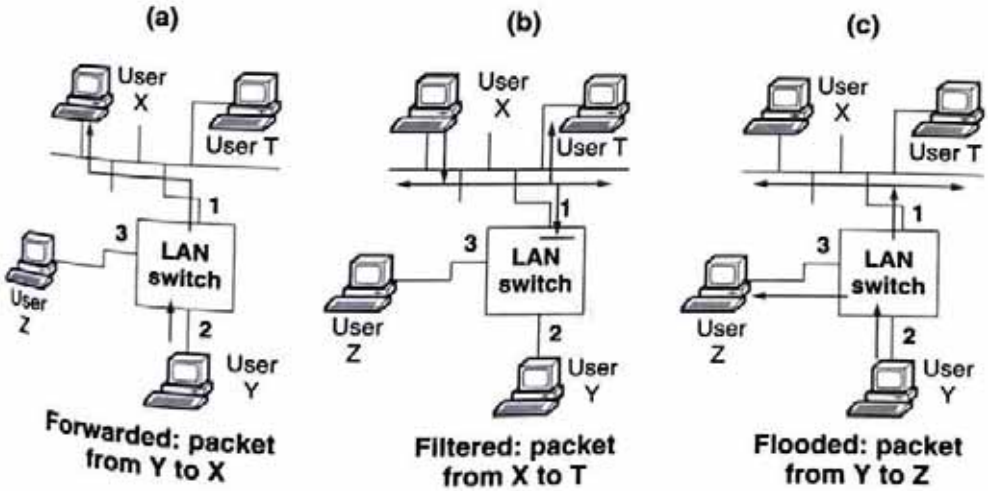


Figure 9.12 Bridge functional diagram (this figure may be seen in color on the included CD-ROM).

Automatically learned SAT entries will be aged out (i.e., deleted) if a source address becomes inactive for a certain period of time. This allows the MAC host to be moved from one location to another without the tedious requirement to reconfigure the SAT manually.

Figure 9.13 illustrates the three functions performed by a bridge. In Figure 9.13(a), user Y attached to bridge port 2 sends a frame to user X attached to bridge port 1. Upon receiving the frame at port 2, the bridge looks up X in the SAT and found it associated with port 1. The frame is thus forward to user X through port 1. In Figure 9.13(b), user X is sending a frame to user T. User X's frame is intercepted by the bridge at port 1. The bridge looks up the SAT and find that both user X and user T are both attached to port 1. So it filters the frame at port 1. A third example is shown in Figure 9.13(c). In this case, user Y's frame for user Z



Source address table	
MAC address	Port
X	1
Y	2
T	1

Figure 9.13 Illustration of bridge operation (this figure may be seen in color on the included CD-ROM).

arrives at bridge port Z. Since Z is not contained in the SAT, that frame for user Z is flooded by the bridge to all ports except the incoming one. This way, Z will eventually receive its frame. When Z starts to transmit frames to other users (e.g., Z responds to the received frame), Z's SA will be learned by the bridge so that the next frame bounded for Z will not need to be flooded again.

Another situation an incoming frame is flooded is when the received frame is a broadcast or multicast frame. It can be seen that forwarding and filtering help to preserve the bandwidth in other parts of the network where the frame does not need to be flooded. As far as the end users are concerned, the existence of the bridge is transparent as all the LANs (collision domains) interconnected by a bridge form a single broadcast domain. An outgoing user frame will "magically" arrive at its destination host no matter which bridge port it is attached to. There is no address translation or frame encapsulation required when Ethernet frames are forwarded from one section of the network to another section.[6] Thus Ethernet bridging is also called transparent bridging.

9.4.3 Spanning Tree Protocol

It is not difficult to imagine that two or more bridged LANs can be transparently joined to form a larger LAN network by interconnecting bridge ports. When multiple bridges are connected together, there is a possibility to form loops of forwarding paths. Forwarding loops cause a problem called broadcast storm. An example is shown in Figure 9.14. Imagine a broadcast frame arriving at one bridge port. This frame will be broadcast to all other outgoing ports to arrive at another bridge. Each bridge seeing the broadcast frame will broadcast it to all outgoing ports. We end up with a situation that the broadcast frame is circulating and replicating itself exponentially in the network, eventually exhausting all the bandwidth resources.

Another problem of having loops in a system is that a host can be reached through multiple paths. This creates confusions in the bridge learning and forwarding logic. Nonetheless, the availability of multiple paths offers redundancy to allow network resilience in the case of link failures, because traffic can take alternate route to the destination.

The solution to the above problems is to avoid multiple forwarding paths from being formed in a bridged network using the STP. In the STP, all bridge ports regularly send out Bridge Protocol Data Units (BPDUs) to its link partner to exchange the topology information. The BPDUs are well-formed Ethernet frames using one of the aforementioned reserved multicast protocol destination addresses. (BPDUs use the MAC address 01–80-C2–00–00–00h). Each bridge port will only exchange BPDUs with its link partner, which will not be forwarded. Any host receiving a BPDU without being able to understand it will simply discard the

[6] Ethernet MAC address space is large enough to give each host a universally unique six-octect address.

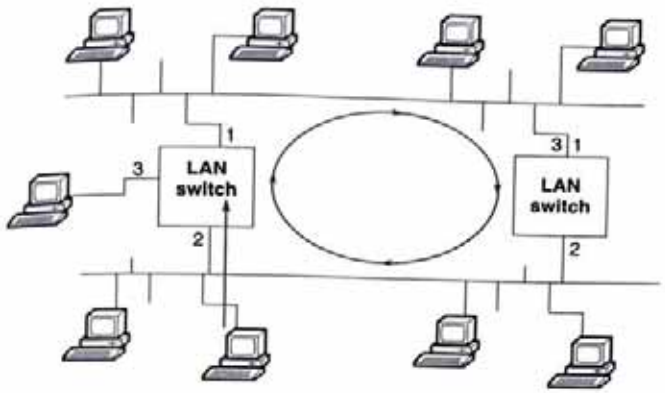


Figure 9.14 Loops formed by multiple bridges. Broadcast traffic sent from a host will keep looping in the network; eventually use up the bandwidth resources (this figure may be seen in color on the included CD-ROM).

BPDU. After exchanging enough BPDUs, a root bridge is elected by the bridges participating in the STP. The redundant links are disabled from forwarding traffic (i.e., user data frames) by putting the ports connecting the link ends into the blocked state so that each LAN is connected to the root bridge only through a designated port on a designated bridge. Bridges elect the root bridge, designated bridge and designated ports based on a set of priority criterions such as port speed, bridge and port IDs, and/or manually provisioned cost parameters. For space reasons, the details of the spanning tree algorithm will not be discussed here. Interested readers should refer to IEEE 802.1D [27] for the details.

An example is given here in Figure 9.15. Figure 9.15(a) shows the physical connectivity of a LAN with multiple interconnected bridges. Multiple paths of forwarding loops are possible in this physical topology. After running the spanning tree protocol, the ports marked with crosses [Figure 9.15(b)] are put into blocking mode. The forwarding loops are removed and an active tree topology with Bridge 1 as the root bridge is formed.

One should realize that links blocked from forwarding traffic are still existent in the resultant physical network. Whether a port is in the active forwarding state or blocked state, STP is running continuously with every port constantly exchanging BPDUs with its link partner. The blocked ports only block user data frames from being forward.

When an active forwarding link goes down, the expected BPDU frames will be lost and the ports on its two ends will time out. This will trigger the STP to send advertisement messages to all connected bridges to recalculate the new active topology. Redundant links which were blocked before may then be activated (i.e., changed into forwarding state) to restore the traffic.

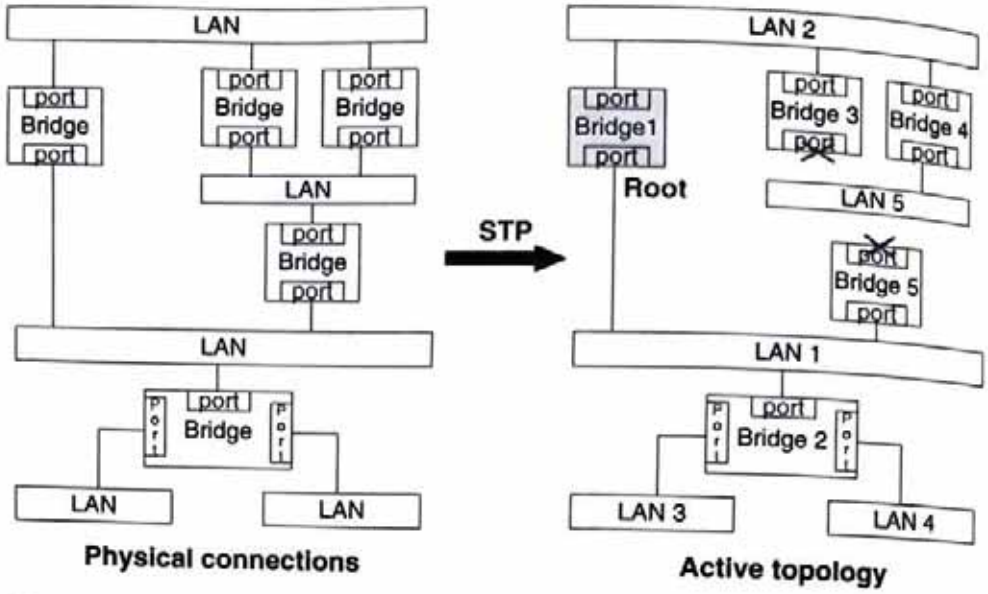


Figure 9.15 (a) Physical connections of a local area network connected with multiple bridges. (b) After running STP, the ports marked with crosses are set into block state so that forwarding loops are removed in the resulting active tree topology with Bridge 1 as the root of the tree (this figure may be seen in color on the included CD-ROM).

9.4.4 Limitations of STP

Standard spanning tree protocol usually takes tens of seconds to restore the traffic in the case of a link failure. For burst mode links, when the network is inactive, there is no physical signal activity in the transmission channel. A link down can only be determined by BPDU time out. In the case of P2P full duplex links, when a physical link goes down, it will cause the keep-alive idle symbols to be lost or PCS coding violation, so that bridges do not need to wait for BPDU time out to notice a link failure.

A later variation of the spanning tree protocol called Rapid Spanning Tree Protocol (RSTP) was standardized as IEEE 802.1w [29], which restores traffic in a matter of a few seconds after a link failure.

The STP has the advantage of self-configuration. Once bridge ports are connected, there is no need for operator configuration and provision. However, for a given physical connectivity, the active logical tree topology formed by STP is static and cannot be adapted to the actual traffic pattern. An example is shown in Figure 9.16. In this figure, N1 to N5 represent bridges connected in a physical ring topology. Under normal conditions, the STP will elect bridge N1 as the root bridge and block the link between bridges N3 and N4 to remove the forwarding loop (i.e., the link between N3 and N4 will not forward traffic unless some other links in the ring becomes broken). As a result, any traffic between LAN3 and LAN 4 needs to travel across a long link through bridge N1, even though there is a direct path between bridge N3 and N4. The root bridge N1 can easily become the network performance bottleneck. In contrast, Layer 3 protocols use other mechanisms to

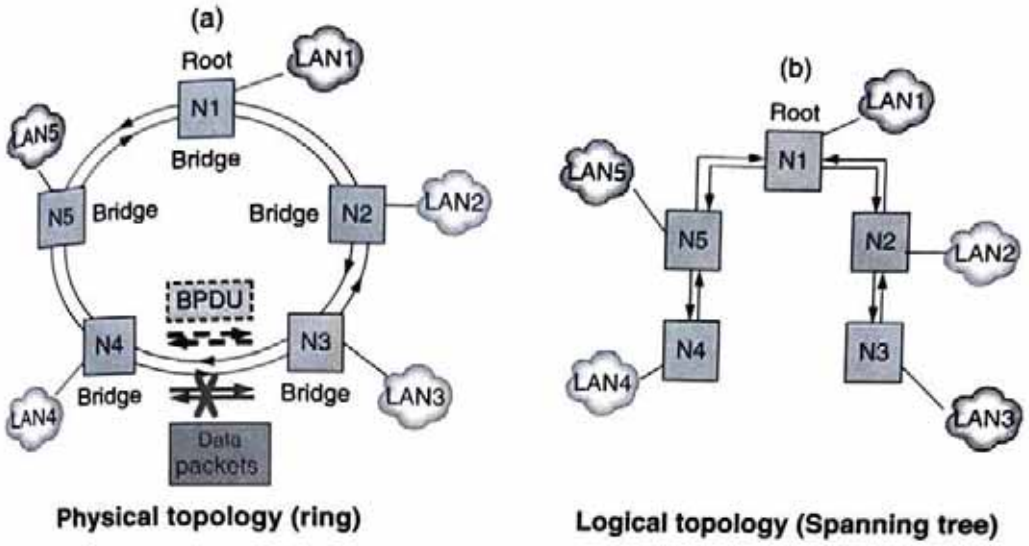


Figure 9.16 Inefficient use of link resources in STP. (a) Physical topology and (b) logical topology after running STP (this figure may be seen in color on the included CD-ROM).

avoid infinite data loops and broadcast storms.[7] They can also make use of multiple routing paths for load distribution.

MPLS (Multiprotocol Label Switching) is another method to perform IP/Ethernet traffic switching and engineering. In a packet-switching world, to achieve SONET-like rapid reroute and traffic restoration (on the scale of 50 ms) in response to link failures, MPLS fast reroute can be employed. Packet streams in an MPLS network are routed on circuit-like label switched paths (LSPs), which are formed between the source and destination nodes. A back-up LSP on a diverse route is precalculated when an LSP is formed. Path statuses are monitored by end-nodes exchanging "Hello" messages. An example of MPLS fast reroute protocol is the RSVP-TE (Resource Reservation Protocol–Traffic Engineering) protocol [30]. By default, in RSVP-TE, a "Hello" message is exchanged between the nodes at ends of a link every 5 ms. A node receiving the "Hello" message should respond to its link partner with "Hello Ack." A link is declared failure when proper "Hello Ack" is missed in 3.5 "Hello" intervals (i.e., 16.5 ms). When the primary path fails, traffic is quickly rerouted to the backup path.

9.4.5 VLAN and VLAN Stacks

VLAN Basic

As explained before, network hosts connected by bridges form a single broadcast domain. The virtual bridged LAN (VLAN) technology [31] segregates a

[7] A method called time to live (TTL) is widely used in routing protocols to prevent infinite loops and remove orphanage packets.

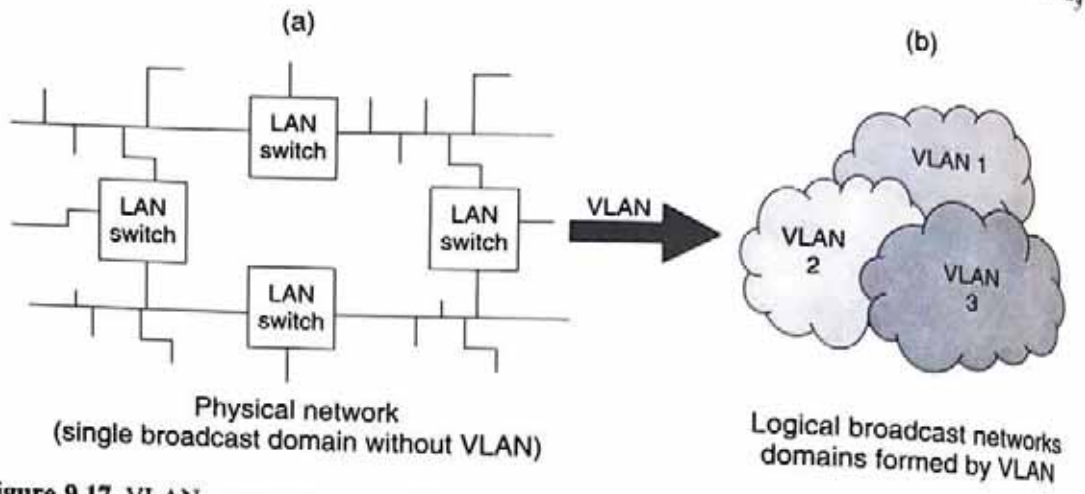


Figure 9.17 VLANs segregate physically connected LAN network (a) into multiple logical broadcast networks (b) (this figure may be seen in color on the included CD-ROM).

network physically connected by bridges into multiple logical broadcast domains (Figure 9.17).

VLAN offers the following network advantages:

(1) It limits broadcast traffic to smaller groups and improves network performance.
(2) It provides network privacy and security by separating traffic belonging to different organizations.
(3) It eases network management by allowing operators to assign network ports to different VLAN groups.

VLAN bridges can be implemented so that each logical broadcast domain can have its own separate spanning tree, thus allowing operators to have better control of the resulting logical network topology and traffic distribution [29].

VLAN bridges need to classify and tag data frames so that they can be segregated according to the VLAN they belonged to. This is achieved by adding a four-octect VLAN tag (also called Q-tag) after the SA field of the original Ethernet frame as shown in Figure 9.18. The first two octets correspond to the length/type field of the original Ethernet frame and contain a length/type field value of hexadecimal value 0×81-00. The next two octets represent the tag control information which includes a 3-bit user priority field, a 1-bit canonical format indicator (CFI), and a 12-bit VLAN ID (VID). The 3-bit user priority field can be used to implement eight service quality classes. The CFI field was designed for use with token ring technologies and has no significance any more except for backward compatibility. The 12-bit VID allows 4094 different VLANs to be supported. (The 0 VID is used to represent priority frames and VID 0×FFF is reserved.) Frames belonging to a particular VLAN will only be broadcast/forwarded to hosts on the same VLAN.

Figure 9.19 shows the block diagram and operation of a VLAN bridge. In a VLAN capable environment, there are three types of Ethernet frames: (1) untagged

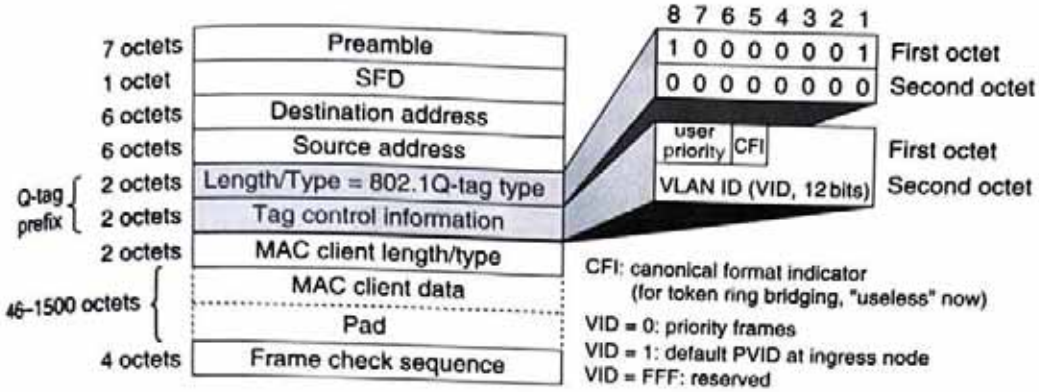


Figure 9.18 VLAN tagged Ethernet frame format (this figure may be seen in color on the included CD-ROM).

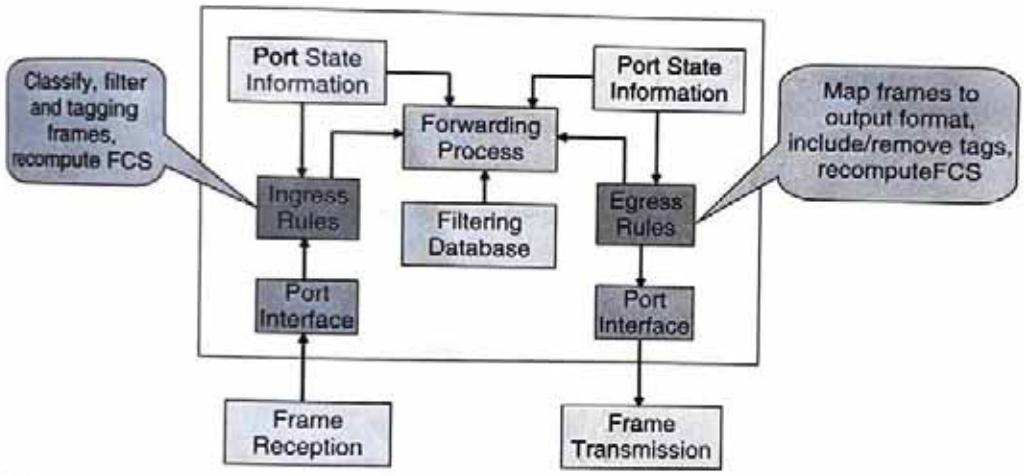


Figure 9.19 Block diagram of a VLAN bridge (this figure may be seen in color on the included CD-ROM).

frame, (2) priority-tagged frame (VID = 0), and (3) VLAN-tagged frame. Similar to the transparent bridging philosophy described previously, as far as end users are concerned, the existence of VLAN is transparent.

All end user frames in access networks are normally untagged frames. When a user frame is received by a VLAN bridge port shown in Figure 9.19, it is processed by a set of ingress rules. The ingress rules add a VLAN tag to the user frame based on some provisioned rules. The most common type of VLAN classification is port-based in which each access port is assigned with a provisioned VID. VLANs can also be associated with MAC addresses or upper layer protocol identifiers encapsulated within Ethernet frames. In the latter two cases, the ingress rules also need to perform filtering and classification functions. Another function performed by the ingress rule is recalculation of the FCS after inserting the VLAN tag. At the output of a VLAN bridge port, a set of egress rules are performed to remove the tag and recalculate the FCS. Thus the end users have no idea of the existence of VLANs.

In a port-based VLAN, the operator can easily change a user from one broadcast domain to another broadcast domain by changing the VID of the port that the

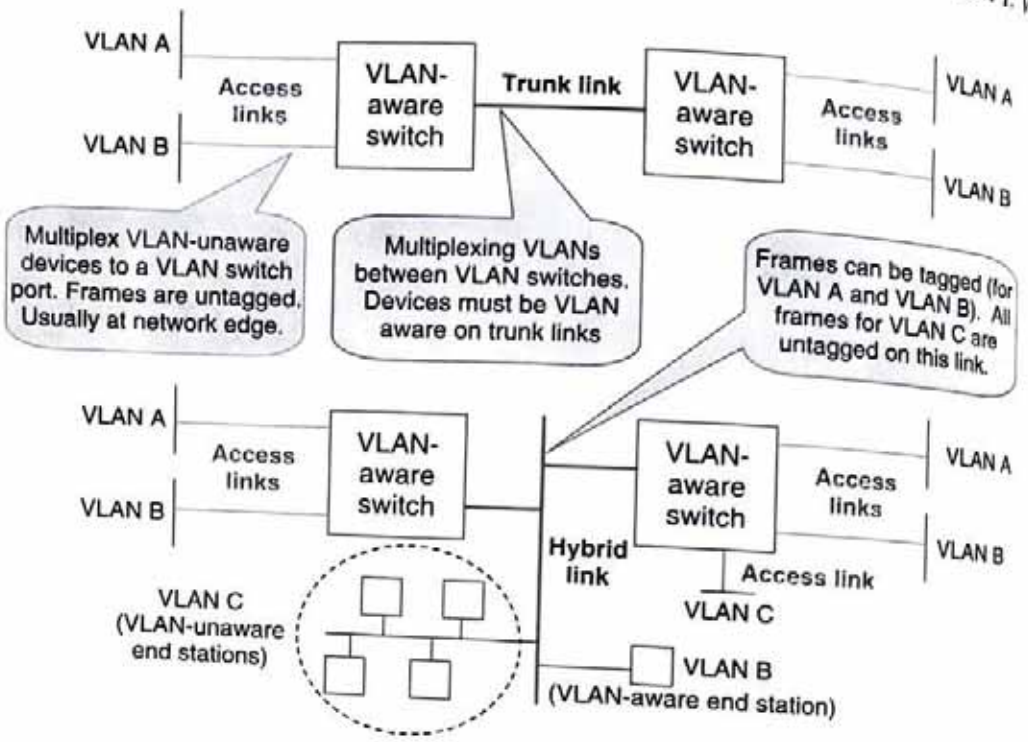


Figure 9.20 Access, trunk and hybrid links in a VLAN-enabled network (this figure may be seen in color on the included CD-ROM).

user is connected to. An IT manager can also separate users in different departments on different VLANs so that they do not have access to each other's network without going through a gateway.

Figure 9.20 shows three different types of links in a VLAN-enabled network. Frames transported in access links are traditional Ethernet frames with no VLAN tags. These links are usually connected to end users. The trunk links are usually links between VLAN bridges. All the frames on trunk links are VLAN tagged. In other words, trunk links are in effect multiplexing links of different VLANs. A third type of link is the less common hybrid link which can transmit both VLAN-tagged and untagged frames. In this case, all the untagged VLAN frames must belong to one and only one VLAN.

VLAN Stacks

The VLAN idea can be used by service providers to provide virtual layer-2 connectivity services to users over wide area backbone networks. However, two issues must be solved.

(1) The 12-bit VID supports only about 4000 different VLANs. This is very limiting in a carrier environment.
(2) VLAN has already been widely deployed in corporate LANs. Customers subscribing to the services would like to preserve their VIDs and be able to manage their own VID space.

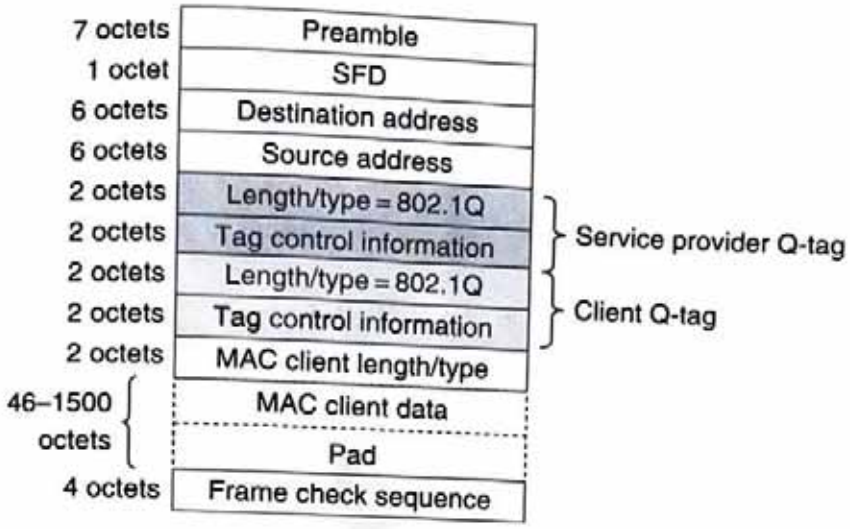


Figure 9.21 Q-tag stacking in IEEE 802.1ad provider bridges (this figure may be seen in color on the included CD-ROM).

The simple way to resolve the second problem is to stack another VLAN tag, called service VLAN tag (or S-tag), on top of the customer VLAN tag (or C-tag) as shown in Figure 9.21. This technique is also called Q-tag-in-Q-tag, or QiQ, and is standardized as IEEE 802.1ad Provider Bridges [32]. The QiQ technique nested customer VLANs (C-VLANs) inside service VLANs (S-VLANs) to achieve C-VLAN transparency from a customer point of view.

9.5 CARRIER ETHERNET

The cost advantage of Ethernet and the convergence of voice, data, and video services on packet-oriented network infrastructure made Ethernet services the fastest growing services in the telecommunication industry. Instead of requesting traditional TDM-based leased line services, more and more customers are now requesting Ethernet leased lines which not only are cheaper and but also have wider bandwidths.

Traditionally, Ethernet has been managed by corporate IT personnel. Ethernet service definitions and management are new territories to telecom service providers. The new challenges facing a carrier class Ethernet transport system includes scalability, OAM, availability, and security. We will touch some of these subjects in the following sections.

9.5.1 Scalability

To solve the limited 802.1ad VLAN address issue, the IEEE 802.1ah [33] provider backbone bridge (PBB) standard was created. In the IEEE 802.1ah standard, a service provider SA and DA is stacked on top of the customer addresses. This provides a virtually unlimited address space for operators to support as many

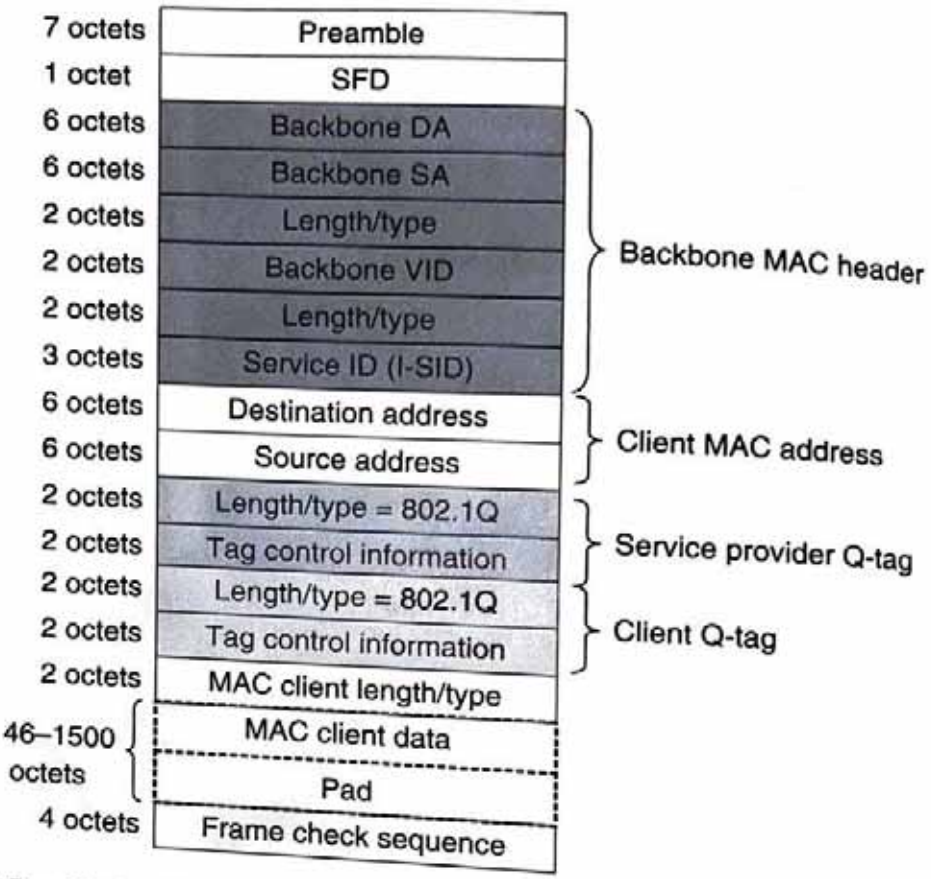


Figure 9.22 Provider backbone bridge (PBB) MAC frames (this figure may be seen in color on the included CD-ROM).

customer VLANs as they want. Since the customer MAC addresses are nested within service provider addresses, this method is also nicknamed as MAC-in-MAC or MiM. The Mac-in-Mac stacking is shown in Figure 9.22. Interested readers are referred to [33] find more details from IEEE 802.1ah.

9.5.2 Ethernet Transport

The simplest Ethernet service is the P2P Ethernet signal transport over long distances, traditionally called transparent LAN. Figure 9.23 shows two approaches to Ethernet signal transport over a long distance WDM optical network. The first approach puts native Ethernet frames directly on optical wavelengths. The symbols transmitted on the physical link use Ethernet PCS line coding (i.e., 8B10B for Gigabit Ethernet and 64B66B for 10 Gigabit Ethernet). In this case, the transport signal is identical to the signal presented at the UNI and complete physical layer transparency is achieved.

Another approach to transport Ethernet frames is to make use of the widely deployed legacy transport networks which were built with other technologies such as SONET/SDH [34] or ITU-T G.709 OTN (optical transport network) [35]. Ethernet frames are encapsulated in other transport frames such as the SONET SPE

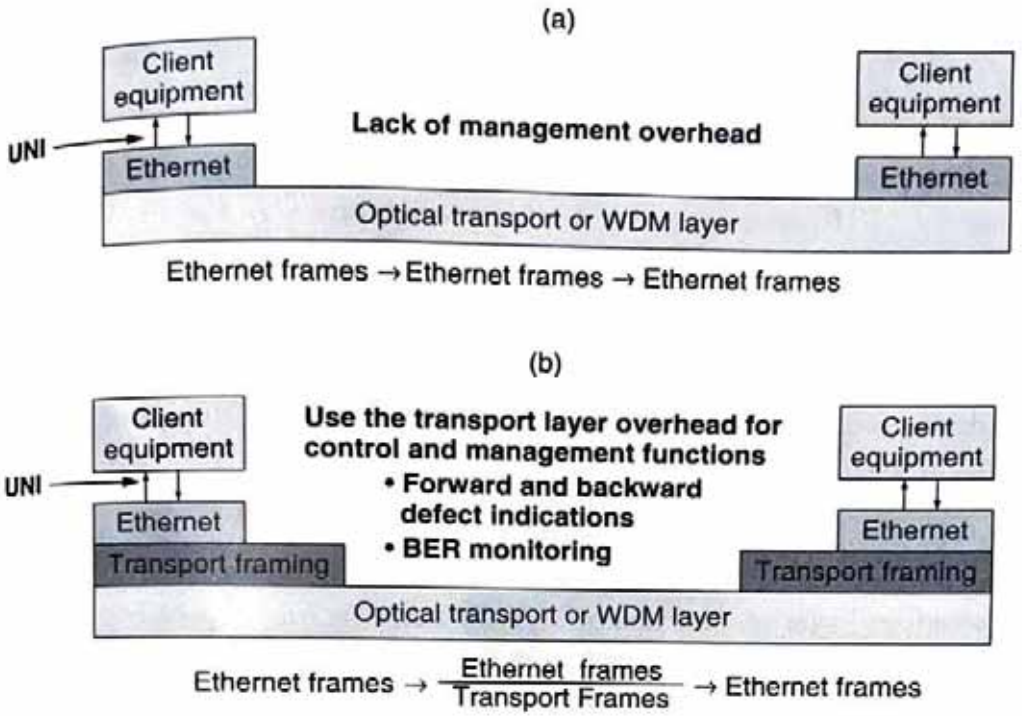


Figure 9.23 (a) Direct native Ethernet transport over WDM optical layer and (b) Ethernet transport over transport framing (this figure may be seen in color on the included CD-ROM).

(synchronous payload envelope) before being placed on the optical layer [36, 37]. In this approach, Ethernet is only used as the UNI between the client equipment and the transport network. Legacy transport networks were built with extensive management facilities and mechanisms for fault tolerance and fast recovery, which complements the corresponding inadequacies of Ethernets, at least during the transitional period before the Ethernet community develops its own.

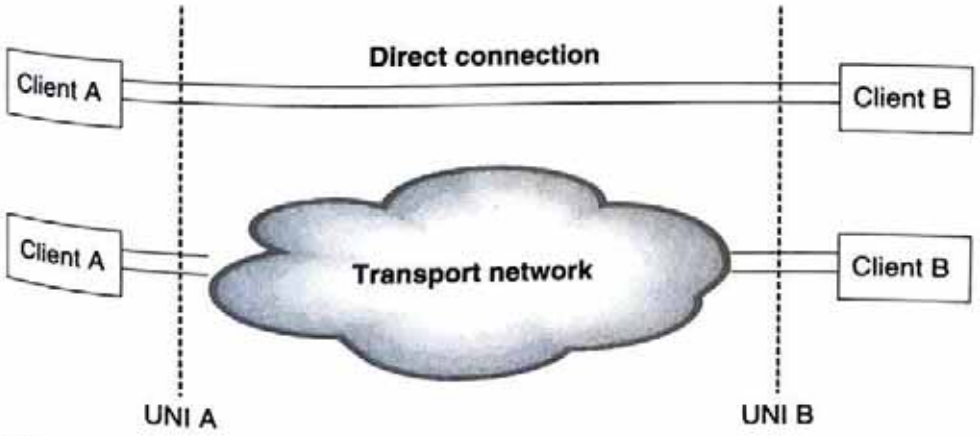


Figure 9.24 Client equipment joined by direct fiber connection (top) and transport network (bottom). From a user's perspective, the transport network should be transparent to client equipment, i.e., ideally it should behaves as a pair of wires joining UNI A and UNI B as in the top diagram (this figure may be seen in color on the included CD-ROM).

From a user's perspective, the underlying technology employed for transporting the client Ethernet frames should be transparent. Figure 9.24 shows the comparison of client equipment joined by direct fiber connections between clients' Ethernet interfaces (top) and by a transport network (bottom). Ideally, from a customer's perspective, the transport network between the two dash lines which indicate the UNI demarcation between a service provider and the customer network, should behave as a pair of wires.

In reality, the ideal transparency as described above cannot always be achieved and may not always be the desired behavior either. Despite the difficulty to maintain complete transparency, the transport equipment should provide MAC frame transparency at the minimum, i.e., no filtering and dropping of customer Ethernet frames should occur within the transport network.

Figure 9.25 shows an example of multiplexing dual Gigabit Ethernet (GbE) over a SONET OC-48 link. A SONET OC-48 link has a data rate of 2.488Gb/s. Although the MAC data of GbE runs at 1.0 Gb/s, after the 8B10B PCS encoding, a GbE interface has a symbol rate of 1.25 Gbaud/s. So it is impossible to multiplex two GbE physical layer signals onto an OC-48 payload and maintain both client interfaces at line rate at the same time. As indicated in the figure, the client signals are terminated at the MAC layer to strip off 8B10B encoding before being multiplexed using POS [36, 37]. Physical layer functions implemented using Ethernet PCS control codes such as auto-negotiation [18, Clause 37] are terminated locally. In addition, the clocks at the Ethernet interfaces need to be decoupled from the SONET clock because of the differences in clock rate and accuracy requirements.[8]

In return for the loss of physical layer transparency, one has obtained the SONET manageability in the transport network, as well as better optical layer utilization by multiplexing two Ethernet streams onto one single SONET wavelength. In addition, forward error correction (FEC) such as that available from

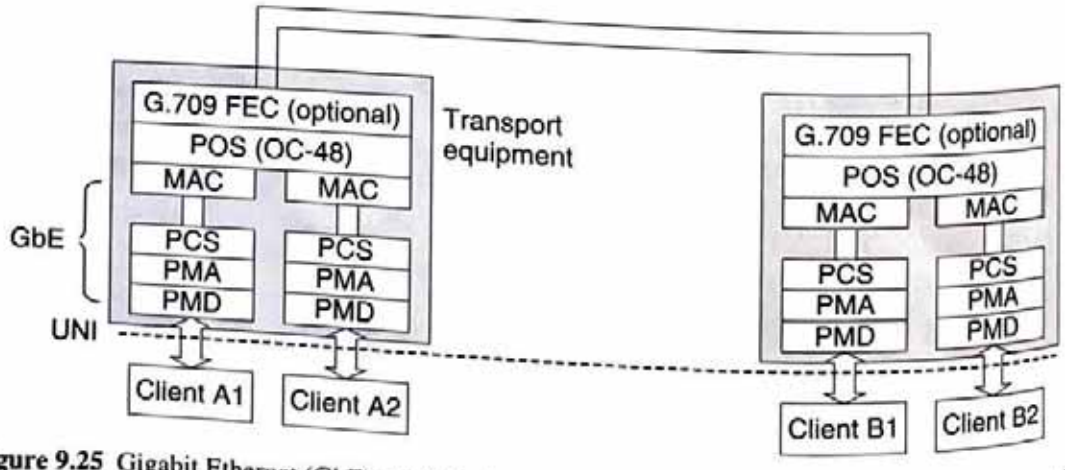


Figure 9.25 Gigabit Ethernet (GbE) multiplexing using packet over SONET (POS) (this figure may be seen in color on the included CD-ROM).

[8] Ethernet uses asynchronous transmitters. Each receiver recovers the clock from the received signal. The transmitter clock and receiver clock are completely independent.

G.709 OTN can be added to increase the signal transmission distance without intermediate regeneration.

In a later section, we will see the implications of client/transport interface decoupling in different network protection switching scenarios.

9.5.3 Generic Framing Procedure

GFP is standardized as ITU-T G.7041 [12, 38], which is used to encapsulate packet data (including Ethernet frames) for transport. There are two different types of GFP: frame based GFP (GFP-F) and transparent GFP (GFP-T), which are shown in Figure 9.26 and Figure 9.27.

As shown in Figure 9.26, GFP-F only encapsulates the contents of Ethernet frames. It starts with a payload length indictor and header error control (HEC) fields, which also marks the beginning of GFP frames. Ethernet line coding overhead, the redundant preamble and SFD fields are removed in GFP-F encoding and recreated in GFP-F decoding. Therefore, GFP-F helps to preserve the transport bandwidth.

GFP-T was designed to transparently transport 8B10B code words. In addition to the actual data words, the 8B10B code words also include idle symbols transmitted during IFGs (interframe gaps) between adjacent Ethernet frames, and the physical layer control codes such as the link negotiation words used by Gigabit Ethernet. Not all the 10-bit words are used in 8B10B codes. GFP-T converts fixed-length blocks of 8B10B code words and convert them into 64B65B codes [38].

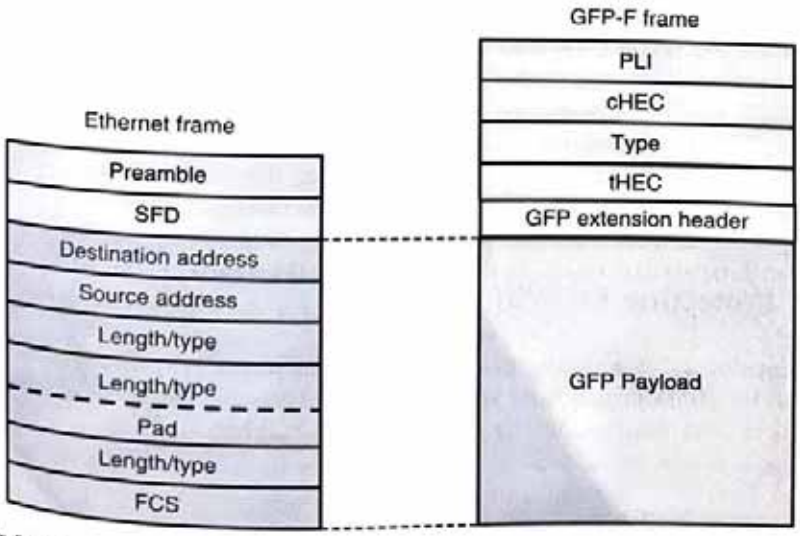


Figure 9.26 Frame-based GFP (GFP-F). PLI: payload length indicator, cHEC: core header error control, tHEC: type header error control (this figure may be seen in color on the included CD-ROM).

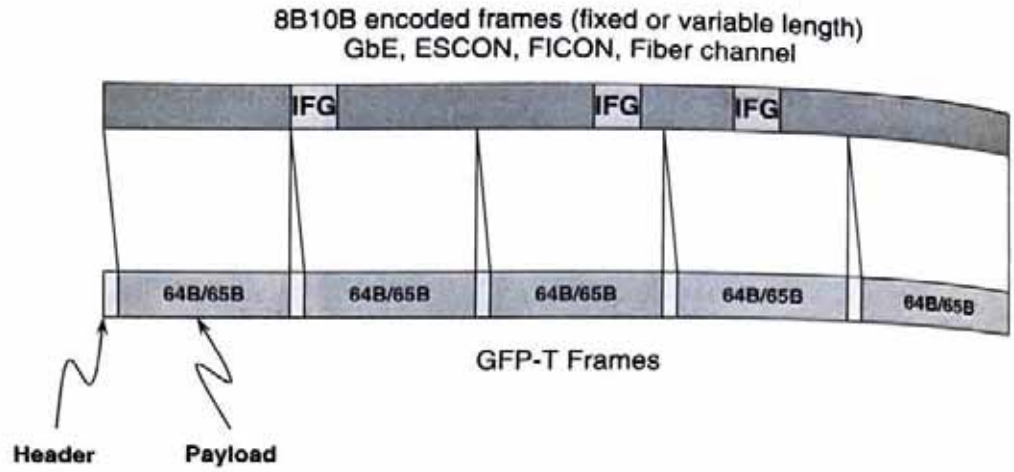


Figure 9.27 Transparent GFP (GFP-T). IFG: interframe gap (this figure may be seen in color on the included CD-ROM).

All the physical layer control codes in 8B10B codes are included in 64B65B codes and recreated at the other end of the link.

The advantage of GFP-T is that it transparently preserves the end-to-end 8B10B physical layer signaling with a reduced overhead of only 1.5% as opposed to 25%. Moreover, there is no need to wait for the whole frame to be received before encapsulation and thus reduces the transport latency.

GFP frames are usually further encapsulated in SONET or OTN frames. Ethernet speeds increase by multiple of 10 from generation to generation. However, SONET signal rates increase by multiples of 4. The lowest SONET rate, OC3, is 155 Mb/s. This mismatch makes it difficult to efficiently map Ethernet signals on SONET signal hierarchies. To make efficient use of the transport network bandwidth, the ITU-T VCAT standard G.7043 [13] was created to allow GFP frames to be inversely multiplexed to SONET/SDH tributaries with 1.5 Mb/s VC-1.5 granularity. The ITU-T G.7042 [14] LCAS allows dynamic adjustment of the number of inverse multiplexed tributary streams and thus the bandwidths used for carrying Ethernet traffic in a transport network.

9.5.4 Protection Switching

There are many different ways to perform Ethernet protection switching at various network layers. In this section, we briefly describe some approaches and considerations besides the spanning tree protocol described before.

Link Aggregation

Link aggregation [18, Clause 43] is a method to increase data throughput by bundling multiple Ethernet links in parallel to form a link aggregation group

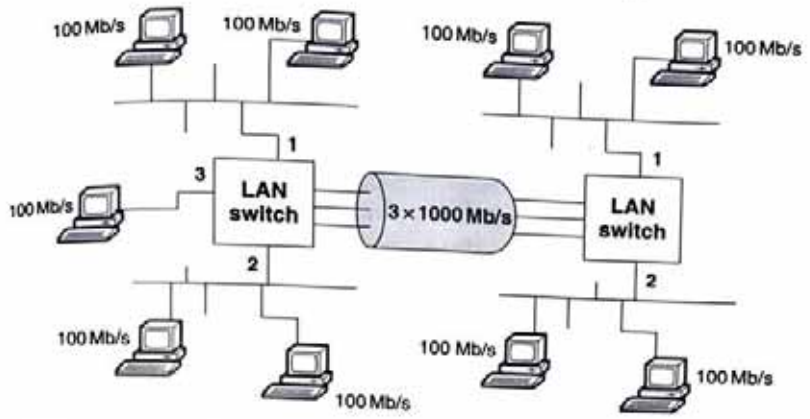


Figure 9.28 Link aggregation increases the throughout by bundling parallel Ethernet links (this figure may be seen in color on the included CD-ROM).

(LAG) as shown in Figure 9.28. In this figure, the three 1000 Mb/s links form a single logical link with an effective throughput of 3000 Mb/s.

In normal bridge connections, parallel links will form forwarding loops and the STP will block all but one link for forwarding traffic. In a LAG, all the parallel links represent a single logical link. All the Ethernet interfaces at one end of a LAG represent a single logical MAC interface with one shared MAC address, which can be the MAC address of one of the parallel interfaces.

One of the requirements in Ethernet frame delivery is to maintain the order of frames as there is no sequence number embedded in Ethernet frames for reordering out of sequence frames at the destination node. The parallel links in a LAG create the opportunity for out-of-order frame delivery at the egress point. To avoid this issue, link aggregation makes use of higher layer protocol signatures (such as IP source and destination addresses, and TCP/UDP port numbers) in load balancing so that Ethernet frames belonging to the same application stream (i.e., identified by the same IP SA, IP DA, and/or TCP port number) are always sent through the same component link in an LAG. Thus the frame delivery order is preserved from source to destination on a per application basis. A hash function based on upper layer protocol signatures is usually used to calculate the link in an LAG that will be used for a particular stream of data frames.

Before 100GbE becomes available, LAG is used by many service providers to aggregate multiple lanes of 10 GbE traffic. There are, however, many problems associated with LAG, which motivate the industry to go for a full-fledged 100 GbE [39–42]:

(a) LAG distributes traffic over parallel links via flow-based hash mechanism, which cannot be used to re-order frames from parallel links and cannot guarantee equal distribution of load.
(b) "Special" traffic (multicast, broadcast, control traffic, etc.) usually traverses a single component link, and load balance is lost.

(c) The unpredictable link removal and insertion make LAG operational cumbersome.
(d) Transponder cost for multiple 10 GbE LAG in WAN is very high.

Link aggregation offers resiliency to link failures. If one of the component links in an LAG fails, its traffic will be redistributed to the remaining links, thus providing a graceful degradation. The disadvantage of link aggregation is that it only works as P2P links and with interfaces of the same speed.

Protection in the Presence of the Transport Layer

An advantage of using an optical transport layer for Ethernet backbone is the ability to implement protection switching in the transport layer. Transport equipment generally offers much faster protection switching than data equipment in a network failure, usually within 50 ms. Understanding the interaction between the data layer and the transport layer is important in designing wide area Ethernet backbones that gives the best protection performance.

Figure 9.29 shows a pair of routers/bridges connected back-to-back directly by an Ethernet link. In general, routers or bridges will also continuously verify the link integrity by exchanging keep-alive PDUs (the so-called Hello message). To reduce the bandwidth overhead, such keep-alive PDUs are exchanged with a very low frequency. By default, the "Hello" message interval in a 802.1D bridge is set to 2 sec. When a predetermined number of PDUs are not received after a timeout period, the link is declared lost. As a result, routers/bridges will send topology advertisement messages to the rest of the network to recalculate the routing table or the new spanning tree so that an alternative data path can be found. In addition to PDU timeout, when the link is broken as shown in the figure, both port interfaces connecting to the link detects the physical link down condition immediately. A physical link down can trigger new forwarding path calculation in real time, thus minimizing the network unavailable time.

When a transport layer is present, as mentioned before, the transport layer and the Ethernet layer are physically decoupled. Consequently, a failure in the transport interface may not necessarily cause link down at the Ethernet interface as shown in Figure 9.30. In such a scenario, the client devices have no choice but to rely on the PDU timeout mechanism to detect the link outage (which can cause a delay as long as 20 sec). To overcome this problem, the transport equipment could bring down the client interface intentionally when a transport link fault is detected.

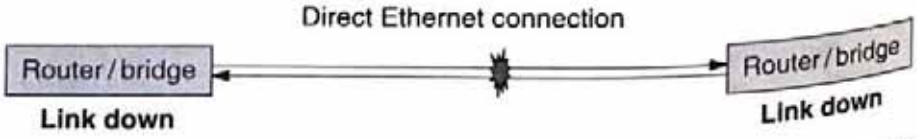


Figure 9.29 A physical link failure will trigger link down at both port interfaces immediately (this figure may be seen in color on the included CD-ROM).



Figure 9.30 A physical link failure on the transport side may not necessarily cause the client interface link down (this figure may be seen in color on the included CD-ROM).

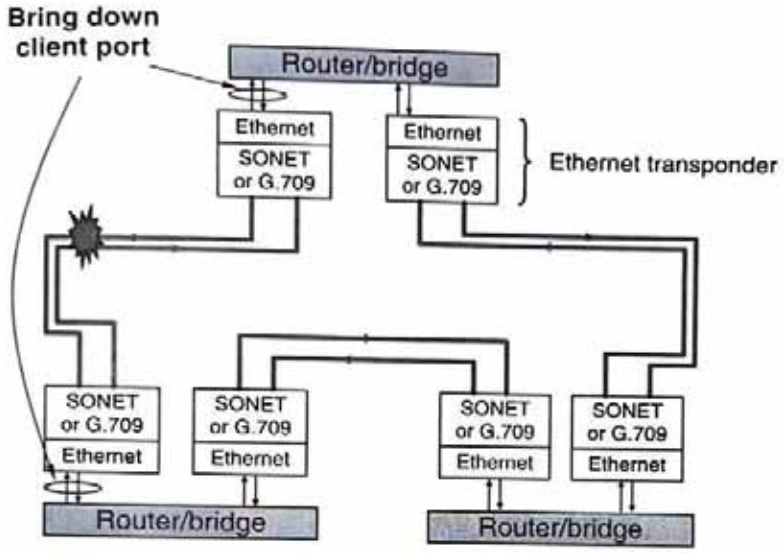


Figure 9.31 A wide area Ethernet ring network using SONET/G.709 for transport. Restoration is performed at Layer 3 or Layer 2 by routers or bridges. To speed up traffic restoration time, when the transport link is lost, the associated client ports are brought down (this figure may be seen in color on the included CD-ROM).

Figure 9.31 shows a wide area Ethernet using SONET/G.709 as P2P transport layer only. Protection and restoration is performed by the router/bridges. In this case, a failure at the transport interface brings down the associated client ports to speed up the restoration time.

Figure 9.32 shows a scenario that the transport link is protected. The tail-end switches move the transport link to the protection path within 50 ms when the working path fails. This happens at a timescale much faster than timeout period of Layer 2/Layer 3 PDUs. In this case, if the transport equipment brings down the client interface upon detecting transport link failures, the client equipment will sense the link outage and start to recalculate the spanning tree or routing table right away. For L2 bridges, this will also cause the SAT to be flushed out and relearned, increasing the volume of broadcast traffic in the network temporarily. Therefore, the decoupling between the client and the transport interfaces is actually beneficial because it avoids racing between the data networking layer and transport layer in restoring the traffic when a failure in the transport network occurs.

Thus Ethernet transport equipment (or transponders) should be built so that the ability to bring down the client Ethernet interfaces can be enabled or disabled at appropriate times, depending on the actual network configuration.

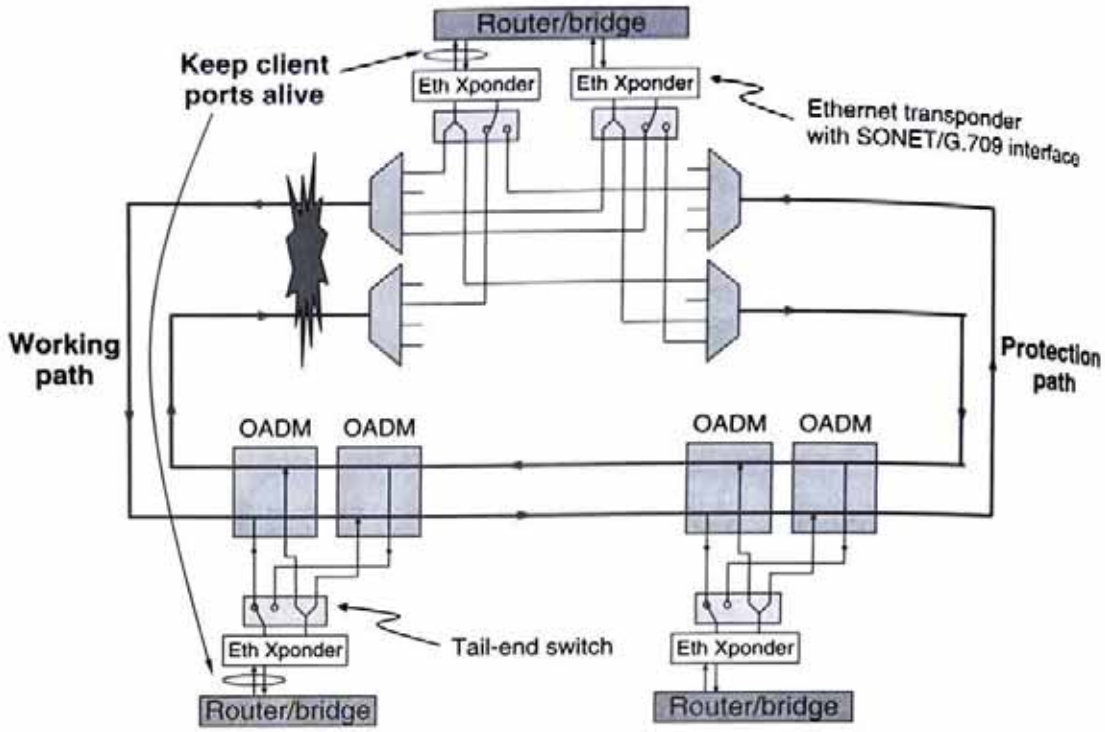


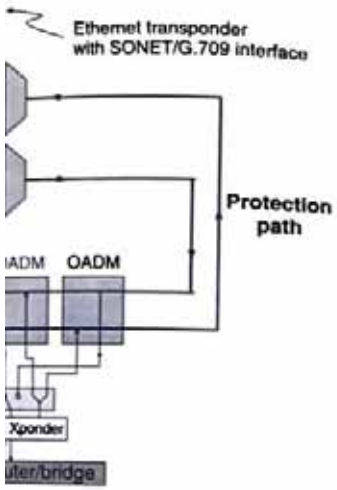
Figure 9.32 Ethernet transport with optical layer protection. A protected transport link should not cause the client link to go down to avoid racing conditions in protection switching (this figure may be seen in color on the included CD-ROM).

9.5.5 Ethernet Service Models and Service Level Agreements

It has been a challenge for carriers to offer Ethernet services because of the lack of standards in SLA definitions. The MEF was initially formed by equipment vendors to tackle Ethernet service models, SLA, and OAM issues [11]. The business value of a common set of languages and standards to specify Ethernet services was quickly recognized by carriers [both telecom operators and cable TV multiple service operators (MSOs)] who have later become the dominating participants at MEF.

MEF does not specify the technology used to implement networks. Instead, it defines the bandwidth and PDU behaviors at the UNI between customers and service providers, or the NNI (network–network interface) between service providers. MEF specifies Ethernet services as P2P (called E-Line) and multipoint-to-multipoint (called E-LAN) in terms of EVCs (Ethernet virtual circuits) [43]. From a customer's perspective, E-Line behaves as an Ethernet transport link whereas E-LAN behaves as a bridged network. EVC attributes include performance specifications such as bandwidth profile, packet loss rate, packet delay, and delay variations.

On the physical layer, Ethernet port interface speeds are all fixed. The effective data throughput, however, depends on actual frame lengths and IFG widths which can be stretched to limit the data throughput and perform ingress bandwidth shaping. MEF uses committed information rate (CIR), excess information rate (EIR), committed burst rate (CBR), and excess burst rate (EBR) as bandwidth profile attributes to specify SLA [43]. The CIR and EIR represent average bandwidth throughput over defined timing periods



rotected transport link should not cause
n switching (this figure may be seen in

rvice Level Agreements

l services because of the lack of
y formed by equipment vendors
ssues [11]. The business value of
y Ethernet services was quickly
nd cable TV multiple service
inating participants at MEF.
mplement networks. Instead, it
l between customers and service
etween service providers. MEF
:) and multipoint-to-multipoint
ircuits) [43]. From a customer's
r link whereas E-LAN behaves
ormance specifications such as
d delay variations.
s are all fixed. The effective data
s and IFG widths which can be
ss bandwidth shaping. MEF uses
rate (EIR), committed burst rate
le attributes to specify SLA [43].
hput over defined timing periods

and are measured in bits per second, whereas CBR and EBR are measured in bytes per second. MEF10.1 [43] specifies three levels of bandwidth profile compliances. Green represents service frames within the limits of CIR/CBR and will be guaranteed (i.e., subject to SLA). Yellow represents service frames exceeding CIR/CBR but within EIR/EBR. These are allowed but not subject to SLA. Red represents service frames exceeding EIR/EBR which are not allowed and will result in packet loss.

In addition to Ethernet service attribute model definition, MEF technical documents also cover the aspects of control plane, management plane [44], circuit emulation on Ethernet [45], mobile data back hauling using Ethernet, protection switching etc. Because of space limitation, it is impossible to thoroughly discuss MEF works in great details. Interested readers could download MEF technical specifications at MEF website free of charge [46].

9.6 ETHERNET PASSIVE OPTICAL NETWORKS

9.6.1 Ethernet Passive Optical Network Architecture

PON (passive optical network) is an access network technology with point-to-multipoint (P2MP) connectivity. More details on PON technologies can be found in Chapter 10 by Wagner. The P2MP connectivity makes EPON (Ethernet PON) very different (both from a transmission viewpoint and protocol view point) from the traditional CSMA/CD and P2P full-duplex Ethernets.

A PON is characterized by a passive remote node (RN) to distribute signals from an optical line terminal (OLT) located at a central office to a number of optical network units (ONUs) at customer sites [47]. Ideally, the fiber plant from the OLT to the ONUs is completely passive. A TDM-PON uses a passive power splitter as the RN. The same signal from the OLT is broadcast to different ONUs by the splitter. Signals for different ONUs are multiplexed in the time domain. ONUs recognize their own data through the address labels embedded in the downstream broadcast signal. EPON falls into the category of TDM-PON. A WDM-PON uses a passive WDM coupler as the remote terminal. Signals for different ONUs are carried on different wavelengths and routed by the WDM coupler to the proper ONU in a virtual P2P fashion. Since each ONU only receives its own wavelength, a WDM-PON has better privacy and better scalability. However, WDM devices are significantly more expensive, which made WDM-PONs economically less attractive at this moment.

The EPON protocol layering diagram can be found in Figure 9.2, along side with standard modern Ethernet layering. Figure 9.33 shows the architecture of an EPON infrastructure. Ethernet frames are used to carry the data between the OLT and ONUs in an EPON. The upstream traffic and downstream traffic are separated on 1.31 μm and 1.49 μm wavelengths, i.e., by wavelength division duplex. Usually, an OLT will serve 16–32 ONUs, which are separated by up to 20 km away from the OLT. The optical power budget between the OLT and ONU eventually limit the transmission distance as well as the number of ONUs that can be supported.

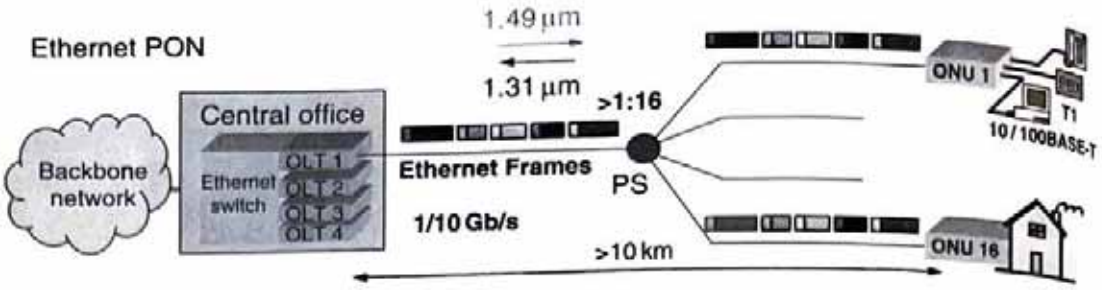


Figure 9.33 Architecture of an EPON infrastructure (this figure may be seen in color on the included CD-ROM).

Current EPON standard (802.3ah) specifies 1000 Mb/s data throughput (1.25 Gbaud/s physical symbol rate) on the feeder link between the OLT and ONU. The downstream signal is broadcast to all ONUs in continuous mode. Upstream signals from different ONUs all merge at the power splitting RN. Because a single OLT receiver is shared among all ONUs, when an ONU is not transmitting, it should turn off its transmitter to avoid interfering with other ONUs' upstream signal. Without coordination, upstream frames from different ONUs will collide in time.

9.6.2 Multipoint Control Protocol

In order to avoid collision, all the upstream transmission is scheduled by the OLT centrally. The CSMA/CD mechanism cannot be used in a PON system for the following reasons: (1) the directional power splitter makes carrier sense and collision detection impossible because without using special tricks no ONU can monitor the optical transmission from other ONUs on the same PON; and (2) the data rate and distance covered by a PON system greatly exceeds the limits imposed by the CSMA/CD protocol. The CSMA/CD protocol becomes very inefficient under conditions of high bandwidth and long transmission distance [48].

In EPON, scheduling is performed by the MPMC (multipoint MAC control) layer using the Multipoint Control Protocol (MPCP) [18, Clause 64]. As indicated in Figure 9.2, the MPCP protocol entities in the OLT and ONU form a master-slave relationship with the OLT as the master. In the MPCP protocol, the OLT schedules the starting time and duration for an ONU to transmit upstream data bursts using the Gate MPCPDU (Multipoint Control Protocol Data Unit). ONUs inform the OLT their buffer status using the Report MPCPDU. Based on the reported information, the OLT can dynamically allocate the upstream bandwidth to make the most efficient use of the shared link between the OLT and ONUs.

Since ONUs are located at different distances from the OLT, signals from different ONUs will experience different delays before reaching the OLT. It is therefore important to establish a timing reference between the OLT and an ONU so that after accounting for the fiber delay, when the ONU signal arrives at the OLT, it arrives at precisely the same moment that the OLT intends for the ONU to transmit. The timing reference between the OLT and ONUs is established through the *ranging* process.

Ranging measures the round trip delay between the ONU and OLT, also using the Gate and Report MPCPDUs, which have time stamps embedded. From the time stamps in Gate and Report MPCPDUs, the OLT measures the round trip time (RTT), which is then stored and used to adjust the time that data frames from an ONU should be transmitted. All ONUs are thus aligned to a common logical time reference after ranging so that collision does not occur in a PON system.

From time to time, an EPON OLT will periodically broadcast Discovery Gate messages to discover unregistered ONUs. A new ONU joining the network detects the Discovery Gate and responses with a Register Request to the OLT. After sending the Discovery Gate, an OLT must reserve a time period called discovery window for ONUs that have not been ranged to response. The size of the discovery window depends on the maximum differential delays between the closest ONU and the furthest ONU. Optical signal delay in 1 km of fiber is 5 μs. Therefore, for 20 km of differential distances between ONUs, an RTT difference of 200 μs needs to be reserved in the discovery window. It should be realized that if the upper bound and lower bound of ONU distances are known to the OLT (e.g., through management provision), then instead of reserving a ranging window covering the maximum allowed separation between an ONU and an OLT, the size of ranging window can be reduced to cover only the maximum differential distance among ONUs. If multiple ONUs attempt to join the PON at the same time, collision may occur during discovery. This is resolved by ONUs backing off with a random delay in EPON.

If the register request is properly received by the OLT, the OLT will issue the Register message to the ONU, followed by a Gate message. The Gate message schedules the ONU to transmit an upstream Register Acknowledgment which completes the new ONU registration. Figure 9.34 depicts the auto-discovery

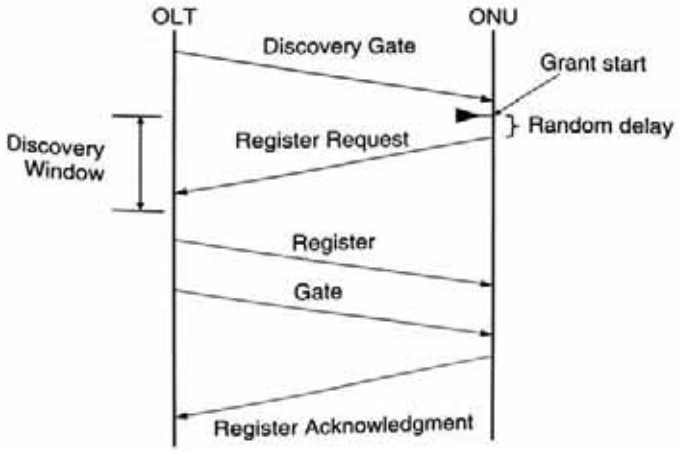


Figure 9.34 Auto-discovery process in EPON (reprinted with permission from IEEE Std. IEEE Std. 802.3, 2005, *Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*, Copyright [2005] by IEEE)

process in EPON. During operation, the ONU and OLT may continuously monitor the fluctuation of RTT due to changes such as temperature fluctuation, and perform fine adjustment by updating the RTT register value.

9.6.3 Point-to-Point Emulation in EPON

In an Ethernet environment, L2 connection is achieved using IEEE 802.1-based bridges [27]. As explained earlier, a bridge performs L2 forwarding function by examining the SA and DA of each received frame. If both of them are connected to bridge through the same port, the bridge filters out the packet without forwarding it. This helps to preserve the bandwidth in other parts of the network and improves the network performance.

In an EPON system, the P2P symmetric Ethernet connectivity is replaced by the asymmetric P2MP connectivity. Because of the directional nature of the remote node, ONUs cannot see each other's upstream traffic directly (Figure 9.35). In a subscriber network, this directional property provides an inherent security advantage. Nevertheless, it also requires the OLT to help forwarding inter-ONU transmissions.

Without any treatment, an IEEE 802.1 bridge connected to the OLT would see all the inter-ONU frames with SA and DA belonging to MAC entities connected to the same bridge port, and would thus determine that they were within the same broadcast domain. As a result, the switch would not forward the traffic between different ONUs connected to the same OLT.

To resolve this issue, a point-to-point emulation (P2PE) function has been created in the RS. The P2PE function maps EPON frames from each ONU to a different virtual MAC in the OLT, which is then connected to a higher layer entity such as L2 switch (Figure 9.36).

The P2PE function is achieved by modifying the preamble in front of the MAC frame to include a logical link ID (LLID) [18, Clause 65]. The modified preamble with the LLID is used in the PON section between the OLT and ONUs. The format of the modified EPON preamble is shown in Figure 9.37. It starts with an SLD

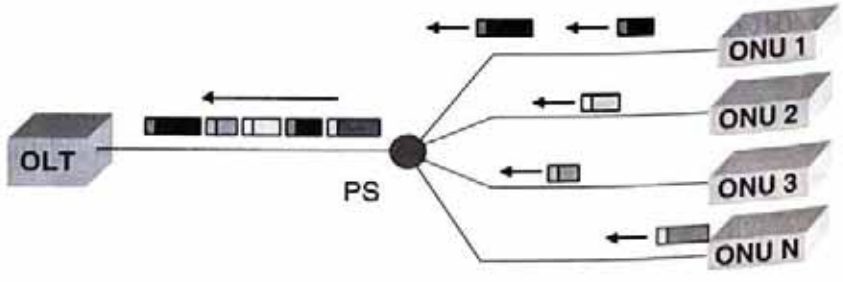


Figure 9.35 Although all the ONU traffic arrives at the same physical port at the OLT, because of the directional power splitting coupler used at the remote node, ONUs cannot see each other's traffic without the forwarding aid of OLT (this figure may be seen in color on the included CD-ROM).

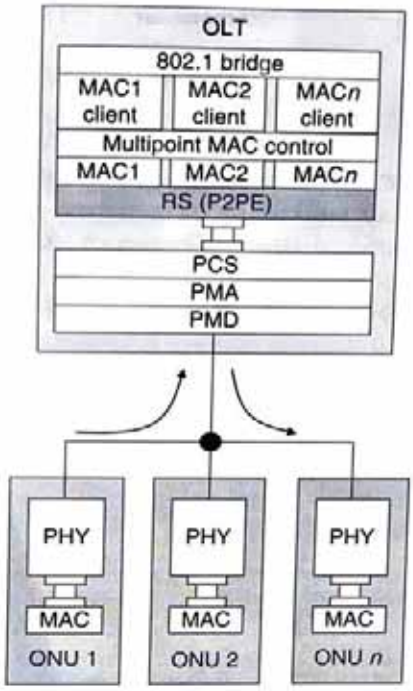


Figure 9.36 Point-to-point emulation in EPON.

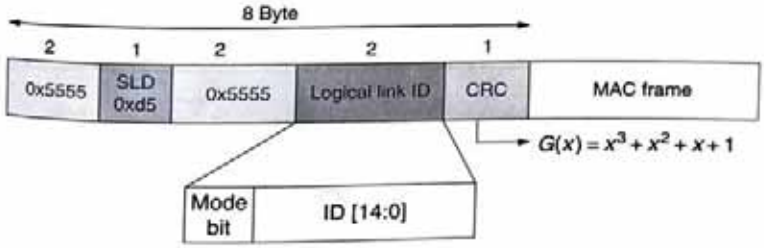


Figure 9.37 Modified preamble with LLID for point-to-point emulation in EPON (this figure may be seen in color on the included CD-ROM).

(start LLID delimiter) field, followed with a two-byte offset and a two-byte LLID. A one-byte CRC field protects the data from the SLD to the LLID inclusive. The first bit of the LLID is a mode bit indicating broadcast or unicast traffic. The rest of the 15 bits are capable of supporting 32 768 different logical ONUs. As mentioned earlier, the actual number of ONUs that can be supported per PON is limited by the power budget. LLIDs are assigned to ONUs at ONU registration time.

The mode bit is set to 0 for P2PE operation. Figure 9.38 shows principle of EPON P2PE. When the mode bit is set to 1, the OLT uses the so-called SCB

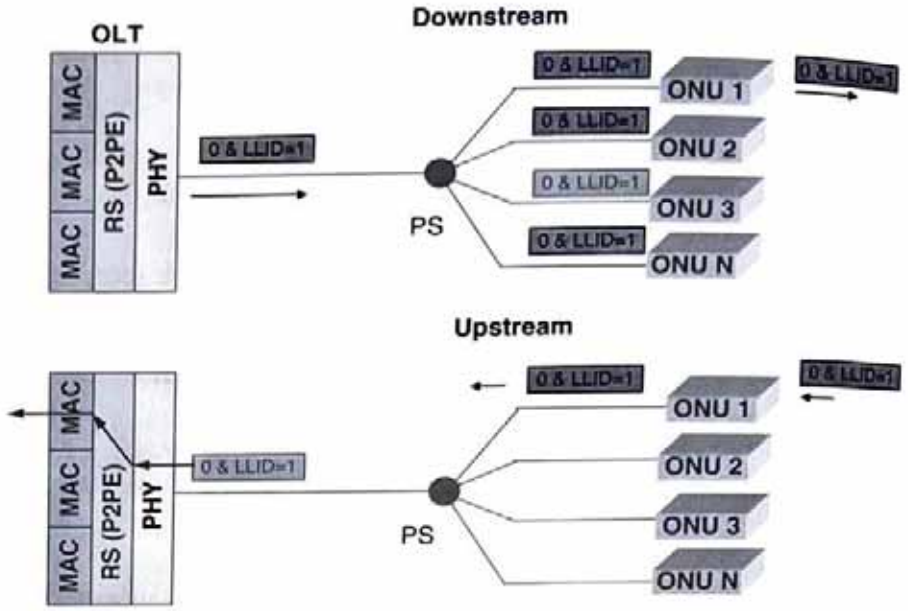


Figure 9.38 EPON point-to-point emulation operation (this figure may be seen in color on the included CD-ROM).

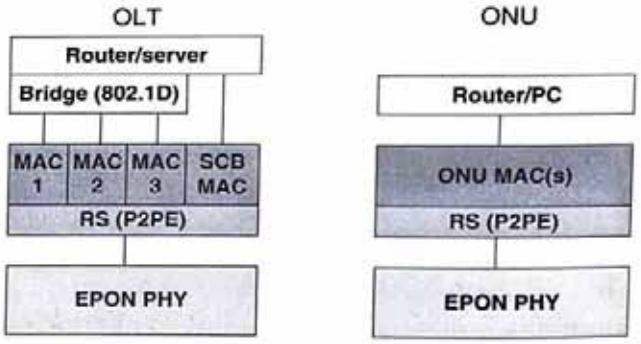


Figure 9.39 Point-to-point and single copy broadcast (SCB) MACs in an EPON model (this figure may be seen in color on the included CD-ROM).

(single-copy broadcast) MAC to broadcast traffic to all ONUs. It takes the advantage of native EPON downstream broadcast operation. To prevent broadcast storm in L2 switches, EPON standard recommends avoiding the connection of the SCB port to 802.1 switches, and use it only to connect to L3 routers[9] or servers for the purpose of disseminating broadcast information. Figure 9.39 illustrates the SCB MAC and emulated P2P MACs in an EPON model.

[9] L2 switches uses STP to ensure no multiple paths exist between two nodes and thus avoid the possibility of forming loops and creating broadcast storms. When both the emulated P2P link and SBC link exist between the OLT and ONU, STP will get confused. Unlike L2 switches, L3 routing protocols can make use of multiple signal paths for load balancing. They can also use the TTL field to avoid loops.

9.6.4 Burst Mode Operation and Loop Timing in EPON

Ethernet protocol is a burst mode protocol. However, modern P2P Ethernet uses dedicated transmitting and receiving paths between a hub and Ethernet workstations. Such a system maintains the clock synchronization between the receiver and transmitter by transmitting idle symbols when there is no data to be sent. Therefore, even though the Ethernet protocol itself is bursty, the physical layer of modern P2P Ethernets is no longer bursty. Although the preamble has been preserved in modern P2P Ethernet, they have no practical significance except for backward compatibility with first-generation Ethernet devices.

Since EPON upstream physical connectivity is bursty, preambles are needed again to help the OLT burst mode receiver to synchronize with the ONU. Moreover, preambles are modified in EPON to carry the LLID used in P2PE [18, Clause 65].

To maintain low cost, traditionally all Ethernet transmitters are running asynchronously on their own local clock domains. There is no global synchronization. A receiver derives the clock signal for gating the received data from its received digital symbols. Mismatches between clock sources are accounted for by adjusting the IFG between Ethernet frames.

In an EPON system, the downstream physical link maintains continuous signal stream and clock synchronization. In the upstream direction, to maintain a common timing reference with the OLT, ONUs use loop timing for the upstream burst mode transmission, i.e., the clock for upstream signal transmission is derived from the downstream received signal.

9.6.5 PCS Layer and Forward Error Correction

EPON defines a symmetric throughput of 1.0 Gbps both in the upstream and downstream directions, and adopted the 8B/10B line PCS coding used in the IEEE 802.3z gigabit Ethernet standard [18, Clause 36]. To take advantage of the low-cost silicon processing capability, EPON has included FEC as an optional choice in the physical layer so that a relaxed optical PMD specification, a higher splitting ratio, or a longer transmission distance can be achieved.

The use of FEC is optional in EPON. The IEEE 802.3ah standard defines RS (255, 239) block codes in the EPON PCS layer [18, Clause 65]. Parity bits are appended at the end of each frame. Since the clock rate does not change when FEC parities are appended, the data throughput is decreased by about 7% when FEC is used. The RS(255, 239) block code does not change the information bits. This allows ONUs which do not support FEC to coexist with ONUs supporting FEC encoded frames. An ONU with no FEC support will simply ignore the parity bits and run running at a higher bit error rate (BER).

9.7 ETHERNET OAM

OAM is an active field of interest and research in Ethernet. One of the charters of the IEEE 802.3ah EFM study group was to specify the Ethernet OAM sublayer functions. The OAM sublayer is situated above the MAC control layer as an optional layer as shown in Fig. 9.2.

The OAM sublayer [18, Clause 57] implements a P2P slow protocol between two interconnected MAC entities using OAMPDUs. Slow protocol PDUs are identified with an Ethernet Type value of 0x88–09 in hexadecimal. To minimize the protocol overhead, slow protocol PDUs are limited to 10 PDUs per second. The formats of OAMPDUs are show next to the OAM layer block diagram in Figure 9.40.

As shown in the figure, the OAM sublayer multiplexes OAMPDUs with data frames from the regular data MAC layer in the transmit path. In the receive path, the OAM sub-layer parses the incoming frames to the OAM client or regular data MAC client. Functions of the OAM layer include:

- OAM capability discovery
- Link monitoring
- Remote loopback
- Remote fault indication

Network management functions such as protection switching, MIB (Management Information Base) read/write, and authentication are not included in the Ethernet OAM layer.

There are two ways for OAM layer to pass protocol information between link partners. A two-octet flag provides quick indication of critical events such as link

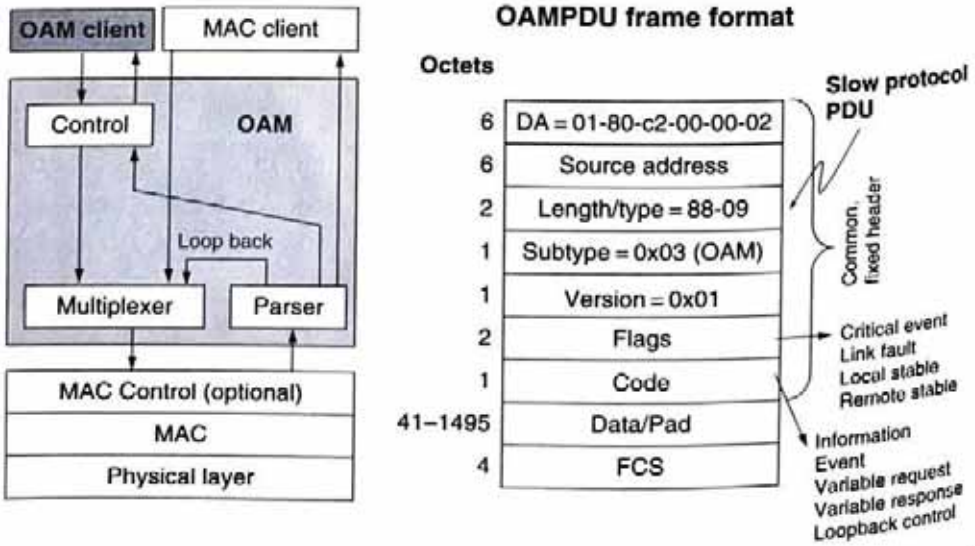


Figure 9.40 OAM sub-layer block diagram and OAMPDU frame format (this figure may be seen in color on the included CD-ROM).

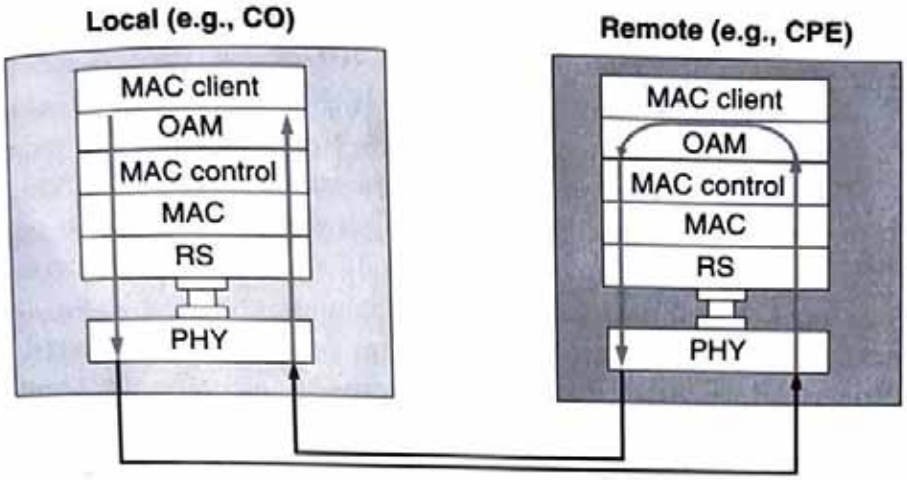


Figure 9.41 OAM loopback function (this figure may be seen in color on the included CD-ROM).

ult, local stable, and remote stable using status bits. The OAM sublayer can also nd event and information to the link partner using OAM variable request and sponse messages.

A very important function implemented in OAM is remote loopback, which shown in Figure 9.41. Figure 9.40 also indicates the loopback path in the AM sublayer. In the case of a network failure, the remote loopback function llows an operator to quickly test the transmission link and narrow down the ault location.

The MEF and ITU are now busy working on service provider OAM (SOAM) unctions for Ethernet services [16, 44], which we cannot cover here because of pace limits.

9.8 LATEST ETHERNET DEVELOPMENTS

9.8.1 10 Gb/s Ethernet Passive Optical Networks

As the human society marches further into the information age, demand for bandwidth keeps increasing. Broadband access has become a norm in industrialized nations. Riding on the tremendous success of the 802.3ah EPON, also called GE-PON, in March 2006, IEEE started the study group on the next generation 10 Gigabit Ethernet passive optical network (10GE-PON) standardization, which became the 802.3av task force [49].

10GE-PON offers 10 Gb/s downstream throughput. Two different speeds for the upstream transmission will be available. In the symmetric design, both the downstream and upstream speeds are 10 Gb/s while 1Gb/s upstream is used in the asymmetric design. This will be the first time that Ethernet systems are designed with asymmetric data throughput in the downstream and upstream directions.

One important consideration in 10GE-PON design is the smooth upgrade from the current base of deployed E-PON or GE-PON. To achieve this, 10GE-PON will

adopt a different wavelength (e.g., 1550 nm C-band has been proposed) for the 10 Gb/s downstream traffic [50]. This 15xx nm 10 Gb/s wavelength is blocked by a wavelength filter at legacy GE-PON ONUs which receive the 1490 nm GE-PON downstream signal. In fact, most of the GE-PON ONUs deployed today already have the blocking filter preinstalled. For those systems without blocking filters, a blocking filter will be installed at ONUs which do not need to be upgraded to 10 Gb/s when the OLT is upgraded to support 10 Gb/s downstream speed. Therefore, Gigabit and 10 Gigabit downstream signals are overlaid in the wavelength domain and extracted with appropriate filters at ONUs. To ensure upstream compatibility with legacy GE-PON, either a separate wavelength is used for the 10 Gb/s upstream signal in the WDM overlay approach, or the OLT receiver will switch between GE and 10GE mode in the time domain automatically (TDM overlay approach), depending on the upstream burst [51].

The higher speeds offered by 10GE-PON can be shared among larger user groups to achieve better economy. This means larger remote node splitting ratios and longer transmission distances. It is envisioned that 10GE-PON may need to support splitting ratios of up to 1:64 or 1:128, and transmission distances of up to 60 km between OLT and ONUs. One goal of the 10GE-PON standard is to achieve an enhanced power budget of 29dB (called Class B++) between the OLT and an ONU [52]. Instead of specifying the remote node splitting ratio, IEEE standardizes the power budget between the OLT and ONU and lets users decide how to use the available power budget for splitting loss, fiber attenuation, and transmission penalties.

In addition to sustaining higher splitting ratios and longer transmission distances, higher power is also required for the higher transmission speed. Theoretically, given everything else is the same, 9.1 dB more received power is required for a 10GE-PON link compared to a GE-PON link, which is 8.24 times faster after accounting for the difference in 8B10B and 64B66B PCS code rates. To achieve the transmission performance of 10GE-PON, many new optical and electronic technologies will be used. First, FEC and APD will be used to improve the receiver sensitivities [53, 54]. Secondly, compared with GE-PON, the dispersion effect increases by 68 times in 10GE-PON. So EDC is being considered to enhance dispersion tolerance and alleviate the dispersion penalties especially for extended reaches [55]. Thirdly, to achieve the power budget requirement, erbium-doped fiber amplifier (EDFA) and SOAs (semiconductor optical amplifiers) are proposed to overcome the signal loss [56, 57].

PON equipment is extremely cost-sensitive. To achieve the best cost structure, EDFA and SOA will be deployed at the OLT instead of distributed at individual ONUs. In the downstream direction, the 10 Gb/s 15xx nm downstream signal will be from a low-cost distributed feedback (DFB) or electroabsorption modulated laser (EML) laser, whose output power is boosted up by an EDFA or SOA to meet the requirement of ONU receivers, which are preferably low-cost and less-sensitive PIN receivers. Use of APD in the downstream direction has not been ruled out either. In the upstream direction, an SOA + APD configuration is used to receive the bursty

upstream signal at the OLT. SOA has the advantage of wideband, very compact in size, and using mass-manufacturable planar technologies. Nevertheless, EDFAs have dominated in the traditional long-haul DWDM market because of their superior noise figure, low polarization dependence, and gain dynamics which results in low interchannel crosstalk. However, EDFAs only work in the C and L bands. Significant advances in SOA have been made in the last several years [58, 59] that practical devices with satisfactory performances are now available. As a matter of fact, the fast SOA carrier dynamics actually makes it better for the bursty PON signals than EDFAs [77]. 10GE-PON upstream signal amplification is an excellent application which could help nurturing the SOA component industry.

Nonlinear effects, which are traditionally seen only in long-haul transmission systems, will now need to be considered in 10 GE-PON. For example, the downstream output power will eventually be limited by SBS (stimulated brillouin scattering) to about 8 to 10 dBm [60] (see also, Chapter 10 on Fiber-Based Broadband Access Technology and Deployment). Another effect is the stimulated Raman scattering (SRS) effect. The original 1490 nm GE downstream signal copropagates with the 1550 nm 10GE-PON downstream signal and serves a Raman pump to the latter [61]. These two wavelengths are separated such that the 1490 nm wavelength forms a quite efficient Raman pump for the C-band 1550 nm wavelength. The strong 10 Gb/s downstream wavelength will deplete the 1490 nm GE-PON downstream signal, causing penalties to this signal, especially when the transmission fiber length is long.

9.8.2 100 Gb/s Ethernet Development

With the continuing growth in broadband access networks and the introduction of higher bandwidth access technologies such as 10-Gigabit Ethernet PONs, backbone capacities also need to scale proportionally. After a few years of deployment, 10-Gigabit Ethernet has now been commoditized in metro and long-haul backbone systems.

Companies such as Google and Yahoo running high-speed ISP backbones and data centers already need links that operate at 10 Gb/s and higher to support their bandwidth-hungry applications. Telecom and MSO carriers also need Ethernet connections with much higher throughput to maintain their fast growing IPTV services [62]. LAG is used to obtain the required throughput with current technologies. As explained before, LAG load balances the traffic across multiple parallel Ethernet links according to higher-layer protocol identifiers using a hashing algorithm. For video streams, which have higher bandwidth granularity and longer connection time, it is more difficult to load balance the traffic. Moreover, LAG requires complicated configuration and management.

Today, more than a 100 companies have expressed interests in participating in the study of higher-speed Ethernet. Traditionally the speed of a new generation Ethernet is always 10 times that of the previous generation. Accordingly the next-generation Ethernet would be 100Gb/s. However, a 40 Gb/s interim standard

has been proposed because of the current technological and economical challenge in 100 Gb/s transmission. Even though 100Gb/s transmission and processing are still a quite a few years away from commercial use and deployment, optical technologies (such as WDM) exist today and 100Gb/s serial data transmission on field optical fibers have been demonstrated [63–65]. High-speed electronics are also available for MAC processing at 100 Gb/s speed [66–68].

Both serial and parallel PHY implementations for 100 Gb/s Ethernet have been suggested at IEEE 802.3 Higher Speed Study Group (HSSG) meetings. The serial PHY approach transmits 100 Gb/s of data on a single wavelength whereas the parallel approach breaks the data into multiple lanes using parallel fibers or wavelengths. For short haul transmission (i.e., a transmission distance shorter than 40 km), parallel PHYs dominate the proposed solutions. For long-haul transmission, there are both serial and parallel PHYs proposed. It should be noted that today's commercial SONET OC-768 systems running at 40 Gb/s serial transmission already requires re-engineering the fiber plant with carefully controlled dispersion maps, as well as new fibers with very low PMD (polarization mode dispersion). Adaptive PMD compensations are required to ensure system availability. Chromatic and polarization mode dispersion effects increase as the square of the data rate. Therefore, 100 Gb/s serial transmission, even though possible, requires extremely demanding component tolerance and very rigorous system tune-up. Consequently, 100 GbE transmissions in metro and long-haul networks should consider not only the transceiver cost at terminals, but also the transmission system infrastructure cost.

Serial 100 GbE transmission has benefited from bandwidth-efficient modulations such as duobinary and DQPSK (differential quadrature phase-shift keying) [64, 69]. For more details of modulation formats, please refer to Chapter 2 by Winzer and Essiambre on "Advanced Optical Modulation Formats". Moreover, FEC and EDC (electronic dispersion compensation, also refer to Chapter 18 (Volume A) by Shanbhag, Yu, and Choma) are used to increase system tolerances to OSNR (optical signal to noise ratio) degradation and dispersion effects.

Parallel 100GbE transmission can easily bundle 10 wavelengths in a 10Gb/s DWDM system. Commercial 10 Gb/s DWDM systems with 100 and 50 GHz channel spacing are mature and readily available. With the current component and technology cost, this approach will offer the fastest time to market and the lowest system cost. However, this brute force approach has a serious problem of offering low spectral efficiency. Figure 9.42 plots the number of 100 Gb/s Ethernet links that can be supported in the C-band of a single-mode fiber versus the transmission spectral efficiency [70]. Therefore, a parallel PHY based on highly spectral efficient 10 G transmission (e.g., with 10 G channel spacing as low as 12.5 GHz) is highly desirable. This approach has the advantage over alternative parallel PHY approaches (4×25 or 5×20 Gb/s) in the fact that the 100 GbE can be transported in many existing 10 G infrastructure without concerns about dispersion map and fiber PMD issues.

Most likely 100 Gb/s Ethernet will be firstly used in data centers for interconnecting high-capacity servers and data switches. We do not expect 100 Gb/s serial interfaces will be economically competitive for the next several years. Photonic IC (PIC)

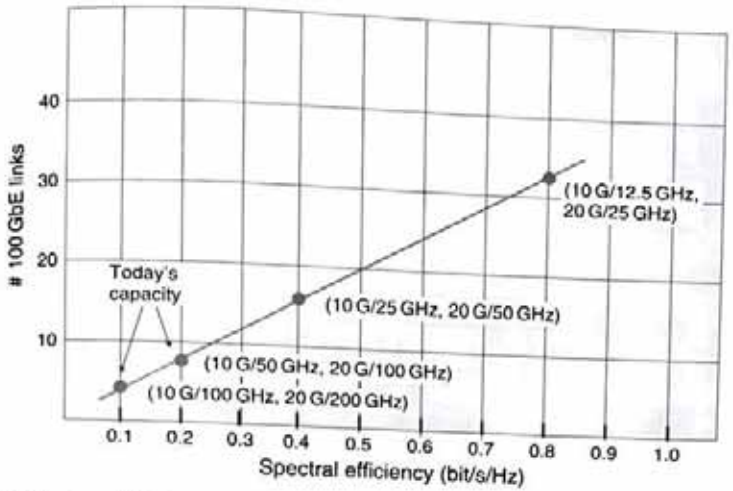


Figure 9.42 Number of 100 Gb/s Ethernet links that can be supported on a single-mode fiber in C-band (this figure may be seen in color on the included CD-ROM).

[78] will be the important technology to realize low-cost and high-density parallel-lane 100 Gb/s Ethernet systems. Both silicon- and InP-based PIC platforms are being considered. The current yield issue to generate InP PICs with channel spacings less than 200 GHz [71] will be a serious road block to their practical use in spectrally efficient 100 GbE systems. Chapter 10 (Volume A) by Welch et al. on "III–V Photonic Circuits and their Impact on Optical Network Architectures" explains PICs in more details. Furthermore, electronic transmission mitigation techniques such as FEC and EDC will be used in 100 Gb/s Ethernet systems. For short distance interconnect applications, spectral efficiency will not be very important so that wavelength spacing can be wide to increase optical component tolerances and reduce system costs. For metro and long distance transmissions, spectral efficiency is very important. So ultra-dense WDM transmission can be the solution to (1) maintain the spectral efficiency and (2) alleviate the fiber plant requirements to support 100 Gb/s Ethernet. It would be most ideal if one can transport 100 Gb/s Ethernet with 10 Gb/s transmission engineering rules and yet maintain the high spectral efficiency [70]. Such systems enable smooth upgrade to 100 Gb/s without fork-lifting upgrades.

9.9 HIGH-SPEED ETHERNET APPLICATION EXAMPLE

9.9.1 An IPTV Example

Recently, video services in the form of IPTV have become the most important high-bandwidth application that drives the growth of broadband networks and high-speed Ethernet. Digital TV signals are much easier to transport than their analog ancestors

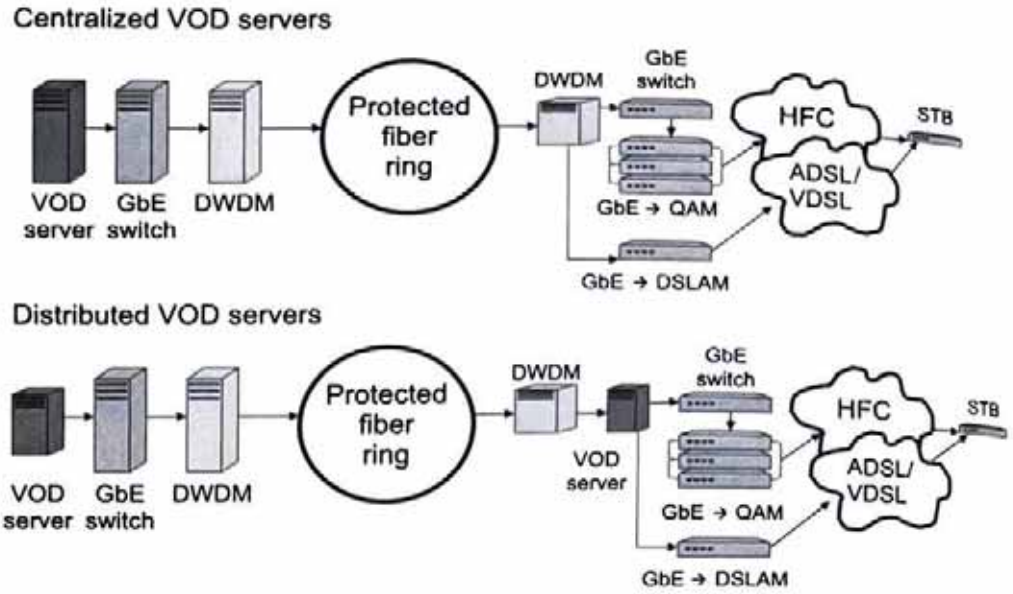


Figure 9.43 VOD delivery network architecture using high-speed Ethernet (this figure may be seen in color on the included CD-ROM).

because of the much lower system linearity and signal-to-noise ratio requirements. As mentioned at the beginning of this chapter, with the new mpeg compressing technology, a Gigabit Ethernet link is capable of carrying 240 streams of standard resolution video signals, each of which requires 3.75 Mb/s bandwidth.

A state-of-the-art approach to implement a VOD network is shown in Figure 9.43, in which traditional SONET wavelengths are replaced by Ethernet wavelengths. Video servers at the head-end node are connected to the WDM optical layer through a Gigabit or 10 Gigabit Ethernet L2/L3 switch [72]. Video signals are transmitted as MPEG over IP packets which are encapsulated in Ethernet frames. At a hub node, an edge quadrature amplitude modulation (QAM) device converts the video signals in Ethernet frames into QAM formats. This architecture not only replaces all the expensive SONET interfaces with low-cost Ethernet interfaces but also improves the network flexibility. With the help of the L2/L3 switch, video signal streams on IP packets can now be arbitrarily switched to any wavelength, and hence any hub node. In effect, the servers at the head end now form a server farm shared by all the hub nodes. The utilization of the expensive video servers is thus improved. Careful readers will realize that this infrastructure for VOD is also the very same infrastructure needed for offering IP data services. Instead of managing separate networks for data and video streaming, carriers now can offer bundled services on a single network with reduced operation and management costs.

9.9.2 Undirectional Ethernet Broadcast

Broadcast is a cost- and bandwidth-efficient way to supply entertainment audio/video services and disseminate information such as stock market quotes and

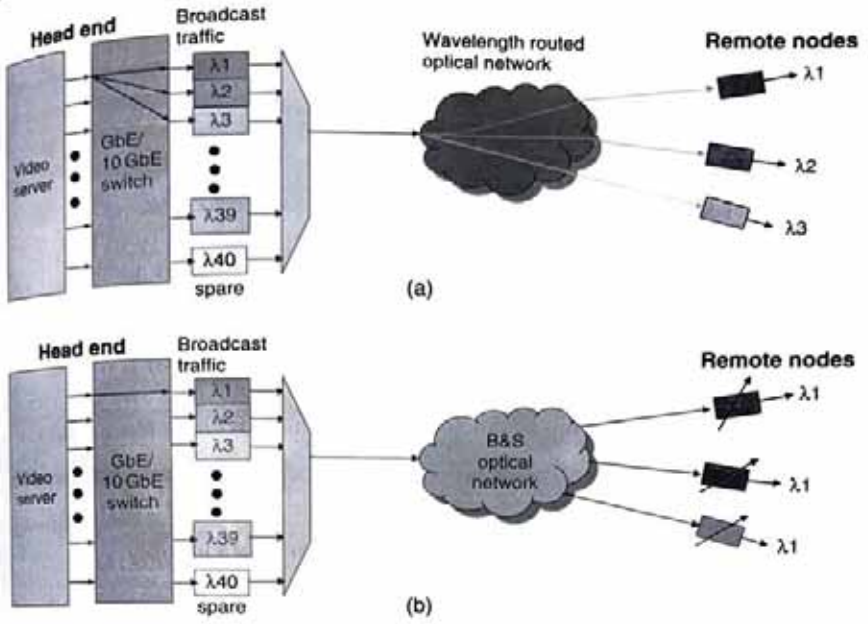


Figure 9.44 Distributing broadcast services in a wavelength-routed network (a) and a broadcast-and-select (B & S) network (b) (this figure may be seen in color on the included CD-ROM).

weather forecast to a large number of users. Broadcast can be easily achieved in the optical layer using optical power splitters.

Figure 9.44(a) shows an example of distributing broadcast video in a traditional wavelength-routed metro WDM network, in which different wavelengths are add/dropped at individual remote nodes. Since each remote node drops a different wavelength, the head-end switch needs to replicate the signal three times using three different wavelength transmitters, which is very inefficient in use of the wavelengths, head-end transmitters, switch ports and switch bandwidths. In a broadcast-and-select network [Figure 9.44(b)] [73], only a single switch port and a single head-end transmitter are required, producing significant cost savings.

Traditional Ethernet connections are always bidirectional. Although by nature, IP routing protocols treats a bidirectional Ethernet connection as two independent unidirectional links, bi-directionality has been inherently assumed in many Ethernet and IP protocols such as STP and ARP (Address Resolution Protocol) [74]. Currently, there is no standard of unidirectional Ethernet interfaces for broadcasting on optical networks. To take the advantage of unidirectional Ethernet broadcast, work is needed to resolve the protocol layer issues such as (1) link discovery, (2) link fault monitoring, reporting and trouble shooting, and (3) SAT and ARP table populating, which are generally easier with bidirectional links. The UDLR (unidirectional link routing) protocol [75], which was originally designed

for satellite networks, solves some of these issues by creating a return tunnel on a separate lower bandwidth link. The physical layer realization of a unidirectional Ethernet link is not difficult because full-duplex Ethernet essentially consists of two independent propagating paths. One just needs to disable the auto-negotiation function and remote fault monitoring on a standard bidirectional Ethernet physical layer. From a capital cost perspective, a unidirectional links saves a laser transmitter at the receiving end and a photo-receiver at the transmitting end.

9.10 CONCLUSION

Ethernet has been firmly established as the technology of choice for building the infrastructure of the information society. To cope with the fast evolving requirements for the rapidly growing Internet, Ethernet is also evolving at a breath-taking speed, with new features and capabilities being proposed and introduced almost every day by many companies, standard bodies and research organizations. Technologies to realize 10 GE-PON and 100 Gigabit Ethernet are now hot items on the active agenda list of the IEEE 802.3 standard group. R&D efforts on Ethernet service and OAM models are solving the issues that carriers are facing in offering Ethernet services. In the very near future, carrier class Ethernet equipment will play a key role in tomorrow's triple- or quadruple-play networks to provide converged services.

LIST OF ACRONYMS

3-R	Reshape, retime, and reamplify
10 GE-PON	10 Gigabit Ethernet passive optical network
APD	Avalanche photodiode
ARP	Address Resolution Protocol
BER	Bit error rate
BPDU	Bridge Protocol Data Unit
CBR	Committed burst rate
CDR	Clock data recovery
CFI	Canonical format indicator
CIR	Committed information rate
CRC	Cyclic redundancy check
CSMA/CD	Carrier sense multiple access with collision detection
C-VLAN	Customer VLAN
DA	Destination address
DFB	Distributed feedBack (laser)
DMUX	Demultiplexer
DWDM	Dense wavelength division multiplexing
EBR	Excess burst rate

issues by creating a return tunnel on a
al layer realization of a unidirectional
fuplex Ethernet essentially consists of
st needs to disable the auto-negotiation
tandard bidirectional Ethernet physical
nidirectional links saves a laser trans-
eiver at the transmitting end.

: technology of choice for building the
To cope with the fast evolving require-
ernet is also evolving at a breath-taking
being proposed and introduced almost
odies and research organizations. Tech-
gabit Ethernet are now hot items on the
ndard group. R&D efforts on Ethernet
ssues that carriers are facing in offering
, carrier class Ethernet equipment will
r quadruple-play networks to provide

mplify
ive optical network

ocol

it

tor
rate
k
ccess with collision detection

aser)

ion multiplexing

EDC	Electronic dispersion compensation
EDFA	Erbium-doped fiber amplifier
EFM	Ethernet for the First Mile
EIR	Excess information rate
EML	Electroabsorption modulated laser
EO	Eletrical–optical
EPON	Ethernet passive optical network
EVC	Ethernet virtual circuit
FCS	Frame Check Sequence
FDDI	Fiber Distributed Data Interface
FEC	Forward error correction
FP	Fabry–Perot
GBIC	Gigabit interface converter
GE-PON	Gigabit Ethernet passive optical network
GFP	Generic framing procedure
GFP-F	Frame-based GFP
GFP-T	Transparent GFP
GMPLS	Generalized multiprotocol label switching
HEC	Header error control
I2C	Inter-IC (bus)
IEEE	Institute of Electrical and Electronics Engineers
IFG	Interframe gap
IP	Internet Protocol
IPTV	Internet Protocol television
ISI	Inter symbol interference
ITU	International Telecommunication Union
LAG	Link aggregation group
LAN	Local area network
LCAS	Link capacity adjustment scheme
LDPC	Low-density parity code
LF	Local fault
LFOS	Local Fault Ordered Set
LLID	Logical link identifier
LMSC	LAN/MAN Standard Committee
LOS	Loss of signal
LSP	Label switched path
MAC	Medium access control
MAN	Metropolitan area networks
MDI	Medium-dependent interface
MEF	Metro Ethernet Forum
MII	Media-independent interface
MiM	MAC-in-MAC
MMF	Multimode fiber
MPCP	Multipoint Control Protocol

age 57 of 214

MPCPDU	Mutlipoint Control Protocol Data Unit
MPEG	Motion Picture Expert Group
MPLS	Multiprotocol label switching
MPMC	Multipoint MAC control
MSA	Multisource agreement
MSO	Multiple service operator (i.e., CATV operator)
MUX	Multiplexer
NNI	Network–network interface
OAM	Operation, administration, and maintenance
OAMPDU	OAM Protcol Data Unit
OE	Optical–electrical
OIF	Optical Internet Forum
OLT	Optical Line Terminal
ONU	Optical network unit
OSI	Open system interconnect
OTN	Optical transport network
P2MP	Point-to-multipoint
P2P	Point-to-point
PAM	Pulse amplitude modulation
PBB	Provider backbone bridge
PCB	Printed circuit board
PCS	Physical coding sublayer
PDU	Protocol Data Unit
PHY	PHYsical Layer
PIC	Photonic IC
PMA	Physical medium attachment
PMD	Physical medium dependent
	Polarization mode dispersion
POS	Packet over SONET
QAM	Quadrature amplitude modulation
QiQ	Q-tag-in-Q-tag
QoS	Quality of Service
RDI	Remote fault indicator
RF	Remote fault
RFOS	Remote Fault Ordered Set
RN	Remote node
RS	Reconciliation sublayer
RS	Reed–Solomon (code)
RSTP	Rapid Spanning Tree Protocol
RSVP-TE	Resource Reservation Protocol-Traffic Engineering
RTT	Round trip time
SA	Source address
SAT	Source address table
SBS	Stimulated Brillouin scattering

SCB	
SDH	
SERDES	
SFD	
SFI-4	
SFP	
SLA	
SLD	
SOA	
SOAM	
SONET	
SPE	
SRS	
STP	
S-VLAN	
TDM	
TTL	
UDLR	
UNI	
UTP	
VCAT	
VID	
VLAN	
VOIP	
WDM	
WIS	
XAUI	
XGMII	
XGXS	

REFERENC

[1] R. M. Metcalf networks," Co
[2] R. Seifert, Gig
[3] R. H. Caro, "E
[4] ISO/IEC 13818
[5] K. G. Coffman *IV B* (I. Kamin
[6] http://grouper.ie
[7] *SFF-8053 Spec* September 200
[8] *INF-8074i Spec* 2001, SFF Com

SCB	Single-copy broadcast
SDH	Synchronous digital hierarchy
SERDES	Serializer–deserializer
SFD	Start frame delimiter
SFI-4	SERDES–framer interface, Release 4
SFP	Small form factor pluggable
SLA	Service level agreement
SLD	Start LLID Delimiter
SOA	Semiconductor optical amplifier
SOAM	Service provider OAM
SONET	Synchronous Optical NETwork
SPE	Synchronous payload envelop
SRS	Stimulated Raman scattering
STP	Spanning Tree Protocol
S-VLAN	Service VLAN
TDM	Time-division multiplexing
TTL	Time to live
UDLR	Unidirectional link routing
UNI	User network interface
UTP	Unshielded twisted pair
VCAT	Virtual conCATenation
VID	VLAN ID
VLAN	Virtual bridged LAN
VOIP	Voice over IP
WDM	Wavelength-division multiplexing
WIS	WAN interface sublayer
XAUI	10 Gigabit Attachment Unit Interface
XGMII	10 Gigabit media-independent interface
XGXS	10 Gigabit extender

REFERENCES

[1] R. M. Metcalfe and D. R. Boggs, "Ethernet: Distributed packet switching for local computer networks," *Communications of the ACM*, 19(7), July 1976.
[2] R. Seifert, *Gigabit Ethernet*, Addison-Wesley, 1998, Reading, Massachusetts.
[3] R. H. Caro, "Ethernet wins over industrial automation," *IEEE Spectrum*, 114, January 2000.
[4] ISO/IEC 13818–5:2005. Also see: http://www.mpeg.org/MPEG/index.html
[5] K. G. Coffman and A. M. Odlyzko, "Growth of the Internet," in *Optical Fiber Telecommuncations IV B* (I. Kaminow and T. Li, eds), Academic Press, 2002, pp. 17–59, San Diego, California.
[6] http://grouper.ieee.org/groups/802
[7] *SFF-8053 Specification for GBIC (Gigabit Interface Converter)*, SFF Committee, Rev. 5.5, September 2000, available from: http://www.schelto.com/t11_2/GBIC/sff-8053.pdf
[8] *INF-8074i Specification for SFP (Small Formfactor Pluggable) Transceiver*, Rev 1.0 May 12, 2001, SFF Committee, available from ftp://ftp.seagate.com/sff/INF-8074.PDF

[9] http://www.xenpak.org/
[10] http://www.xfpmsa.org/
[11] http://www.metroethernetforum.org/
[12] ITU-T G.7041, *Generic framing procedure (GFP)*, August 2005.
[13] ITU-T G.7043, *Virtual Concatenation of Plesiochronous Digital Hierarchy (PDH) signals*, July 2004.
[14] ITU-T G.7042, *Link Capacity Adjustment Scheme (LCAS) for Virtual Concatenated Signals*, March 2006.
[15] ITU-T G.8031, *Ethernet Protection Switching*, June 2006.
[16] ITU-T Y.1731, *OAM Functions and Mechanisms for Ethernet Based Networks*, May 2006.
[17] Optical Internetworking Forum, Document OIF2005.204.17, *User Network Interface (UNI) 2.0 Signaling Specification: Common Part*, June 26 2007.
[18] IEEE Standard 802.3, 2005 Edition, *Carrier Sense Multiple Access with Collision Detection (CSMA/CD) access method and physical layer specifications.*
[19] C. F. Lam, "Beyond Gigabit, Application and Development of High-Speed Ethernet Technology" in *Optical Fiber Telecommuncations IV B* (I. Kaminow and T. Li, eds), Academic Press, 2002, pp. 514–563.
[20] Website address of Small Form Factor (SFF) Committee, http://www.sffcommittee.com/ie/#
[21] A. Ghiasi et al., "Experimental Results of EDC Based Receivers for 2400 ps/ nm at 10.7 Gb/s for Emerging Telecom Standards," *OFC/NFOEC* 2006, paper OTuE3, http://www.asipinc.com/pdf/OFC06_EDC10c.pdf
[22] IEEE 802.3an, Physical Layer and Management Parameters for 10 Gb/s Operation, Type 10GBASE-T, 2006.
[23] ANSI/TIA-TSB-155, "Additional Guidelines For 4-Pair 100 Ω Category 6 Cabling For 10gbase-T Applications," October 2004, http://www.ieee802.org/3/an/public/nov04/TR42.7-04-10-142a-TSB155_d1.1a.pdf
[24] SFF-8431 Specifications for Enhanced 8.5 and 10 Gigabit Small Form Factor Pluggable Module "SFP+", Revision 2.0, 26 April 2007, ftp://ftp.seagate.com/sff/SFF-8431.PDF
[25] OIF-ITLA-MSA-01.1, Integrable Tunable Laser Assembly MSA, Novomber 2005, http://www.oiforum.com/public/documents/OIF-ITLA-MSA-01.1.pdf.
[26] J. Conradi, "Bandwidth-Efficient Modulation Formats for Digital Fiber Transmission Systems," in *Optical Fiber Telecommuncations IV B* (I. Kaminow and T. Li, eds), Academic Press, 2002, pp. 862–901.
[27] IEEE Standard 802.1d, *Media Access Control Bridge*, 1993.
[28] IEEE 802.3as, *Frame Format Extensions*, 2006.
[29] IEEE 802.1w, *IEEE Standard for Local and metropolitan area networks— Common specifications Part 3: Media Access Control (MAC) Bridges — Amendment 2: Rapid Reconfiguration*, 2001.
[30] D. Awduche, "RSVP-TE: Extensions to RSVP for LSP Tunnels," *RFC* 3209, December 2001.
[31] IEEE 802.1Q 2005, *IEEE Standard for Local and Metropolitan Area Networks Virtual Bridged Local Area Networks.*
[32] IEEE 802.1ad, *IEEE Standard for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks Amendment 4: Provider Bridges*, May 2006.
[33] IEEE 802.1ah, *Provider Backbone Bridge*, Draft 3.5, April 2007.
[34] Bell Communications Research – GR253-Core, *SONET Transport Systems: Common Generic Criteria*, Issue 2 Revision 2, January 1999.
[35] ITU-T G.709, *Interfaces for the Optical Transport Network (OTN)*, March 2003.
[36] IETF RFC 1619, Point to Point Protocol (PPP) over SONET/SDH Specification, May 1994.
[37] IETF RFC 1662, PPP in HDLC like framing, July 1994.
[38] IEEE Communications Magazine, (Feature Issue on *Generic Framing Procedure and Data over SONET/SDH and OTN*) May 2002.
[39] V. Saxena, "*Bandwidth drivers 100G Ethernet*," IEEE HSSG, January 2007, available from http://grouper.ieee.org/groups/802/3/hssg/public/jan07/Saxena_01_0107.pdf.

[40] D. Lee, *"Saturating 100G and IT pipes,"* IEEE HSSG, March 2007, available from http://grouper.ieee.org/groups/802/3/hssg/public/mar07/lee_01_0307.pdf.

[41] A. Bechtel, *"A web company's view on Ethernet,"* IEEE HSSG, March 2007, available from http://grouper.ieee.org/groups/802/3/hssg/public/mar07/bechtel_01_0307.pdf.

[42] T. Seely, *"Carrier hurdles to meeting 10GE demand,"* IEEE HSSG, March 2007, available from http://grouper.ieee.org/groups/802/3/hssg/public/mar07/seely_01_0307.pdf.

[43] MEF 10.1, *Ethernet Services Attributes Phase 2*, November 2006, available from http://metroethernetforum.org/PDFs/Standards/MEF10.1.doc

[44] MEF17, *Service OAM Framework and Requirements*, April 2007, available from http://metroethernetforum.org/MEF17.doc.

[45] MEF3, *Circuit Emulation Service Definitions, Framework and Requirements in Metro Ethernet Networks*, April 2004, available from http://metroethernetforum.org/PDFs/Standards/MEF3.doc.

[46] MEF Technical Specification Website: http://metroethernetforum.org/page_loader.php?p_id = 29

[47] N. J. Frigo, "A survey of fiber optics in local access architectures," in *Optical Fiber Telecommunications, III A* (I. Kaminow and T. L. Koch, eds), Academic Press, 1997, pp. 461–522.

[48] A. S. Tanenbaum, *Computer Networks*, 3/e, Prentice Hall, 1996.

[49] IEEE802.3av 10GE-PON website: http://grouper.ieee.org/groups/802/3/av/index.html

[50] S. Tsuji, "Issues for Wavelength Allocation," IEEE802.3av Task Force Meeting, Knoxville, TN, September 18–19, 2006, http://grouper.ieee.org/groups/802/3/av/public/2006_09/3av_0609_tsuji_1.pdf.

[51] G. Kramer, "10G EPON – 1G EPON Coexistence," *IEEE802.3av Task Force Meeting*, Monterey, CA, January 15–17, 2007, http://grouper.ieee.org/groups/802/3/av/public/2007_01/3av_0701_kramer_1.pdf

[52] D-S Lee, "Overview of TX/RX technology for high split EPON systems," *IEEE802.3av Task Force Meeting*, November 13–16, 2006, Dallas, TX, http://grouper.ieee.org/groups/802/3/av/public/2006_11/3av_0611_lee_1.pdf.

[53] F. Chang, "10G EPON Optical Budget Considerations," 10 Gb/s PHY for EPON Study Group Meeting, San Diego, CA, July 17–20, 2006, http://grouper.ieee.org/groups/802/3/10GEPON_study/public/july06/chang_1_0706.pd

[54] M. Takizawa, "Progress and Issues Concerning the Power Budget Discussions in the Power Budget Ad Hoc," *IEEE802.3av Task Force Meeting*, Orlando, FL, March 12–16, 2007, http://grouper.ieee.org/groups/802/3/av/public/2007_03/3av_0703_takizawa_1.pdf

[55] F. Chang, "10G EDC for SMF," *IEEE802.3av Task Force Interim Meeting*, Knoxville, TN, September 18–19, 2006, http://grouper.ieee.org/groups/802/3/av/public/2006_09/3av_0609_chang_1.pdf

[56] L. Spiekman and D. Piehler, "Semiconductor Amplifier for Passive Optical Networks," *IEEE802.3av Task Force Meeting*, November 13–16, 2006, Dallas, TX, http://grouper.ieee.org/groups/802/3/av/public/2006_11/3av_0611_spiekman_1.pdf

[57] P. Doussiere, E. Mao, and W. Jiang, "Technical Feasibility of EDFA based network architecture for 10GEPON," *IEEE802.3av Task Force Meeting*, November 13–16, 2006, Dallas, TX, http://grouper.ieee.org/groups/802/3/av/public/2006_11/3av_0611_doussiere_1.pdf

[58] L. H. Spiekman, J. M. Wiesenfeld, A. H. Gnauck et al., "8 × 10 Gb/s DWDM transmission over 240 km of standard fiber using a cascade of semiconductor optical amplifiers," *IEEE Photon. Technol. Lett.*, 12(8), 1082–1084, August 2000.

[59] K. Morito, S. Tanaka, "Record high saturation power (+22 dBm) and low noise figure (5.7 dB) polarizationinsensitive SOA module," *IEEE Photon. Technol. Lett.*, 17(6), 1298–1300, 2005.

[60] S. Ten, "SBS Degradation of 10 Gb/s Digital Signal in EPON: Experimental and Model," *IEEE 802.3av Task Force Meeting*, Monterey, CA, 15–16 January 2007, http://grouper.ieee.org/groups/802/3/av/public/2007_01/3av_0701_ten_1.pdf

[61] S. Ten and M. Hajduczenia, "Raman-Induced Power Penalty in PONs using 0-order Approximation," *IEEE 802.3av Task Force Meeting*, Monterey, CA, January 15–16, 2007, http://grouper.ieee.org/groups/802/3/av/public/2007_01/3av_0701_ten_2.pdf

[62] "Higher Speed Study Group Call for Interests," *IEEE 802.3 HSSG Meeting Presentation*, San Diego, CA, July 18, 2006, http://grouper.ieee.org/groups/802/3/cfi/0706_1/CFI_01_0706.pdf

Page 61 of 214

[63] J. Peter Winzer, Greg Raybon, and Marcus Duelk, "107-Gb/s optical ETDM transmitter for 100 G Ethernet transport," *ECOC 2005*, Post-deadline paper Th4.1.1, 2005.

[64] P. J. Winzer et al., "2000-km WDM transmission of 10 × 107-Gb/s RZ-DQPSK," *ECOC 2006*, Post-deadline paper Th4.1.3, 2006.

[65] E. Lach and K. Schuh, "Recent Advances in Ultrahigh Bit Rate ETDM Transmission Systems," *J. Lightw. Technol.*, 24(12), 4455–4467, December 2006.

[66] Y. Suzuki, Z. Yamazaki, Y. Amamiya et al., "120-Gb/s multiplexing and 110-Gb/s demultiplexing ICs," *IEEE J. Solid-St. Cir.*, 39(12), 2397–2402, December 2004.

[67] M. Belhadj, "More on the feasibility of a 100 GE MAC," *IEEE 802.3 HSSG Meeting*, Monterey CA, January 17–19, 2007, http://grouper.ieee.org/groups/802/3/hssg/public/jan07/belhadj_01_0107.pdf

[68] F. Shafai, "Technical Feasibility of 100 G Designs," *IEEE802.3 HSSG Meeting*, Ottawa, ON, CA, April 17–19, 2007, http://grouper.ieee.org/groups/802/3/hssg/public/apr07/shafai_01_0407.pdf

[69] M. Duelk and S. Trowbridge, "Serial PHY for Higher-Speed Ethernet," *IEEE802.3 HSSG Meeting*, Knoxville, TN, September 18–21, 2006. http://grouper.ieee.org/groups/802/3/hssg/public/sep06/duelk_01_0906.pdf

[70] W. I. Way, "Spectral-Efficient 100 G Parallel PHY in Metro/Regional Networks," *IEEE802.3 HSSG Meeting*, Monterey, CA, January 17–19, http://grouper.ieee.org/groups/802/3/hssg/public/jan07/way_01_0107.pdf

[71] S. Khodja, "*100 GbE Silicon Photonics Platform Considerations*," *IEEE HSSG*, January 2007, available from http://grouper.ieee.org/groups/802/3/hssg/public/jan07/khodja_01_0107.pdf

[72] C. F. Lam, "Metropolitan Area Ethernet Development, Applications and Management," *OECC 2005*, Seoul, Korea, July 4–8, 2005.

[73] C. F. Lam, "Next Generation Metropolitan Area WDM Optical Networks," *APOC 2006*, paper 6354–42, 2006.

[74] IETF RFC 826, David C. Plummer, An Ethernet Address Resolution Protocol, November 1982.

[75] IETF RFC 3077, E. Duros, et al., Link-Layer Tunneling Mechanism for Unidirectional Links, March 2001.

[76] OIF-SFI4–02.0, *SERDES Framer Interface Level 4 (SFI-4) Phase 2: Implementation Agreement for 10 Gb/s Interface for Physical Layer Device*, http://www.oiforum.com/public/documents/OIF-SFI4–02.0.pdf

[77] B. Stefanov, L. Spiekman and D. Piehler, "Semiconductor Optical Amplifiers – High Power Operation," *IEEE802.3av Task Force Meeting*, Orlando, FL, March 13–15, 2007, http://grouper.ieee.org/groups/802/3/av/public/2007_03/3av_0703_stefanov_1.pdf

[78] R. Nagarajan, et al., "Large-Scale Photonic Integrated Circuits," *IEEE J. Sel. Top. in Quantum Electron.*, 11(1), 5–65, 2005.

10

Fiber-based broadband access technology and deployment

Richard E. Wagner

Corning International, Corning, New York, USA

Abstract

After more than 20 years of research and development, a combination of technological, regulatory, and competitive forces are finally bringing fiber-based broadband access to commercial fruition. Three main approaches, hybrid fiber coax, fiber to the cabinet, and fiber to the home, are each vying for a leading position in the industry, and each has significant future potential to grow customers and increase bandwidth and associated service offerings. Further technical advances and cost reductions will be adopted, eventually bringing performance levels and bandwidth to Gb/s rates when user demand warrants while keeping service costs affordable.

10.1 INTRODUCTION

The use of fiber-optic technology in telecommunications systems has grown over the past 25 years, since its introduction as a transmission media for linking metropolitan central offices in 1980. In the decade after that, long-distance networks deployed fiber-based systems extensively, followed by significant construction of metropolitan interoffice network infrastructure. During that time the reliability and availability of transport systems improved dramatically due largely to the low failure rates of the technology. As a result, researchers and developers dreamed of a time when fiber-optic technology could be economically applied in access networks [1] to replace the copper-based systems extending from central offices to residences and businesses (see Figure 10.1 for a diagram [2] of such scenarios). A perfect example of this visionary dream was articulated by Paul Shumate and Richard Snelling in an *IEEE Communications Magazine* article published in 1989 [3], where they predicted that a fiber-to-the-home (FTTH) solution could be at economic parity with copper-based solutions by 1995 if a carrier were to deploy 1–3 million lines (see Figure 10.2 for a replication of their prediction chart). They went on to explain that it could be possible

Optical Fiber Telecommunications V B: Systems and Networks
Copyright © 2008, Elsevier Inc. All rights reserved.
ISBN: 978-0-12-374172-1

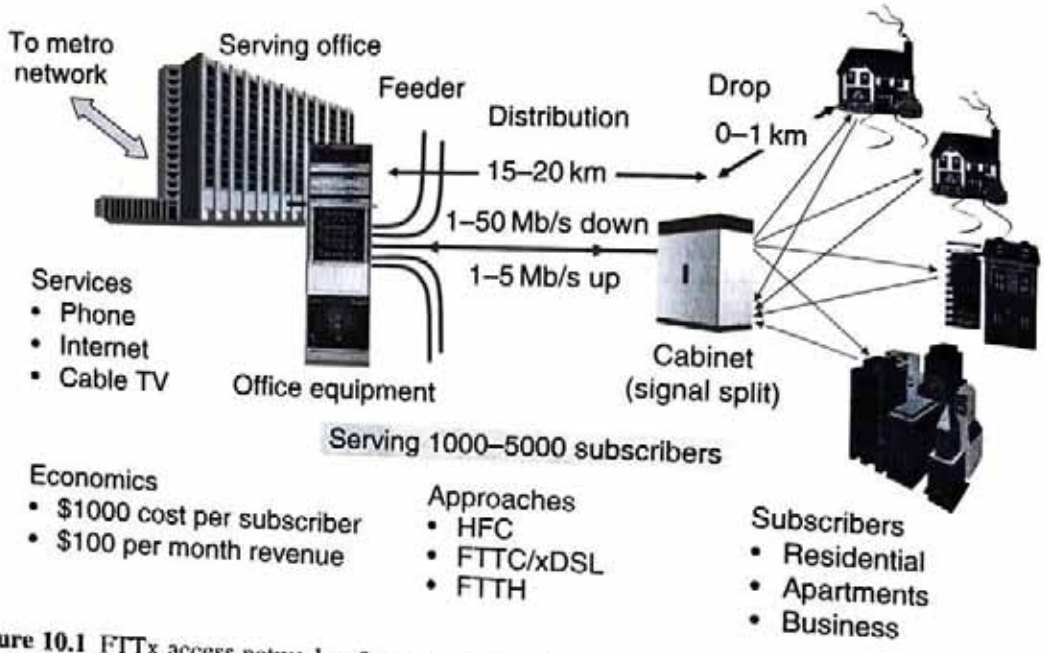


Figure 10.1 FTTx access network reference models (this figure may be seen in color on the included CD-ROM).

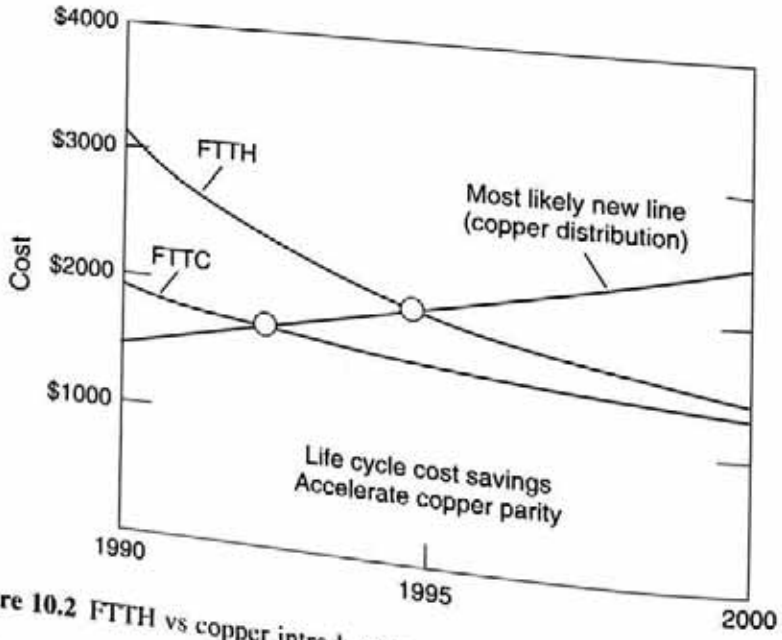


Figure 10.2 FTTH vs copper introduction cross-over date (C. 1991 Shumate).

to convert the entire US network to FTTH over a 25-year period, producing a broadband network for the Information Age. They further noted that there were many details to be worked out, including standardization, network interfaces, powering at the customer end, video capability, and network evolution. This work was pioneering and prophetic, as we now know today.

But what those authors, and others with them, did not foresee was the dramatic effect of the telecommunications consent decree on competition in the United States, the dampening effect of the federal communications commission (FCC) policy allowing competing carriers to lease infrastructure at cost or below, the

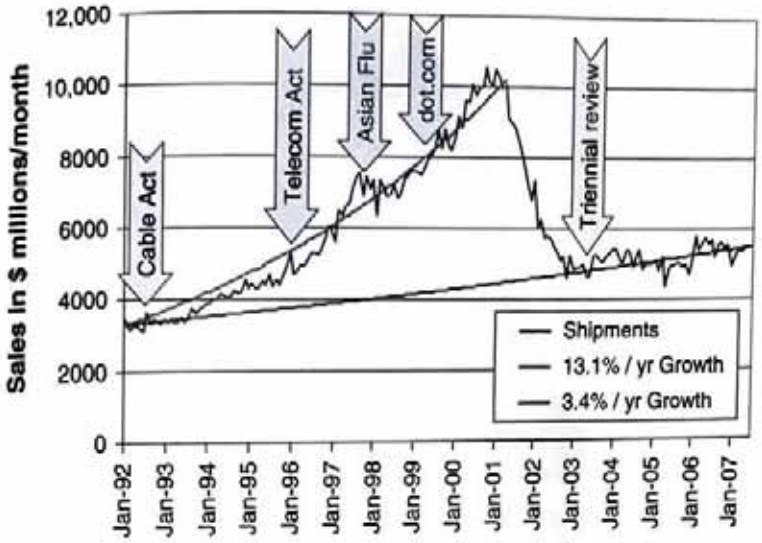


Figure 10.3 Telecom bubble—telecom equipment shipped in the united states vs time (this figure may be seen in color on the included CD-ROM).

delaying effect of the telecom "bubble" bursting [4] (Figure 10.3), and the bandwidth growth driver of the Internet when used for business and entertainment.

Now, in Asia, North America, Europe, and globally, we are seeing significant capital spending on fiber-based broadband access infrastructure [5, 6]. Some of the earliest spending was in Japan, where government funding and initiatives helped to push carriers toward fiber-based access network technologies. Soon after, in the United States, spending was driven by three events: the FCC has made an historic policy decision in 2003 to free fiber-based solutions from competitive regulations, cable TV (CATV) operators have increased telecommunications industry competition by offering high data rate Internet service and telephone service over the Internet, and users have grown accustomed to using the Internet for sharing large digital files containing e-mail messages, photos, music, video, and news clips.

In response to these drivers, the three major local exchange carriers in the United States (Verizon, BellSouth, and SBC) issued a common request for information to broadband access system suppliers, trusting that the resulting volume manufacturing incentives would bring the cost of deploying such systems in line with associated service revenues. Now it appears that this risky move was justified, because in North America we have seen about 8 million fiber-based broadband access lines deployed to homes since then, with more than 1.5 million customers taking the services being offered. Globally there are, by the end of 2007, about 11 million fiber-based access customers. The overall result of this is lowered costs for the technology and its installation, significant technological advancements in the equipment and installation methods, and improved service offerings to customers at affordable rates.

In alignment with this trend, European carriers are now committing funds to fiber-based access networks as well, and in a few years China will ramp up spending for this purpose.

This chapter will focus on the fiber-based approaches to broadband access worldwide, including some of the drivers for deployment, the architectural options, the capital and operational costs, the technological advances, and the future potential of these systems. Three variants of fiber-based broadband access, collectively called FTTx (fiber to the x) in this chapter, have emerged as particularly important. They are hybrid-fiber-coax (HFC) systems, fiber-to-the-cabinet (FTTC) systems, and FTTH systems.

10.2 USER DEMOGRAPHICS

One of the most dramatic and unexpected drivers for fiber-based broadband access has been the explosive growth of the internet and its associated applications. As early as April 1993, when the Mosaic browser became available, the public began to use the internet for transfer of text messages, photos, and data files. Building on that initial application, the number of Internet users has grown to 211 million in the United States alone, and represents about 70% of the population now, while globally there are about 1.2 billion internet user representing about 18% of the world population [7, 8] (Figure 10.4). Not only has the number of users grown, but individual usage has expanded exponentially, until today we routinely exchange music, photos, videos, and software files as large as 10's of megabytes each.

A close look at the bandwidth available to access the Internet shows that it was gated by the technology being offered [9–11] (Figure 10.5), with new ones adopted in cycles of roughly 5–6 years each [2] (Figure 10.6). For example, the earliest users bought phone modems, initially at 9.6 kb/s, then increasing in performance to 14.4 kb/s, 28 kb/s, and finally 56 kb/s [10]. Typically, a group of early users (indicated as Initial Users in Figure 10.6, representing the top 25% of users) adopt the new technology first; the mainstream follows, and finally everybody else (Slow Adopters in Figure 10.6.

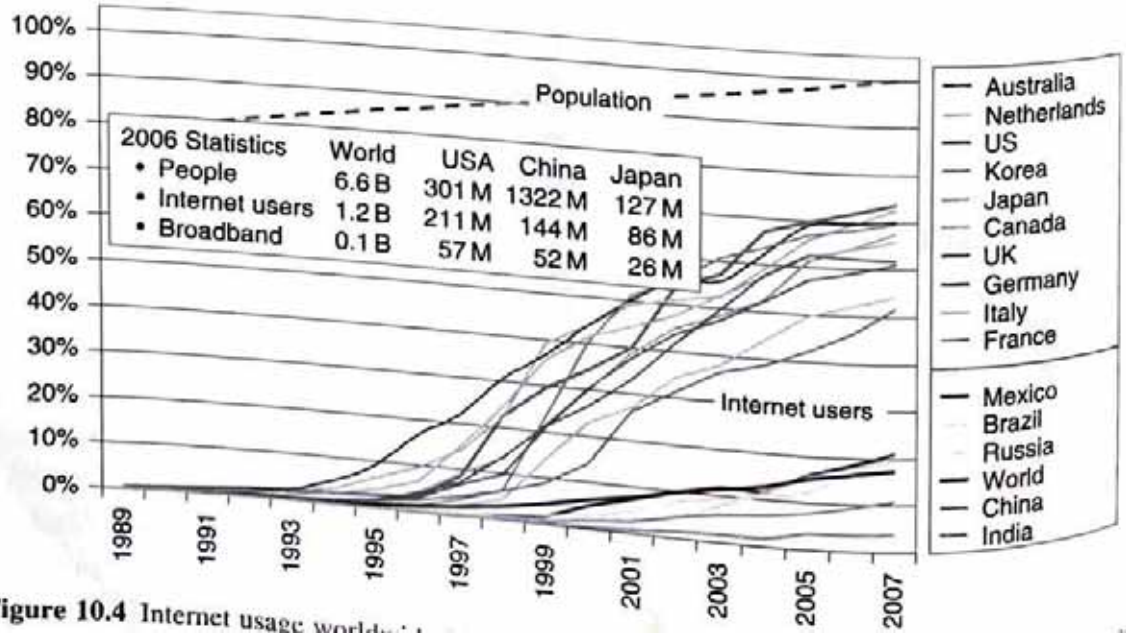


Figure 10.4 Internet usage worldwi

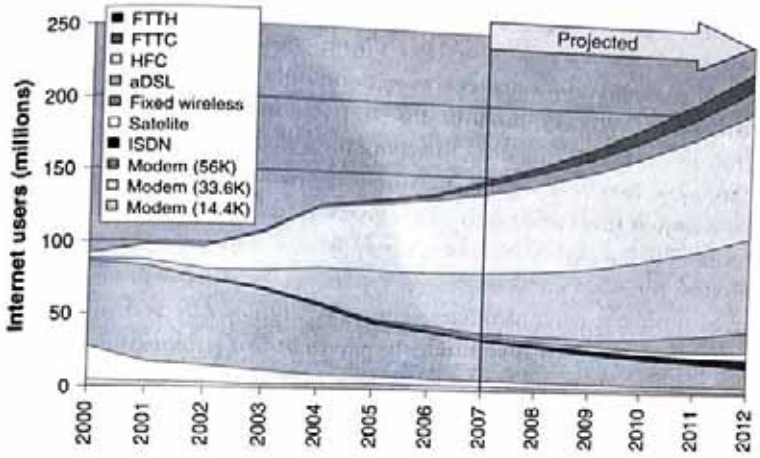


Figure 10.5 US residential access technology adoption over time (this figure may be seen in color on the included CD-ROM).

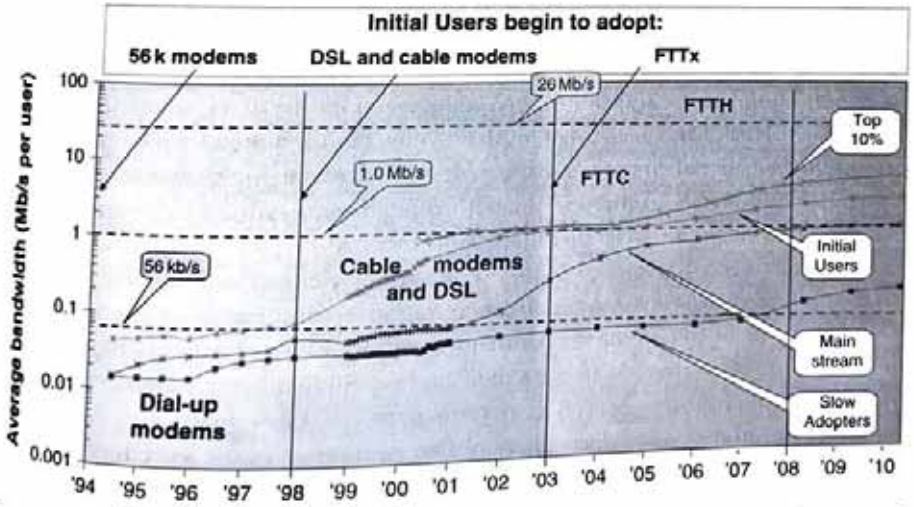


Figure 10.6 US user downstream connection speed trend and projection over time (this figure may be seen in color on the included CD-ROM).

representing the bottom 25% of users) takes advantage of the technology. While the entire adoption cycle takes 8–10 years from introduction to 50% adoption, newer technology is introduced before even half of the users have the old in place. By the time the Slow Adopters begin to use one approach, another newer approach is being offered and adopted by a new set of Initial Users. The cycle of technology adoption by Initial Users, Mainstream Users, and Slow Adopters has experienced two full cycles now, and is into its third cycle: phone modems, then cable modems and asymmetric digital subscriber line (aDSL) modems, and now fiber-based access.

In Figure 10.6, phone modems were introduced in 1993 and by 2001 the overall average user bandwidth exceeded 56 kb/s. Cable modems, introduced in late 1997, were adopted similarly with average user bandwidth exceeding the 1 Mb/s cable modem bandwidth by 2006. Now that FTTH has been introduced, with primary service offerings ranging from 5 to 30 Mb/s, initial users are beginning to subscribe to this technology, especially at 10–15 Mb/s, which will take several years to reach the 25% penetration level and a few more years to achieve the mainstream. We can expect that by 2010 a significant number (perhaps a few million) of those users will be looking for something more, very likely 100 Mb/s service or higher. It is interesting to note that each of these technology cycles has resulted in bandwidth offerings about an order of magnitude larger than the previous, even though the price of the service offerings has only doubled for each cycle.

If we extend this trend into the future, we can expect at least two more cycles of technology adoption by users (Figure 10.7). One of these would offer 100 Mb/s service, projected by this data to have a few million subscribers by 2010, and another higher service offering of 1 Gb/s Ethernet projected to be penetrating the customer base by 2016. The earlier cycle that utilized cable modems and DSL technologies was characterized by asymmetric traffic, in which the downstream signals from the service provider to the user employed a much higher data rate than the upstream signals—for both technological and service reasons—because subscribers were more likely to view content than to generate content. The recent and future cycles, though, need to support nearly symmetric traffic because subscribers desire to share high-resolution images, large files, and video recorder clips with other users—which requires upstream data rates that are similar to downstream rates.

Beyond service offerings of 1 Gb/s with symmetric traffic, it is difficult now to envision what might be the next advance, because completely unimagined applications, spurred by the increased bandwidth offerings, are very likely to arise in the interim. But it is interesting to note that the raw bandwidth of high-definition TV (HDTV) signals for large-screen displays is nearly 10 Gb/s, leading to the expectation that even higher access speeds could be of interest in the very long term.

Against this backdrop of continued bandwidth enhancements stands the much less variable location of the users. For example, residential users in metropolitan

Technology	Per user	Start	50%	
Phone modems	<100 kb/s	1993	2001	History
Cable modems	1 Mb/s	1998	2006	
FTTx approaches	10 Mb/s	2004	2012	
Next generation fiber technology	100 Mb/s	2010	2018	Forecast
Big broadband technology	1 Gb/s	2016	2024	

Figure 10.7 Projection of user demand for bandwidth, showing a Gb/s target eventually (this figure may be seen in color on the included CD-ROM).

993 and by 2001 the overall
ms, introduced in late 1997,
exceeding the 1 Mb/s cable
n introduced, with primary
s are beginning to subscribe
l take several years to reach
ve the mainstream. We can
ew million) of those users
Mb/s service or higher. It is
has resulted in bandwidth
previous, even though the
cycle.
t at least two more cycles
ne of these would offer
w million subscribers by
Ethernet projected to be
cycle that utilized cable
y asymmetric traffic, in
r to the user employed a
r both technological and
to view content than to
, need to support nearly
high-resolution images,
which requires upstream

ffic, it is difficult now to
letely unimagined appli-
very likely to arise in the
th of high-definition TV
s, leading to the expecta-
the very long term.
ements stands the much
al users in metropolitan



target eventually (this figure

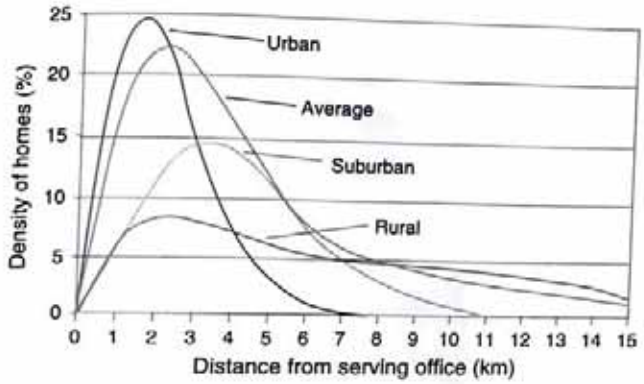


Figure 10.8 Distribution of homes from a serving office in the United States (this figure may be seen in color on the included CD-ROM).

areas in the United States are typically located within 5 km of the central offices that serve them, and in suburban and rural areas this range extends at the greatest to about 10 and 20 km, respectively [12, 13] (Figure 10.8). In Asia and Europe, where the population density is higher, the users are typically closer to the central offices that serve them than they are in the United States. This density of users is not changing much over decades because it is determined by the cultural and socioeconomic norms of suitable living arrangements. So the technologies that carriers deploy need to be capable of serving customers within this rather fixed range, while being able to be upgraded to provide ever higher user bandwidth. In practice, carriers have tended to provide options that offer cost or performance advantages in shorter-reach, higher density metropolitan locations compared to suburban and rural areas. But until recently they have paid little attention to the looming issue of extending the user bandwidth to Gb/s rates.

Ultimately, the introduction of FTTx technologies and their subsequent adoption by mainstream users will have an effect on the traffic load offered to the metropolitan networks that serve the users [14]. This means that the metropolitan interoffice network demands will be advanced by a couple of years compared to a scenario where such technology is not introduced (see Figure 10.9). Such an effect is just now beginning to be felt by service providers as a result of their introduction of FTTx technologies.

10.3 REGULATORY POLICY

Still, with all of these drivers, applications and technology advances available to carriers, they were reluctant to invest in new broadband access infrastructure until recently. In the United States, the 1996 Telecom Act [15] and the regulatory policy of the FCC stipulated that new infrastructure was regulated and had to be offered to competitors at a price set by strict policy rules. This discouraged new network

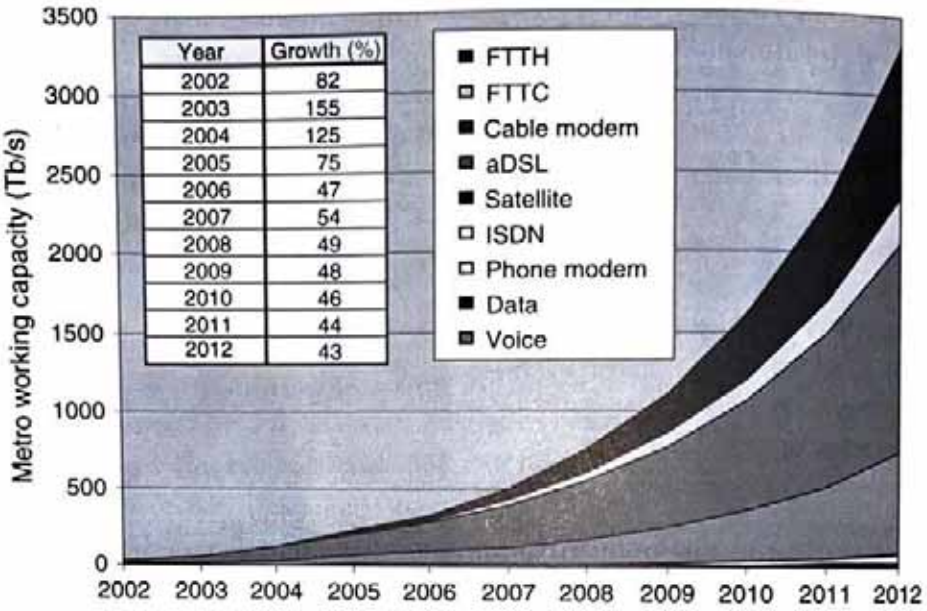


Figure 10.9 US Metro Internet traffic growth over time and in the future, with FTTx factored in (this figure may be seen in color on the included CD-ROM).

builds, because investors feared that their expenditures would be exploited by competitors at costs lower than their own investments would support. Fortunately, this roadblock was removed to a great extent in February 2003, when the FCC adopted new rules for local phone carriers relating to broadband access networks, and issued a notice of proposed rule-making intended to give relief to major carriers of this burden by allowing new FTTx infrastructure to be built without being required to be offered to competitors at regulated prices . This was followed in August 2003 by the actual decree that formalized the FCC policy decision [16]. During this period, SBC, BellSouth, and Verizon issued a common Request For Proposals for FTTx broadband systems, offering encouragement to system houses to design and supply standard FTTx systems. While the initial FCC decision applied to FTTH architectures, follow-up clarifications and decisions related to the 2003 FCC ruling gave adequate relief to some FTTC architectures as well. Somewhat earlier, the Japanese and Korean governments instituted initiatives to help drive their carriers to deploy FTTx technologies, and recently European carriers are finding ways to overcome regulatory issues in the countries they serve. At the same time, municipal governments, independent contractors, and housing builders began to offer FTTx systems as part of their economic development packages. With regulatory relief in the United States, the number of fiber-based access homes passed in North America grew to 8 million in 4 years. In the same time the number of homes connected grew to nearly 1.5 million [17, 18], representing a penetration of about 19% of the homes passed (Figure 10.10). This growth is expected to continue to nearly 40 million homes passed by 2012, with a 44% penetration of homes connected.

Worldwide the number of homes connected has grown in this same period to about 11 million [19–24] and this is expected to grow to 110 million by 2012



e future, with FTTx factored in (this

ures would be exploited by
s would support. Fortunately,
bruary 2003, when the FCC
broadband access networks,
o give relief to major carriers
e built without being required
s followed in August 2003 by
sion [16]. During this period,
uest For Proposals for FTTx
houses to design and supply
n applied to FTTH architec-
the 2003 FCC ruling gave
mewhat earlier, the Japanese
drive their carriers to deploy
e finding ways to overcome
ame time, municipal govern-
gan to offer FTTx systems as
gulatory relief in the United
d in North America grew to
nes connected grew to nearly
ut 19% of the homes passed
to nearly 40 million homes
ected.
rown in this same period to
ow to 110 million by 2012

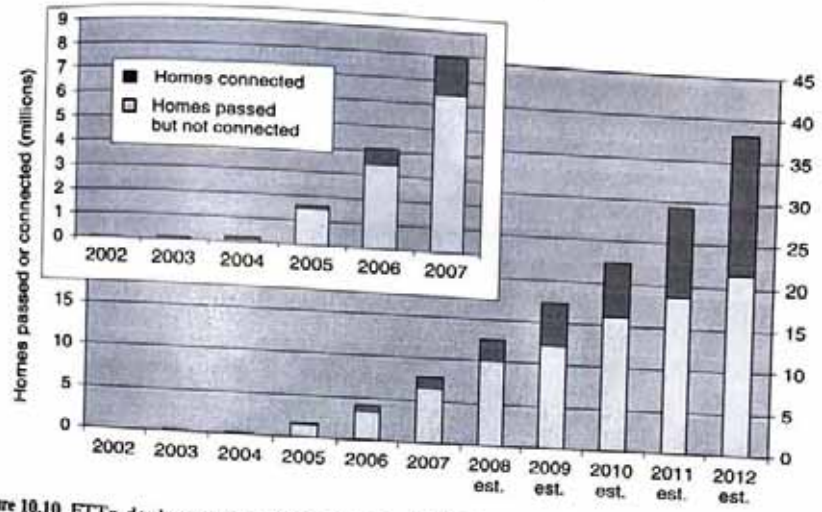


Figure 10.10 FTTx deployments in North America (this figure may be seen in color on the included CD-ROM).

(Figure 10.11). Initially Japan, with its national push, was the leading country to deploy FTTx technology, followed by the United States as a result of the regulatory policy changes that took place there, and then by Europe and finally in a few years by China.

Looking toward the future, there are several global policy issues in active debate today that may impact continued broadband deployment. These include video franchise reform to streamline competitive entry, municipal broadband

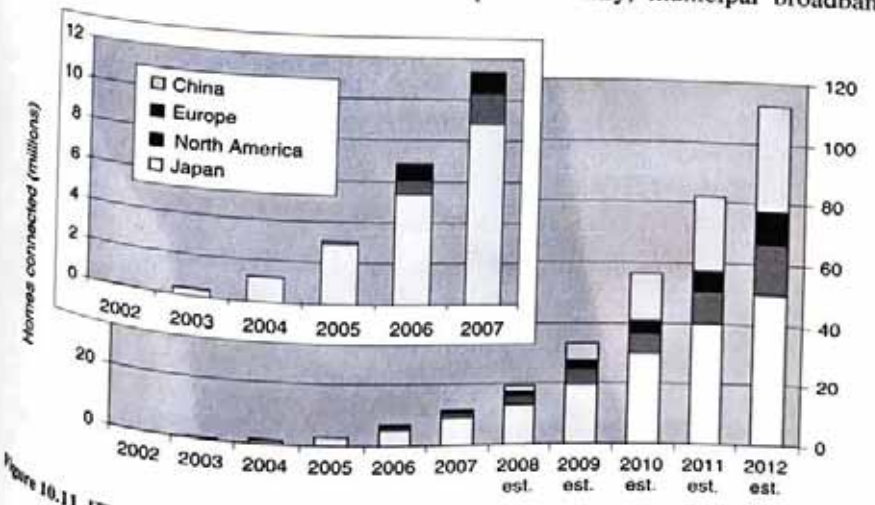


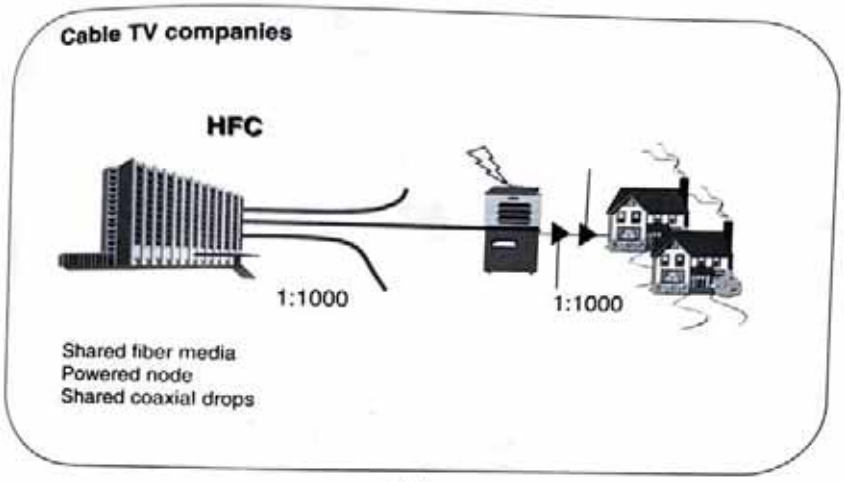
Figure 10.11 FTTx deployments globally (this figure may be seen in color on the included CD-ROM).

deployment allowing local governments to deploy and operate broadband networks, and a universal service fund ensuring fair deployment to all communities regardless of socioeconomic status. It is too early to tell, but the outcome of these policy decisions could encourage further spending by carriers if it is favorable to them.

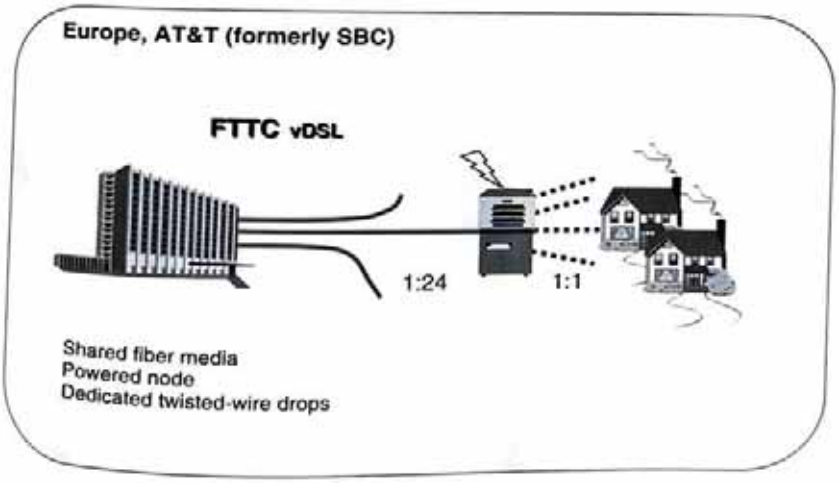
10.4 NETWORK ARCHITECTURES

In the past 15 years, global service providers have gradually pushed fiber technologies farther out from the serving offices and closer to the users. This trend started with the deployment of Subscriber Loop Carrier systems, which simply replaced multiple aging copper lines with fiber transport for the first few km from the serving office. Later, beginning in the early 1990s, CATV operators began to enhance their broadcast TV infrastructure by deploying HFC systems [25], which brought fiber out to within a kilometer or so of the users. Beginning in 1998, BellSouth began to use fiber-to-the-curb systems whenever they needed to refurbish aging copper plant or had new housing start installations, bringing fiber out to within 150 m of the users. These systems are a variant of FTTC systems, with the cabinet placed close to the home at the curb. FTTC is also a hybrid approach, as it uses digital subscriber line (DSL) technology over the existing copper twisted pair infrastructure in the drop portion of the network [26]. In 2002, Japan began deploying FTTH systems, using a point-to-point architecture. Shortly after that, in 2004, Verizon began deploying FTTH systems using a passive optical network (PON) architecture in the USA, pushing fiber all the way to the user. Verizon generalized the approach, calling their system fiber-to-the-premises (FTTP) to allow for the possibility of serving multiple dwelling unit structures (duplex homes and apartments) as well as businesses with the same fiber-based approach. AT&T is now deploying fiber to the node (FTTN) in the United States, a variant of FTTC that brings fiber to within 1.0 km of the user. For the hybrid approaches, longer distances from the termination of the fiber plant to the user have an adverse effect on the bandwidth that can be delivered to the user over copper twisted pairs or coaxial cable.

The network architectures of each of these three variants of fiber-based broadband access are very similar (Figure 10.12). They each have office terminal equipment, large fiber count feeder plant, medium fiber count distribution plant, a drop cable, and customer premises equipment. The three variants differ in a very important respect: the transmission media used for the drop portion of the network. In HFC the drop is coaxial cable, in FTTC the drop is twisted wire pairs, and in FTTH the drop is fiber. They also differ in another important respect: HFC and FTTC require active devices and powering in the outside plant, while FTTH has a totally passive outside plant that requires no power. These variants contribute to differences in the possible user service offerings, as well as differences in the cost structure for deployment and operations.



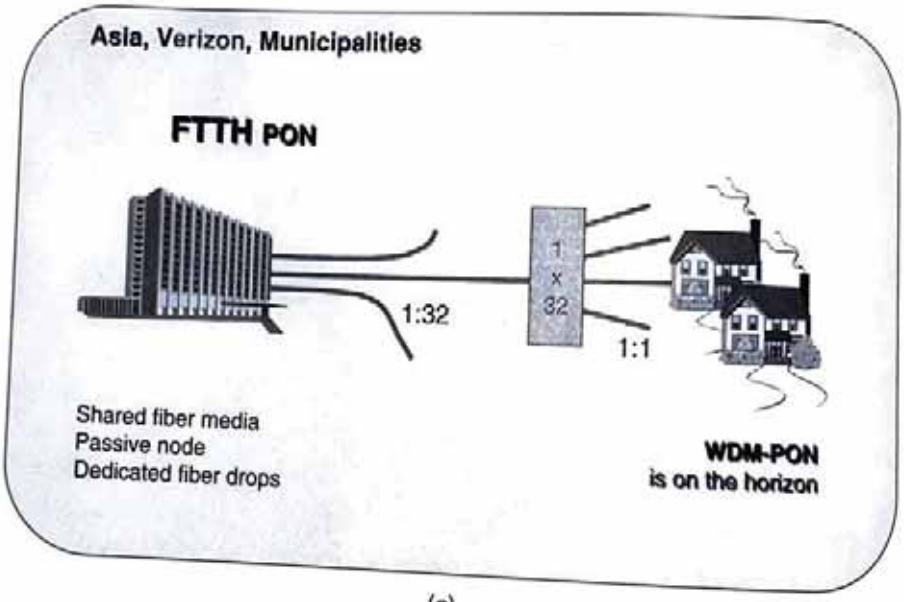
(a)



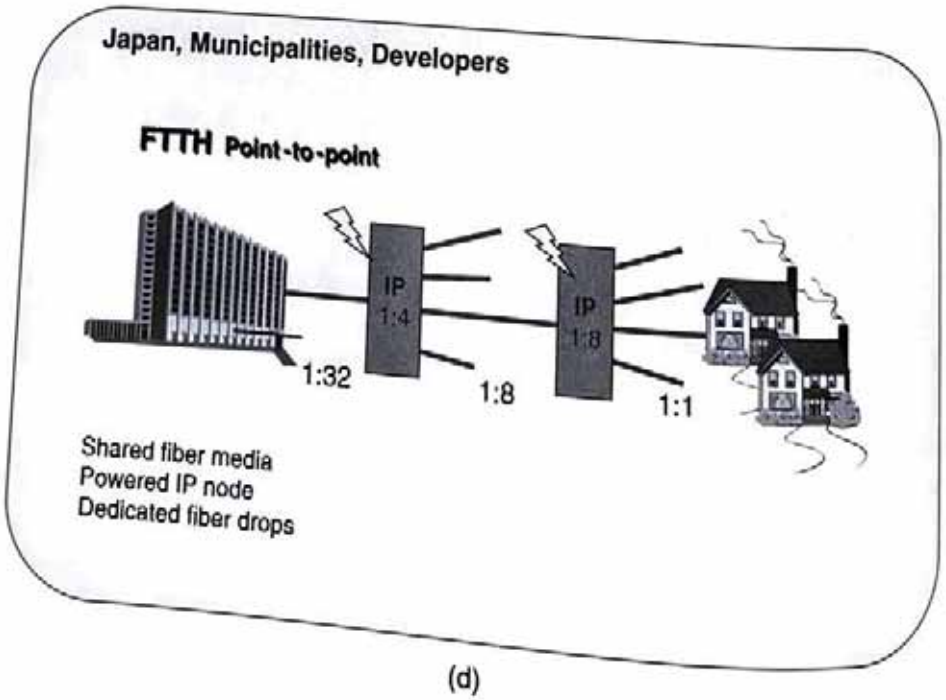
(b)

Figure 10.12 Broadband access network architectures for (a) HFC, (b) FTTC/xDSL, (c) FTTH PON, and (d) FTTH Pt–Pt (this figure may be seen in color on the included CD-ROM).

There are three basic services that each of these architectures is expected to deliver: telephone service, entertainment video service, and high-speed Internet data service. These represent the so-called triple play in FTTx architectures. Since entertainment video has historically been offered using NTSC (National Television System Committee) analog video standards, these systems typically



(c)



(d)

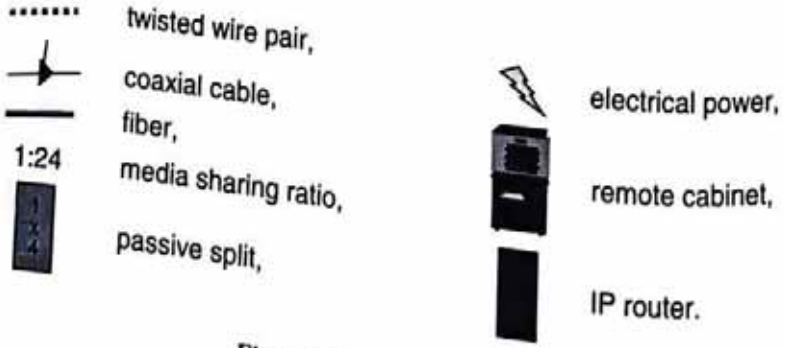


Figure 10.12 (*Continued*)

strive to support both conventional analog video channels as well as digital data channels. In HFC systems, both analog and digital channels are multiplexed onto the same optical carrier; in FTTH systems, the analog and digital signals are multiplexed onto different optical carriers (different wavelengths); and in FTTC systems only digital data channels are supported. Digital TV (DTV) services are normally delivered on the analog channels, while video-on-demand (VoD) is handled by the high-speed data channel. Since FTTC has no analog support, it can not simultaneously carry all of the entertainment channels, but is limited to offering VoD services and a limited number of DTV channels.

FTTC differs from HFC and FTTH in that it must seek to deliver entertainment video services in a digital format. In the long run, this distinction about how entertainment video is supported (either analog or digital) may disappear, since there is a significant trend to shift to digital formats, such as DTV and HDTV. Such a transition to digital television is mandated in the United States by the FCC for over-the-air broadcast TV signals to be completed by February 17, 2009 [27], but for video signals carried in a closed medium (fiber, coaxial cable, twisted pair) there is no equivalent mandated timetable. Other countries have similar timetables for conversion of TV signals to HDTV formats. Still, users are becoming increasingly aware of the advantages of digital video, as they become accustomed to digital video monitors, Video iPODs™, Digital Video Recorders, and other similar visual display appliances. Consequently, it is very likely that by the end of this decade the most common format for video signals will be digital formats that can store and retrieve video content from digital storage devices.

10.4.1 HFC Network Architecture and its Potential

The multiple service operators (MSOs) have deployed HFC networks to more than 40 million customers to date. These architectures include fiber-based transport from a service office to a powered node, which typically supports 500–1000 homes. The modulation format on the transport is subcarrier multiplexed (SCM) analog TV channels on a 1550 nm optical carrier. Since the TV signals are analog in nature, the transmitter must be exceptionally linear to avoid second- and third-order intermodulation impairments among the multiplexed TV channels, and the optical receiver power must be high to avoid noise impairments. This means that high-power erbium-doped fiber amplifier (EDFAs) are used at the serving office, to boost the analog signals to high power and to enable splitting the analog signals to serve many nodes. Every node receives the same set of analog TV channels. In addition, each individual node receives its own channel that is dedicated to Internet service for that node. This means that the unique Internet data must be multiplexed onto the Internet channel for its particular node. Since all users of the node get the same signal, they are all sharing the Internet channel, so none of them can benefit from the full Internet channel bandwidth. At the node, all of the

channels are amplified and split to feed to each subscriber associated with that node. Each subscriber receives the same analog signal as every other subscriber fed by that node, and their cable modem must filter out the TV channels they want to watch and select the Internet packets destined for their account. Cable modems sold today separate the Internet channel from the TV channels and deliver 42 Mb/s of Internet data downstream shared by all users, in compliance with a Data Over Cable Interface Specification (DOCSIS) 2.0 standard. In the upstream direction, Internet data from each subscriber is multiplexed onto a single upstream channel, which in DOCSIS 2.0 is 30 Mb/s shared by all users of the node. The upstream channel is limited by the coaxial amplifiers in the distribution plant, and this is the main limitation of HFC systems for high-speed Internet service.

Over time, there has been a progression of standards for HFC networks, with the DOCSIS cable modem standard being the controlling factor for user bandwidth. The trend has been to increase available downstream user bandwidth, then to increase upstream user bandwidth, then to use more channels for Internet access, and eventually to remove coaxial amplifiers from the distribution plant to allow more channels in the upstream direction. This progression of standardization is intended to increase user bandwidth for Internet service, and that trend is illustrated in Figure 10.13. Note that to achieve such increases, it is necessary to limit the number of customers that a node supports, and correspondingly to move the nodes closer to the user. Eventually, when the drop distances are short enough to remove all amplifiers, and when the node supports 25 users, an HFC network architecture should be able to deliver Gb/s data rate services to fulfill the demands of future users.

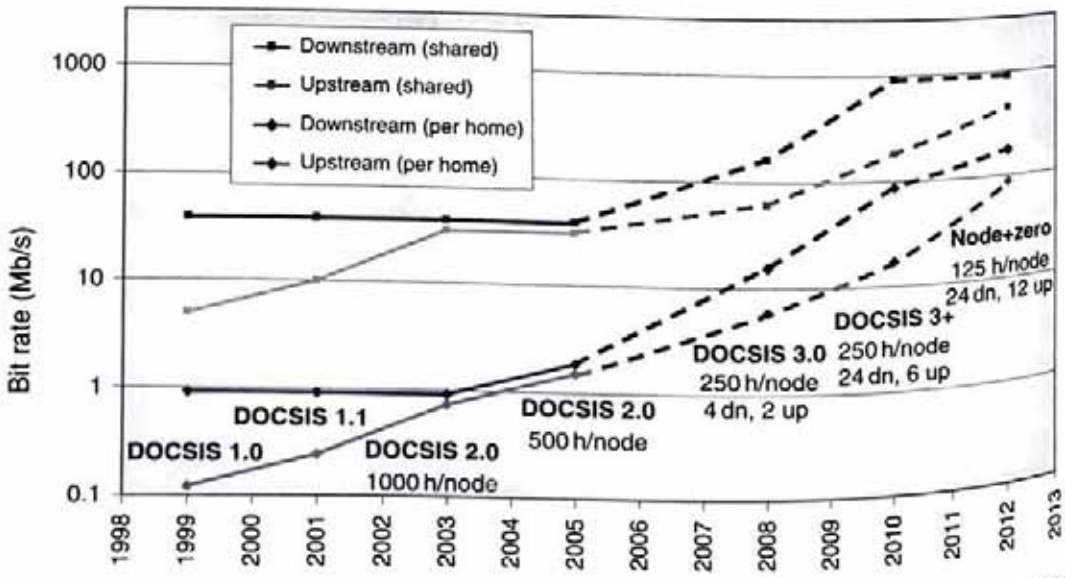
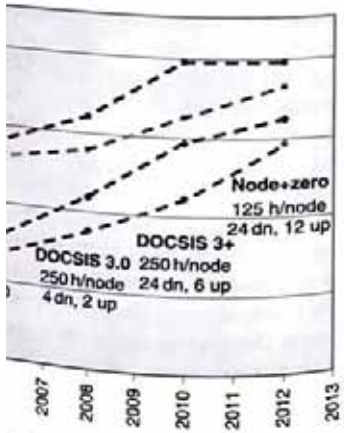


Figure 10.13 Standards progress over time for HFC (CableLabs) (this figure may be seen in color on the included CD-ROM).

ıch subscriber associated with that
ıg signal as every other subscriber
ilter out the TV channels they want
d for their account. Cable modems
ıe TV channels and deliver 42 Mb/s
s, in compliance with a Data Over
ıandard. In the upstream direction,
ed onto a single upstream channel,
ll users of the node. The upstream
he distribution plant, and this is the
l Internet service.

standards for HFC networks, with
controlling factor for user band-
downstream user bandwidth, then
more channels for Internet access,
om the distribution plant to allow
progression of standardization is
net service, and that trend is illu-
h increases, it is necessary to limit
and correspondingly to move the
drop distances are short enough to
pports 25 users, an HFC network
rate services to fulfill the demands



tLabs) (this figure may be seen in color on

10.4.2 FTTC Network Architecture and its Potential

BellSouth has deployed FTTC to more than 1.3 million customers by 2006, AT&T has a plan to deploy FTTN widely to their customers, and BT in the United Kingdom offers this architecture to its customers. These architectures, originally standardized by ANSI (American National Standards Institute) and FSAN (Full Service Access Network Group) with the local exchange carriers input and later standardized by International Telecommunications Union (ITU), use GbE transport from a serving office to a remote cabinet. The optical signal formats and system are simple single-channel digital systems at 1310 nm, with either time-division multiplexed (TDM) or packet multiplexing of individual subscriber's information at the office and corresponding electronic demultiplexing at the remote cabinet in the downstream direction. Normally a single fiber supports the downstream multiplexed signals of 24 subscribers, and at the remote cabinet these are separated and remodulated using DSL formats that can be applied to the individual twisted wire pair drops. A similar arrangement is used in the upstream direction, using a second fiber with electronic multiplexers at the remote cabinet and corresponding demultiplexers at the office.

Over time, the standards and commercial DSL products have evolved to support ever higher user bandwidths. These standards prescribe a discrete multitone (DMT) modulation format, which uses a large number of 4 kHz-spaced orthogonal subcarriers in the band, each modulated at a low rate, which allows the system to cope with severe impairments that may be introduced by the twisted-pair copper medium [28]. Early DSL systems used 1.1 MHz of spectrum on the copper wires to achieve a bit rate of 364 kb/s, but this spectrum use has been increased to 2.2 MHz for aDSL2, then to 12 MHz for very high-speed Digital Subscriber Line (vDSL) and most recently to 30 MHz for vDSL2 technology.

In each case, the noise within the usable spectrum is the limiting transmission impairment, and as the bandwidth increases the total noise increases correspondingly. The noise limitation translates to a distance limit—the distance for which the signal-to-noise ratio produces satisfactory error rates decreases as the usable bandwidth increases. So DSL using 1.1 MHz of spectrum can support drop distances of 2.5 km, while the latest vDSL2 standard using 30 MHz of spectrum supports drop distances of only 300 m. Consequently, the FTTN system of AT&T, with drop distance of 1.0 km can probably support a user bandwidth of about 25 Mb/s, while the FTTC system of BellSouth (now known as AT&T Southeast), with a drop distance of 300 m can probably support a vDSL2 user bandwidth of about 100 Mb/s.

A summary of the standards evolution is depicted in Figure 10.14. In the future, if DSL standards should move to an even higher utilization of copper spectrum of 50 MHz, and if service providers are willing to dedicate two twisted pairs per user and maintain drop distances less than 150 m, it is conceivable that DSL technology can support 1 Gb/s data rate services to individual users.

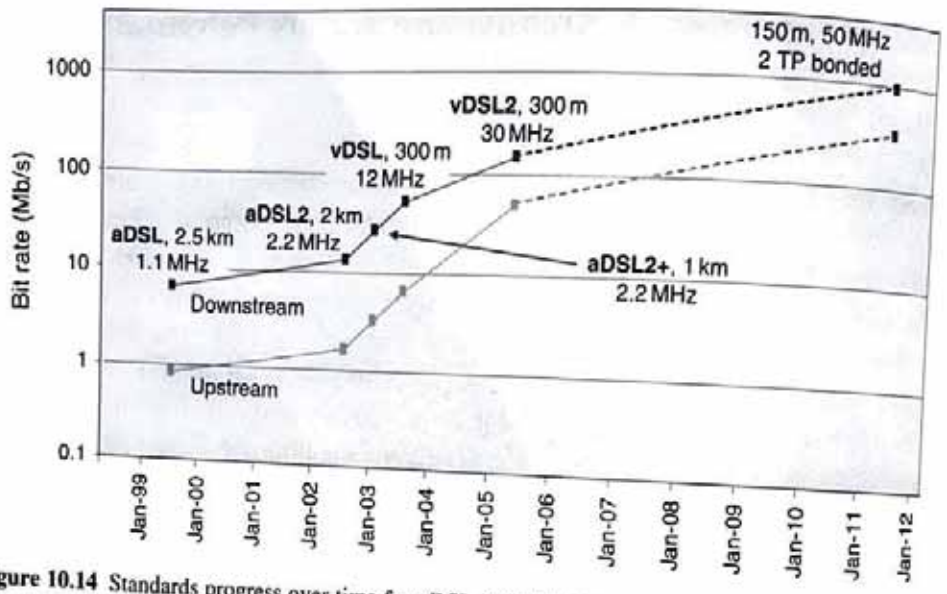
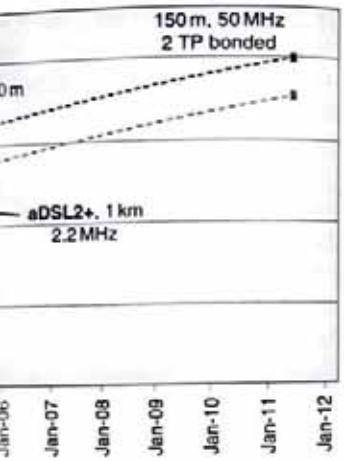


Figure 10.14 Standards progress over time for xDSL (ITU-T) (this figure may be seen in color on the included CD-ROM).

10.4.3 FTTH Network Architecture and its Potential

Verizon began deploying FTTH (they use the term FTTP to include homes, multidwelling units, and businesses) in May 2004 and by July 2007 had passed more than 7 million homes and had an estimated 1.5 million subscribers. Carriers in Asia and Europe, as well as many municipalities globally deploy this architecture as well. The FTTH systems use a PON architecture, transporting common signals from a serving office to multiple users with a 1:32 optical power split at a passive cabinet [29], and then a fiber drop to a network interface unit on the outside of the house. The analog and digital signals are carried on different wavelengths, with the downstream analog signals at 1550 nm and the downstream digital signals at 1490 nm. Remember, though, that analog channels can carry DTV signals via quadrature amplitude modulation. Upstream signals are carried in the same fiber as the downstream signals at a wavelength of 1310 nm, and are coupled to the fiber through coupling filters at each end of the network. The upstream data signals are multiplexed together using time-division multiple access (TDMA) methods, and each user is assigned one or more unique timeslots. A key problem in PON networks is that each user's upstream signal will arrive at a variable time determined by the distance the user is from the serving office and the transit time of the signal in fiber, potentially causing signal contention at the point where the upstream TDMA signals are to be multiplexed. This is handled by an autoranging synchronization signal that prompts each user when to transmit so that their signal arrives at the multiplex point in the correct timeslot.

Initially, the PON standards addressed FTTH systems with a downstream data rate of 622 Mb/s and an upstream data rate of



T) (this figure may be seen in color on the

and its Potential

e term FTTP to include homes. 004 and by July 2007 had passed d 1.5 million subscribers. Carriers ities globally deploy this architec- rchitecture, transporting common vith a 1:32 optical power split at a a network interface unit on the signals are carried on different ls at 1550 nm and the downstream , that analog channels can carry on. Upstream signals are carried in wavelength of 1310 nm, and are at each end of the network. The sing time-division multiple access e or more unique timeslots. A key upstream signal will arrive at a r is from the serving office and the sing signal contention at the point ultiplexed. This is handled by an each user when to transmit so that correct timeslot.

l systems with a downstream data 155 Mb/s using the asynchronous

transfer mode (ATM) format (A-PON), but later this was enhanced to include the analog channel for broadband access (B-PON). Eventually the ATM approach was extended to 2.5 Gb/s downstream and 622 Mb/s upstream (G-PON), and another approach using GbEthernet without an analog overlay was standardized as well (GE-PON). With the ATM format, the user has a well-defined and guaranteed bandwidth and quality of service, while with the Ethernet format the user shares the full bandwidth on a best-effort basis. The ATM-based PON systems are favored by US carriers, since the ATM format is compliant with their legacy transport systems, while the Ethernet format requires other capital-intensive changes to how they build their legacy networks. The user bandwidth characteristics of these various options for FTTH systems are summarized in Figure 10.15, and it is clear that user bandwidths of 100 Mb/s are quite reasonable today, and these can very likely move to Gb/s data rates [30] by making use of the statistical multiplexing inherent in Ethernet protocols.

While each of the FTTx architectures can be extended to provide significantly more bandwidth per user, even up to Gb/s rates per user, inside the home there is still the issue of how to distribute those signals to multiple rooms. The FTTH and FTTC systems provide the user an interface at the side of the home and the HFC systems provide a cable modem interface inside the home, but from a user perspective the signals must reach each PC, Internet appliance, display, and entertainment system in the household for the bandwidth to be useful. In Japan and in China, the FTTH system are installed with the user interface inside the home, allowing this interface to have substantially less environmental stress from temperature and the elements, and offering the possibility of tailoring the interface to the specific needs of the user. In-home networking has been identified as a key

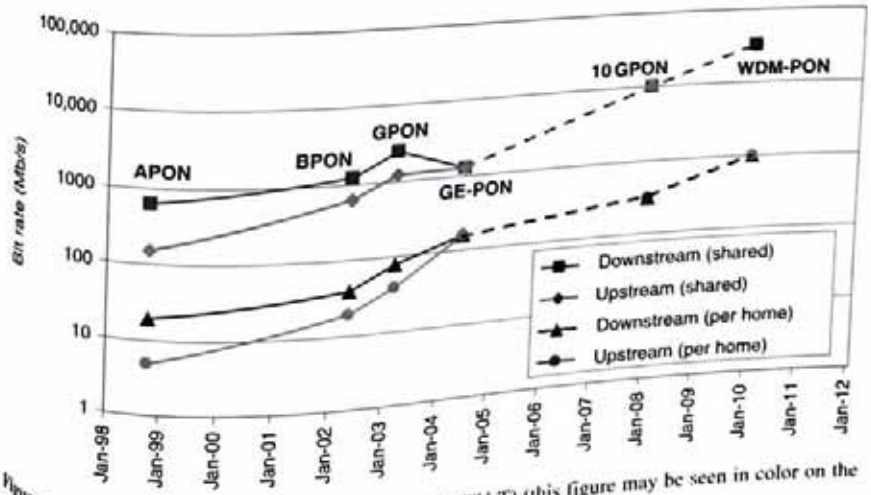


Figure 10.15 Standards progress over time for PON (ITU-T) (this figure may be seen in color on the included CD-ROM).

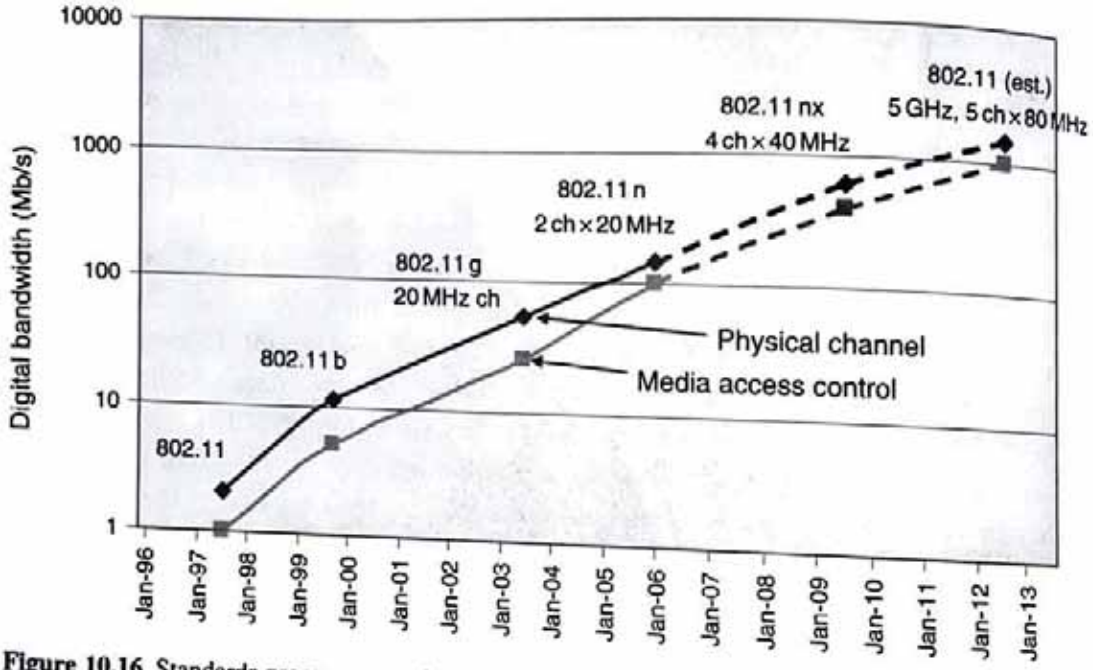
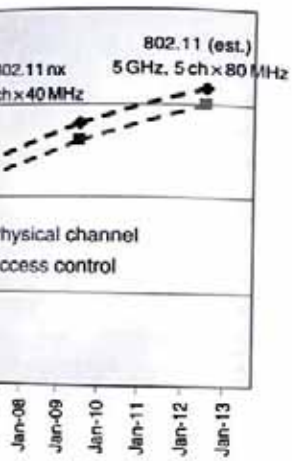


Figure 10.16 Standards progress over time for Wi-Fi (IEEE) (this figure may be seen in color on the included CD-ROM).

problem in the recent introduction of FTTP in the United States and Japan, and has been a lingering issue for HFC installations with multiple PCs. While this is not strictly the carriers issue to address, they will increasingly be in the position to suggest solutions to the homeowner in order to gain the maximum adoption of their service offerings. Fortunately, each of these system interfaces can be equipped with wireless (Wi-Fi, wireless-fidelity) base stations which today are capable of distributing several Mb/s of data anywhere inside a typical home. But as FTTx bandwidth increases over time, the demands of in-home networking capability will grow correspondingly to bandwidths of Gb/s. While the standards activity for Wi-Fi is on track to keep pace with the introduction of FTTx technologies (Figure 10.16), another option is to deploy specially designed fiber [31] in structured cabling arrangements for in-home networks.

10.5 CAPITAL INVESTMENT

While any one of these FTTx approaches can satisfy the current demands and affordability of the current market, they each require substantial capital investment to deploy widely. By way of example, imagine that the estimated 110 million homes is served by fiber-based access in 2012 as shown in Figure 10.11, that this represents a penetration of homes passed of 30%, and that each home passed could be provided for as little as $1000 per household. This would require about $370 billion in cumulative capital expenditures over a 5-year period. This represents a combined annual capital spending for fiber-based access related systems of about



gure may be seen in color on the

l States and Japan, and has
ple PCs. While this is not
ngly be in the position to
he maximum adoption of
ystem interfaces can be
stations which today are
ide a typical home. But as
in-home networking cap-
b/s. While the standards
duction of FTTx technol-
lly designed fiber [31] in

the current demands and
tantial capital investment
he estimated 110 million
in Figure 10.11, that this
each home passed could
would require about $370
period. This represents a
related systems of about

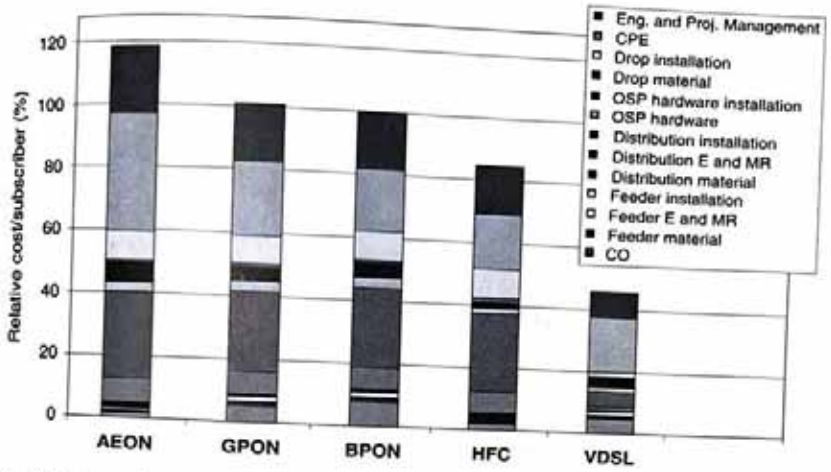


Figure 10.17 Comparison of relative capital costs per subscriber of HFC, FTTC/xDSL, and FTTH options, with B-PON arbitrarily chosen as a reference (this figure may be seen in color on the included CD-ROM).

$75 billion by all carriers globally. But 110 million homes is only about 5% of those available in the developing countries, so it will require a 20-year or more infrastructure commitment to bring fiber to all homes worldwide. A comparison of the FTTx approaches indicates that $1000 per home is a target that is eventually achievable, with the more expensive approaches providing more capability in the long run.

A comparative assessment of the capital costs for FTTx systems is illustrated in Figure 10.17, where it is assumed that 100% of homes passed are taking service, that there is a mix of 70% aerial and 30% buried plant, that the density of users is typical of a combination of urban and suburban customers, and that all of the equipment, installation, project, and subscriber costs are included in the comparison and averaged over all customers. At the lower end, FTTC/vDSL requires the smallest capital expenditures because it capitalizes on the existing telephone copper wire drop plant, but it also provides the most restrictive service options because it does not support analog TV programming. In the mid-range, HFC capital expenditures are higher than for FTTC/vDSL, and the extra expenditure buys the capability for analog TV programming, although HFC systems offer shared and therefore somewhat limited bandwidth for high-speed Internet services. At the upper end of capital expenditures are the FTTH systems, which provide for analog TV services as well as the highest potential digital Internet bandwidth. Because the MSO and ILECs (Incumbent Local Exchange Carriers) already have coaxial cable and twisted-pair drop cables in place, those two options represent lower spending and medium-term risk positions for satisfying future bandwidth demand growth. On the other hand, FTTH requires new infrastructure all the way to the home, making it the most capital intensive but also offering a better longer-term risk for satisfying future growing bandwidth demands.

10.6 OPERATIONAL SAVINGS

Given that FTTx deployments consume so much capital, a carrier is bound to ask: "Why would it benefit my company to take the risk to offer such services?" Of course, the answer is twofold: there are new service-related revenues to reap, and there are also operational cost savings associated with fiber-based systems. Taking a lesson from long-haul and metro transport systems, carriers have learned that fiber systems have fewer failures, higher availability, are more reliable [32], and have more capacity than copper-based systems. So FTTx solutions have a strong historical track-record on which to rely for both good reliability and high bandwidth at long distances. In addition, passive plant is especially attractive, because the number of active devices can be made much smaller in the outside plant where the parts are less accessible, and because passive devices in the outside plant are highly reliable [33]. For this reason alone, fiber-based systems are more attractive than copper-based systems that require periodic amplification and signal shaping. Further, new operational savings can be built into new infrastructure that take into account the ability to use sophisticated computer algorithms to enable faster and more efficient service provisioning, churn, administration, and easier fault location. All of these network operations and "back office" functions add up to significant annual operational savings compared to the way things are done today.

As an example, a comparison between FTTH annual operations expenditures relative to today's copper-based wireline technology [34, 35] is illustrated in Figure 10.18. While details of the customer contact and billing, central office, outside plant and network operations costs have been analyzed for both FTTH and today's technology [34], only the overall operations costs have been estimated for HFC and FTTC solutions [35]. This information indicates that more than $150 per year per subscriber can be saved in operational expenditures for FTTH compared to the labor-intensive wireline operations activities associated with customer

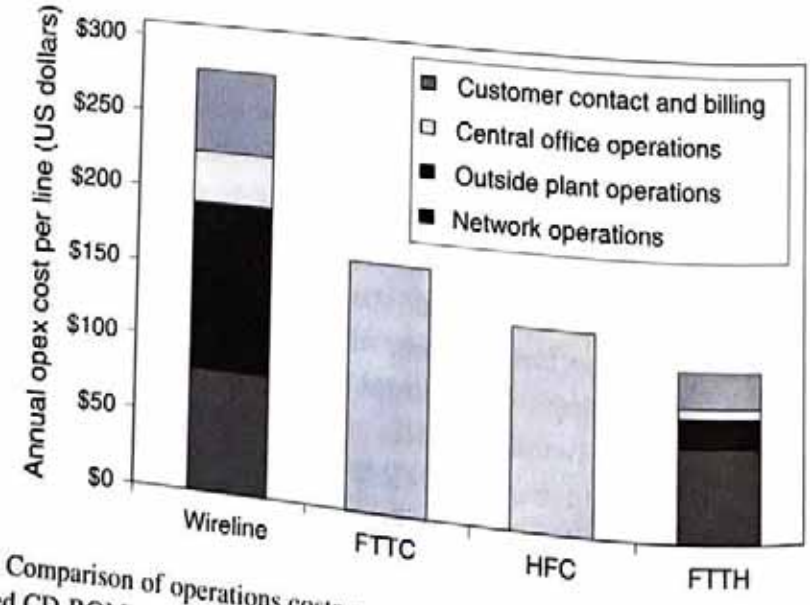


Figure 10.18 Comparison of operations costs per subscriber for FIT (this figure may be seen in color on the included CD-ROM).

service requests, as well as provisioning, terminating, and re-provisioning service. Over a 7-year period, this nearly pays for the installation expenses, without even taking service revenue into account. A similar, but less dramatic, scenario is the case for HFC and FTTC solutions as well, but their reliance on active plant prevents them from reaping the benefits of a fully passive outside plant.

10.7 TECHNOLOGICAL ADVANCEMENTS

Over the last decade, system suppliers have worked at reducing the cost of equipment, the first installed costs of active and passive cable plant and the operations costs of FTTx systems. Both HFC and FTTC solutions have been deployed for most of the decade, and are relatively simple optically, consisting of point-to-point fiber transmission followed by more complex electronic splitting in a cabinet near the customers. Consequently, a major cost reduction driver has been electronics advances of IC integration. For HFC, this has resulted from the DOCSIS 1.0 and 2.0 standards, which allows IC designers to build special-purpose ICs to a common standard and increase the IC volumes dramatically. In a similar vein, the xDSL standardization process has allowed aDSL and vDSL to be implemented in standard IC designs, so that the most recent vDSL2 IC designs are capable of supporting all previous standards by firmware control. As a result of the standardization and volume manufacture of chip sets for HFC and FTTC, the OEM chip costs have been reduced, leading to lower first installed costs for carriers. In both cases, these IC advances also reduce the serving office, outside plant, and customer equipment size, power consumption, and number of active parts in the system. These changes tend to reduce the installation costs as well as the ongoing operations costs of these systems. We can expect similar IC advances in FTTH systems as well when the annual deployment rates are large enough to justify custom IC development costs for FTTH systems—thus bringing down the cost of FTTH equipment in like manner.

10.7.1 Optical Transmitter and Receiver Technology Advances

For all three FTTx options, the photonics parts costs have been driven down by volume manufacturing. A prime example of this is the triplexer that is used in FTTH systems to receive downstream digital and analog signals and transmit upstream digital signals. Only a few years ago triplexers were manufactured using packaged photodiodes, packaged lasers, thin-film filter components, and bulk coupling and splitting optics—resulting in labor-intensive packaging costs and a triplexer price of $350 per unit. Now, with volume demands for triplexers at over million per year, the manufacturing approaches have turned to integrated

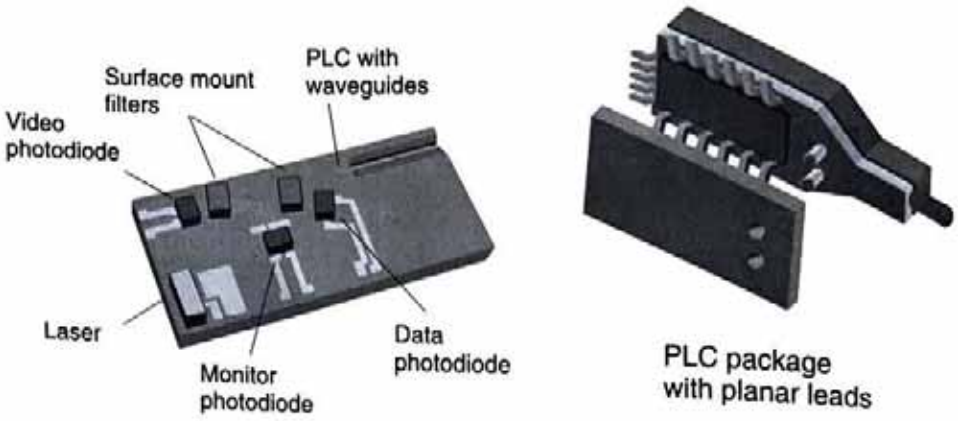


Figure 10.19 Sketch of triplexer solution enabling volume manufacturing (this figure may be seen in color on the included CD-ROM).

solutions, where automated pick-and-place can be applied. For example, planar lightwave circuits [36–39] are used to form the coupling waveguides and wavelength filters, photodiode, and laser chips are flip-chip mounted directly on the PLC substrate, and fiber alignment is done by direct coupling via precision grooves in the PLC substrate [40] (Figure 10.19). These approaches bring the advantage that they can reliably achieve fiber alignment tolerances and good fiber coupling efficiency with a controlled waveguide transition design to the PLC waveguide, by using the precision and yield associated with IC fabrication processes [41]. The result of all these technology improvements is a triplexer that can be offered at a price of $50 or lower to system suppliers. Since the triplexer is a dedicated part (one per subscriber), its cost reduction has a dollar-for-dollar impact on the cost of the overall system. The cost reduction of this single component alone has made a significant impact on lowering the per subscriber cost of FTTH systems.

10.7.2 Cable Management Technology Advances

Installation of optical fiber cable, cabinets and other passive plant components in the feeder, distribution, and drop portions of the outside plant can account for more than half of the cost of installation of an FTTx system. This cost component has received a lot of attention to simplify installation techniques, to speed deployment, and to improve labor efficiency and effectiveness. One way to accomplish this is to integrate much of the time-consuming cable management parts together in the factory [42], where volume manufacturing methods and sophisticated assembly equipment can be easily supported. This means that the distribution cable, branch entry points, and cable terminations must be uniquely designed for a specific location and assembled under controlled conditions in the factory. Since each location has a different set of physical measurements along its route from distribution cable to branch entry point to splitter cabinet, this requires careful network planning and advance engineering effort by the carrier. In any installation, whether

using field installed terminations or terminations installed at the factory, it is necessary to engineer and plan the network for a specific location. But when you integrate the cable assembly in the factory, the engineering measurements must be more accurate because there is no chance to make a final cable length adjustment in the field by cutting the cable shorter.

For cable assemblies that are integrated in the factory, the carrier's craft personnel are sent into the field prior to ordering the outside plant products, where they use precise laser rangefinders, measuring wheels, and pull-tapes to determine the distances of each cable, branch point, and termination on the specific route. These measurements are provided to a few inches tolerance for the cable manufacturer via an online configuration manager, which generates a bill of materials automatically. Then each cable management assembly is custom-manufactured to those demanding specifications and delivered to the installation site—ready to roll off the reel and fasten in place—without any field-related cutting of cables or splicing of cable branch entry points, splice points, or closures [43, 44]. Figure 10.20 shows photos of

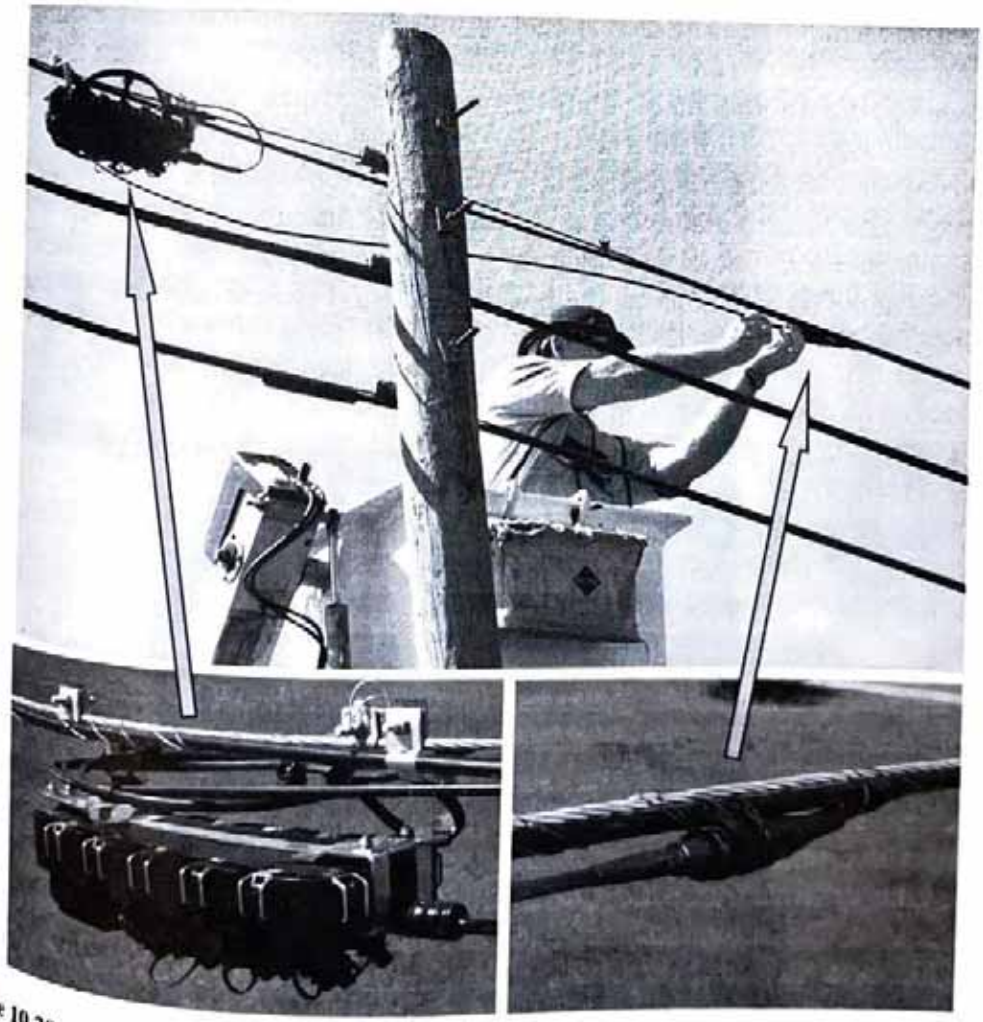


Figure 10.20 Factory installed aerial cable branch point and fiber drop terminal closure (photos) (this figure may be seen in color on the included CD-ROM).

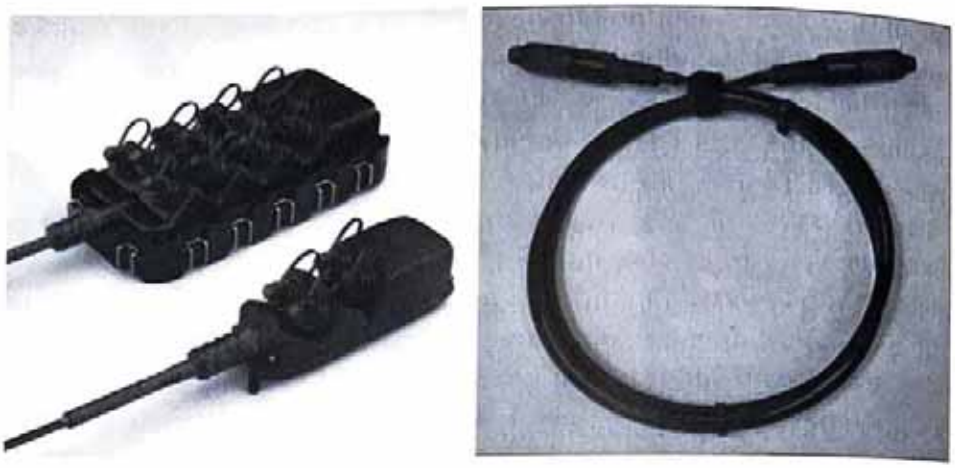


Figure 10.21 Fiber drop terminals and drop cable with robust connectors installed at the factory (photos) (this figure may be seen in color on the included CD-ROM).

aerial cable terminal closure and branch points being installed, and Figure 10.21 shows photos of drop cable terminations and drop cable with robust connectors manufactured for this method. With this technique, installation times can be cut from nearly 2 days to less than 3 hours, and the entire process takes fewer truck rolls, fewer tools, and requires much less time working overhead in a bucket. To further ease installation, and to assure a minimum of future maintenance events, the splitter cabinets, enclosures, branch points, terminals, and drop cables are all fitted with connectors at the factory, making them waterproof and environmentally robust while minimizing splicing in the field.

10.7.3 Outside Plant Cabinet Technology Advances

In a similar fashion, the cabinets that house the splitters are factory assembled and internally preconnected according to an engineering design, so that they can be set in place and attached to their dedicated cables without further craft involvement interior to the cabinet. The initial designs of the cabinets were rather massive, requiring several laborers and a crane to unload and place them. Subsequent technology advances allowed the cabinets to be miniaturized so that they can be lifted by hand and set in place by an individual [45]. An enabling step toward this objective was to improve the bend tolerance of the fiber [19, 46, 47] used in the 1:32 splitters, from 75 mm bending radius to 30 mm bending radius. This was combined with a reduction in the fiber jumper diameter from 2.9 to 2.0 mm, to allow the 1:32 splitters [33] to be reduced in size from 191 mm × 131 mm to 125 mm × 63 mm on the long dimensions, because the excess fiber coiled inside the splitter package could be correspondingly reduced in diameter. Overall this resulted in a 78% reduction in the spatial volume taken up by the splitters (Figure 10.22). At the same time, the layout of the cabinet was improved to make it

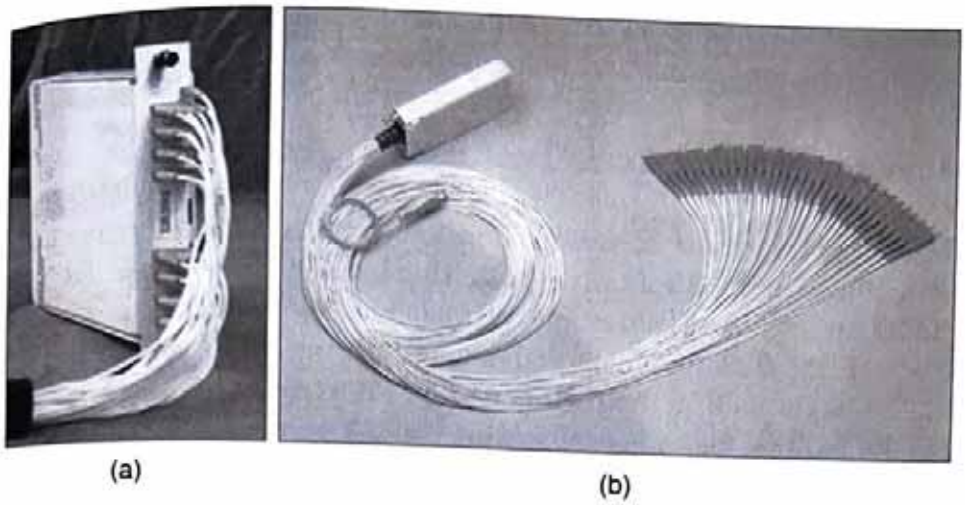


Figure 10.22 Splitter size reduction associated with bend tolerant fiber (a) before and (b) after size reduction (photos) (this figure may be seen in color on the included CD-ROM).

more craft friendly. The splitter size reduction together with the improved cabinet layout then enabled the splitter cabinets to be redesigned to allow a smaller footprint, resulting in a smaller cabinet with less metal, making it much lighter. An example of the cabinet size reduction that resulted from improved fiber bend tolerance is shown in Figure 10.23, where it is clear that the cabinet volume was reduced by more than a factor of four. Such size reductions have diverse impact on costs; ranging from reduced space to store inventory prior to installation, reduced shipping fees, reduced need for heavy equipment to install the cabinets, and reduced labor during cabinet installation. In addition, the smaller cabinets are

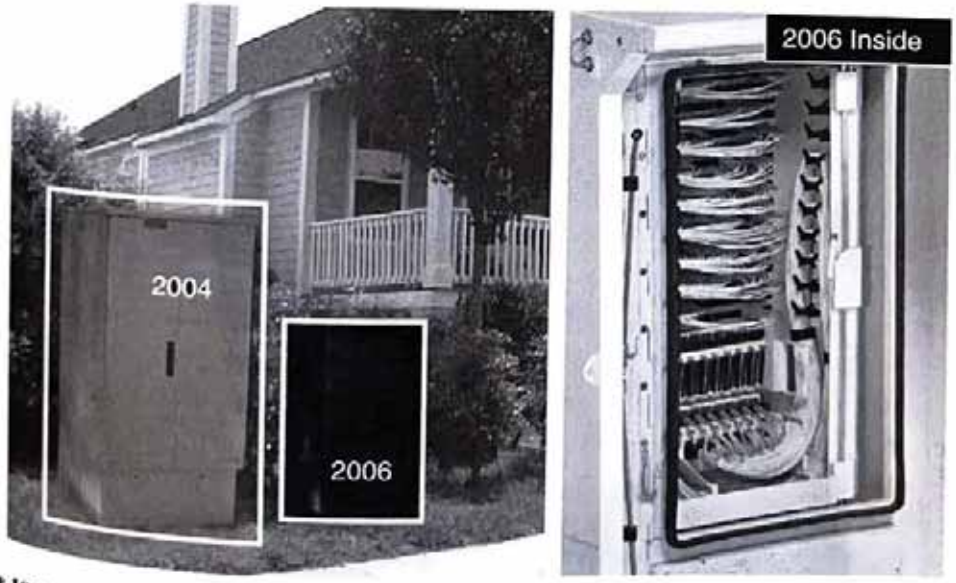


Figure 10.23 Cabinet size reduction associated with bend tolerant fiber, smaller cable diameter, and smaller splitters (photos) (this figure may be seen in color on the included CD-ROM).

less intrusive to the environment, can be more easily placed in a wider variety of locations than the bigger cabinets and experience less vandalism.

10.7.4 Fiber Performance Technology Advances

There is another important technology development that is beginning to play a role in reducing the overall FTTx system cost. This is the discovery that the stimulated Brillouin scattering (SBS) threshold in fiber can be increased, allowing higher launch power for analog signals without introducing more system impairments [48, 49]. To begin with, in both HFC and FTTH systems, the analog signal loss budget is the limiting factor in determining the reach from the serving office out to the active optoelectronics (see Figure 10.24). The loss budget is set at the subscriber end by the receiver noise limitations at a level of −5 dBm to −10 dBm, depending on the system under consideration. Then to get a high enough loss budget for reasonable reach, the launch power has to be very high—in excess of +15 dBm. With standard single-mode fiber (meeting G.652 standards), increasing the launch power above about +17 dBm for the analog signals introduces significant impairments due to the reflected SBS power, which emphasizes all of the key analog impairments [50]. The received noise is increased [carrier-to-noise ratio (CNR) is reduced] as a result of laser phase-to-intensity conversion by the SBS, and intermodulation distortions [CSO and composite triple beat (CTB)] are introduced in the modulated carrier because the optical carrier wave is above the

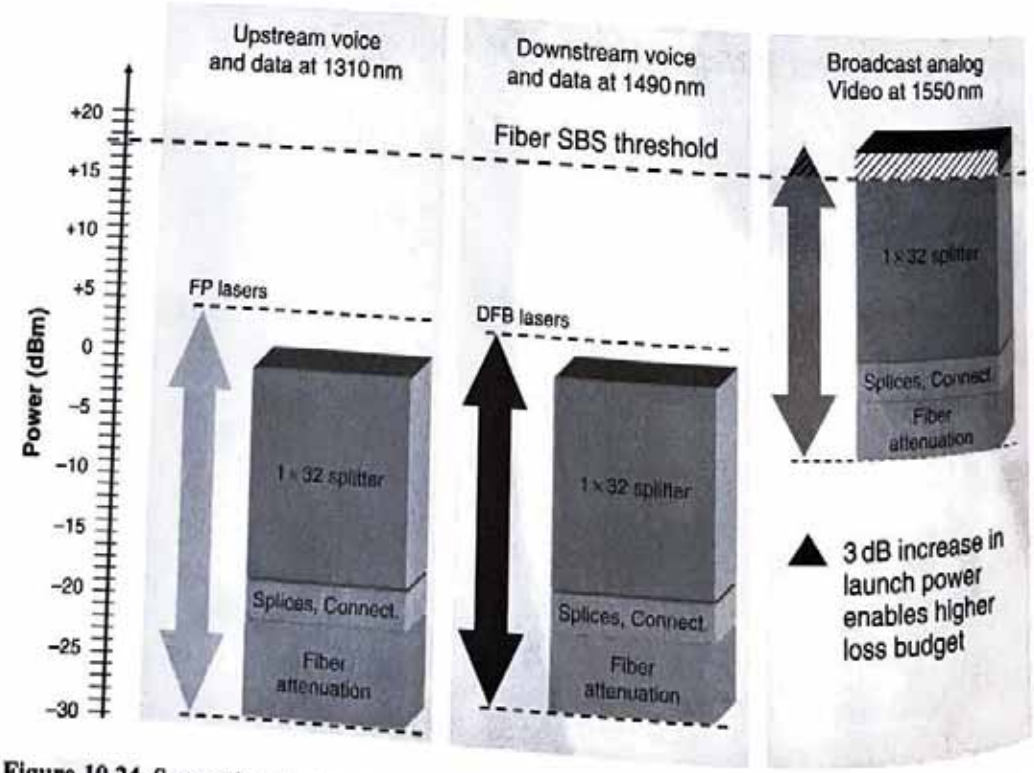


Figure 10.24 System loss budget for analog and digital transmission (this figure may be seen in color on the included CD-ROM).

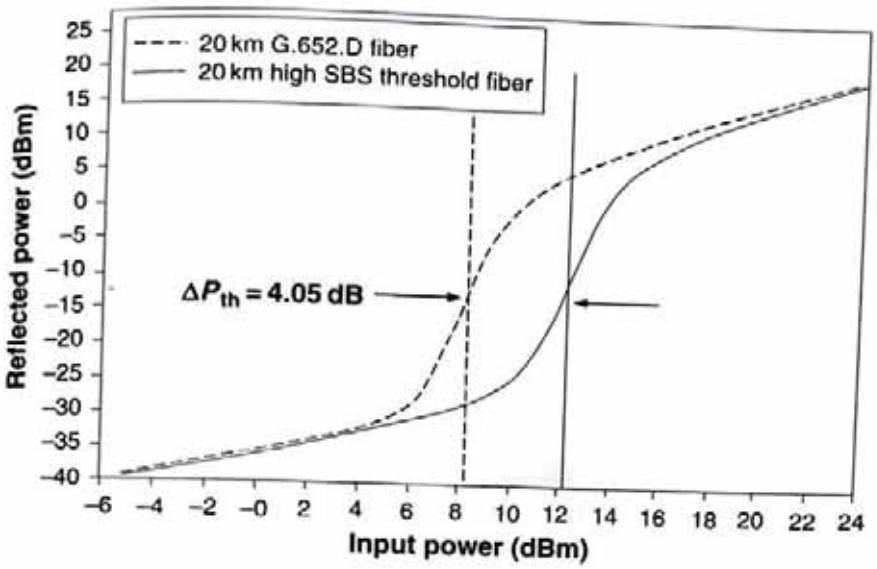
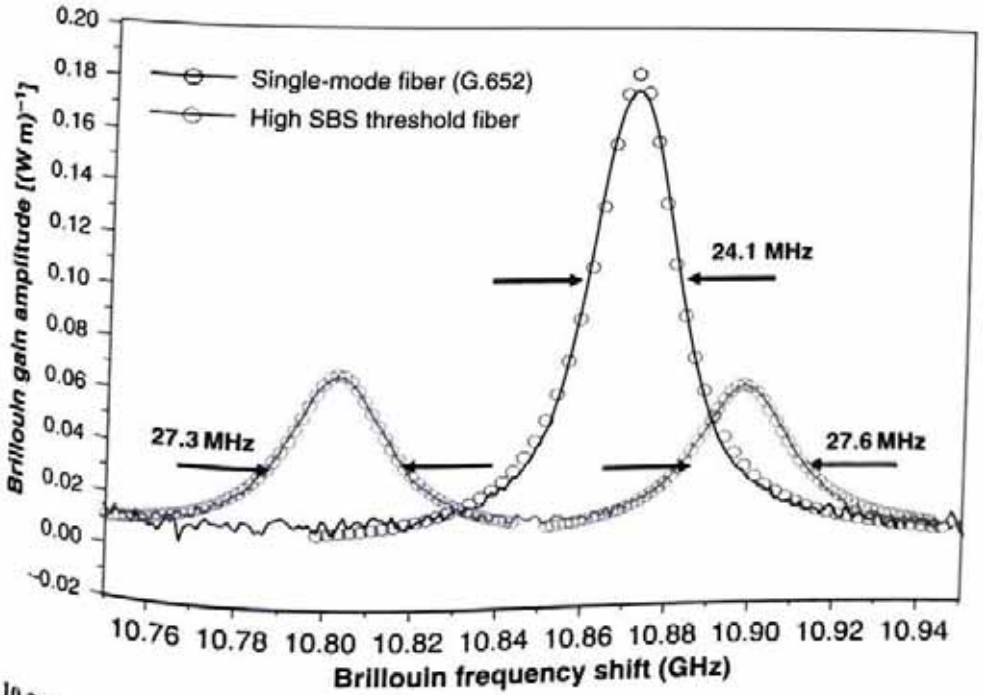


Figure 10.25 Comparison of fiber meeting G.652 standards and high-SBS threshold fiber performance.

SBS threshold and attenuated relative to the modulated sidebands which are below the SBS threshold. Fortunately, this SBS threshold can be controlled to some extent by the design of the fiber profile, which can be optimized to reduce the interaction between the acoustic wave and the optical field in the waveguide and ameliorate the cause of SBS [51, 52] (see Figure 10.25). A suitable fiber design can produce Brillouin gain spectra that are only about 35–40% as strong as the Brillouin gain spectra in standard single-mode fiber [50, 53] (see Figure 10.26).



ure 10.26 Fiber theory for improving SBS threshold (this figure may be seen in color on the luded CD-ROM).

The result is a fiber that can accommodate about 4 dB higher launch power, while still retaining the same analog impairments seen in standard single-mode fiber. Since the fiber design itself avoids excess SBS impairments, cheaper video transmitters can be used because they don't have to compensate for that impairment.

The impact of the high-threshold SBS fiber is to provide relief to the system in several potential ways. In HFC systems, it is possible to launch higher power and place the node farther from the serving office. In FTTH systems, the higher launch power enables twice the splitting with a corresponding increase in cost sharing of key system components [54], or alternatively an increased reach of 10 km or so which enables more subscribers to be accessible to the serving office. Since this fiber performance has to be designed into the system, the impact of this improvement in fiber attributes is only now beginning to be felt in the industry. But laboratory measurements confirm that the system improvements suggested by the Brillouin gain spectrum reductions can in fact be realized (Figure 10.27). Since the improved fiber is completely compatible with standard single-mode fiber, achieving similar or better attenuation, fiber coupling, and splicing attributes, these system cost savings can be passed on to a large extent to system suppliers, to the carriers and ultimately to the subscribers as lower access fees.

Taken together, all of these technological advancements have made a steady impact on the costs of FTTx systems. For example, a progression over time of the system costs for FTTH approaches is shown in Figure 10.28, where the system cost has declined about 15% per year during the 1990s and about 20% per year between 2000 and 2004 [35], to a point where the cost is about $1300 per subscriber in 2006. With continued deployment, increasing volumes, improved system range

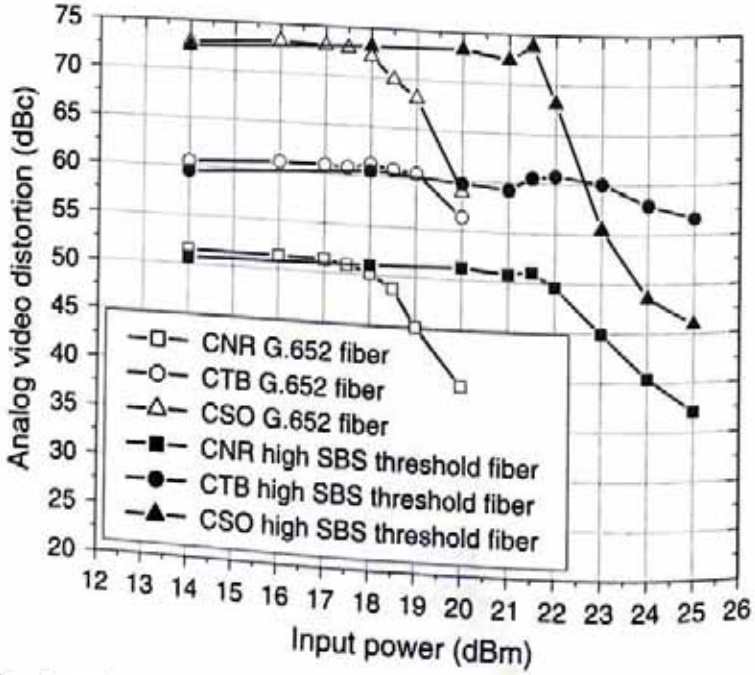


Figure 10.27 Analog impai... ...fiber

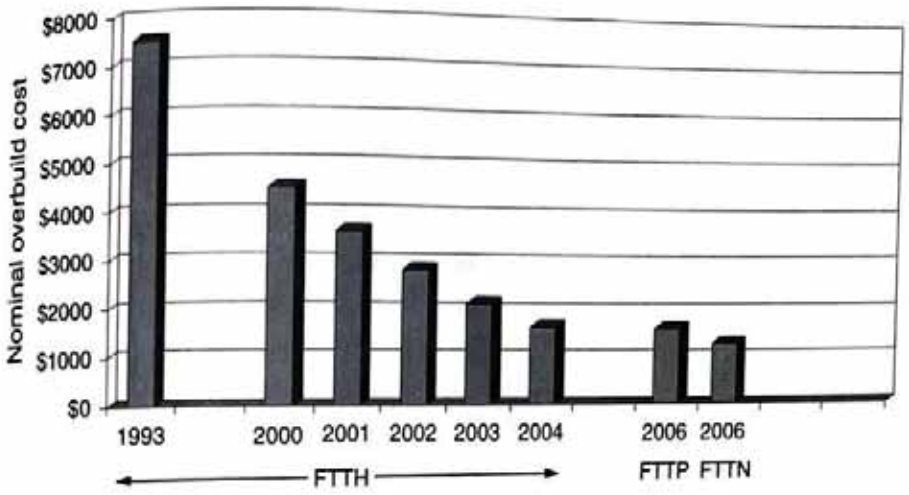


Figure 10.28 Cost decline of FTTx solutions with time (this figure may be seen in color on the included CD-ROM).

and split ratio [13, 55], and competitive pricing the cost per subscriber will likely continue moving downward toward the $1000 target.

10.8 FUTURE BANDWIDTH ADVANCEMENTS

While technology advancements are helping to bring infrastructure and operations costs down, they also introduce the potential for increased user bandwidth performance. This is fortuitous, as the cycle of user technology adoption points toward a 10-fold increase in user bandwidth by 2010 to 100 Mb/s per user [56] for a few million users, and 5 years later to 1 Gb/s per user [57] with a similar user base. These changes in the industry and in user needs are going to drive the technologies that carriers deploy. In fact, the trends in standard products for HFC, FTTC, and FTTH all show the potential for achieving upgrades to 1 Gb/s per user by the time such needs develop, but this requires pushing fiber closer to the home than is currently the case for FTTC and HFC networks. So those service providers will have to consider carefully their evolution and capital spending plans to assure they can continue to meet the competitive needs of users in the long term.

Of course, it is difficult to envision exactly how 1 Gb/s per user could be useful a decade from now. But we have seen technology transform our social and entertainment behavior in the past. Phonograph records gave way to CDs, which have much higher fidelity for music, and letters gave way to e-mail, which is faster and can go to multiple recipients. Film gave way to digital images, which enable instant proofing and sharing with friends and family. Black and white TV gave way to color TV, and now analog TV is being replaced by digital TV, which can enable interactive viewing and game playing. Movie theaters are giving way to DVD players with wide screens, surround sound, and home theater arrangements. Each of these technical advances has ushered in new ways to view and share

information and entertainment content, leading to unexpected demands for higher bandwidth in our daily lives. As the trend continues, there will inevitably be more sharing of digital images, digital music files, digital movie files, digital home movie clips, digital news feeds, with the consumer electronics industry fueling the trend by offering a whole host of innovative communications and entertainment appliances catering to our growing wants and needs. Ultimately, the increased capability and fidelity of sharing and displaying images and movies will lead to the adoption of 1 Gb/s data rates being delivered to our homes and distributed throughout the household, most likely with capabilities for symmetric traffic flows in both directions.

With some imagination, and imperfect knowledge that the historical trends presented here project into the future, it is possible to envision the following scenario playing out over the next two decades.

- *Applications*: all industrial and entertainment content is digital, and able to be delivered to (and from) homes and businesses at Gb/s rates, where it is stored for on-demand use at the subscriber's convenience, and viewed on large-screen digital displays.
- *Technology*: fiber-based systems are used exclusively for broadband access, with many homes and business connected directly to fiber and the remainder connected from cabinets within 150 m of their building.
- *Deployment*: carriers offer integrated multimedia service packages that include entertainment (music and movies), high-speed Internet, cellular communications, as well as voice services. They also offer in-home networking using wireless and structured cabling options, so subscribers can effectively utilize their bandwidth.
- *Economics*: the capital expenditures for broadband access have been largely recovered and the business cases are positive because of reduced operations costs, allowing carriers to compete on price and bandwidth, and to expand infrastructure to support heavy use of Gb/s access rates.

By the time that 1 Gb/s per user is needed and affordable, much of the broadband access infrastructure will already have been installed and in service in the world. Because of this, upgrade strategies are of key importance for how to evolve from today's architectures to those that support Gb/s per user capacity 10–15 years from now. Even beyond this time frame, WDM-PON (wavelength-division multiplexed-PON) and 10G-PON (10 Gb/s-PON) approaches could enable even higher capacity, and complicate the evolution strategy and planning further [58]. While FTTH systems can accommodate such change by replacing or upgrading the transmission equipment at the ends of the access network, the FTTC and HFC systems may require labor-intensive changes in the drop plant as well. These are complex techno-economic issues and worthy of study and experimental investigation right now, although the issues will linger so that there is time to work through the possibilities to arrive at sound solutions before it is too late.

10.9 SUMMARY

After more than 20 years of research and development, a combination of technological, regulatory, and competitive forces are finally bringing fiber-based broadband access to commercial fruition. Three main approaches, HFC, FTTC/vDSL, and FTTH, are each vying for a leading position in the industry, and each has significant future potential to grow customers and increase bandwidth and associated service offerings. No matter which approach wins, or even if all three remain important, construction of the infrastructure needed to serve the entire global network will take one or two decades to complete, because the capital requirements are enormous. During this time it is almost certain that further technical advances and cost reductions will be adopted, bringing performance levels and bandwidth ever higher and keeping service costs affordable. Ultimately, the potential for Gb/s access speeds is on the horizon, and is a target that can be reached economically, when user demand warrants it, through evolution of the infrastructure that is being deployed today.

ACKNOWLEDGMENTS

The author would like to acknowledge the very broad base of work at Corning Incorporated, from which significant content of this paper is drawn. Contributors to that base of work include colleagues from Corning Science and Technology, Corning Optical Fiber, and Corning Cable Systems organizations. In particular, the author would like to thank John Igel, Robert Whitman, Mark Vaughn, Boh Ruffin, and Scott Bickham for specific contributions to this work. In addition, the author would like to thank Paul Shumate and Tingye Li for sharing their perspective on the historical trends for fiber to the home.

LIST OF ACRONYMS

ANSI — American National Standards Institute, standardizing T1E1 for DSL formats

ATM — Asynchronos Transfer Mode, a signal format for combining digital signals

CATV — Cable TV, a means to provide TV via coaxial cable to homes

CNR — Carrier-to-Noise-Ratio, which is the RF carrier strength relative to the noise

CSO — Composite Second Order, an analog impairment due to non-linear effects

CTB — Composite Triple Beat, an analog impairment due to non-linear effects

DMT	Discrete Multi-Tone modulation format, the line code used by aDSL and vDSL systems
DOCSIS	Data Over Cable Interface Specification, a standard to describe cable modems
aDSL, aDSL2	Asymmetric Digital Subscriber Line, a standard for providing digital signals over copper wires used with FTTC systems
vDSL, vDSL2	Very high speed Digital Subscriber Line, a standard for providing digital signals over copper wires used with FTTC systems
DTV	Digital TV, provides digital television with resolution comparable to analog TV
EDFA	Erbium Doped Fiber Amplifier, provides gain in the 1550 nm band
FCC	Federal Communications Commission, a US regulatory body for communications
FSAN	Full Service Access Network Group, broadband access network standards forum
FTTC	Fiber to the cabinet, which brings fiber to a powered cabinet near the subscribers
FTTH	Fiber to the Home, which brings fiber all the way to the side of the home
FTTN	Fiber to the Node, AT&Ts name for FTTC with the cabinet being at a node up to 1.0 km from the subscribers
FTTP	Fiber to the Premises, Verizon's name for expanded FTTH systems to include fiber to business and multiple dwelling units
FTTx	Fiber to the X, describes fiber-based access systems, such as HFC, FTTC, FTTH
G.652	Standard single mode fiber, meeting the ITU-T standard named G.652
GbE	1.0 Gb/s Ethernet, a standard for Ethernet connections
HDTV	High Definition TV, provides high resolution digital television
HFC	Hybrid Fiber Coax, which brings fiber to a powered cabinet near the subscribers
ILEC	Incumbent Local Exchange Carriers, offer local telephone services
ITU	International Telecommunications Union, a standards body for communications
MSO	Multiple Service Operators, companies offering cable TV, internet and phone service
NTSC	National Television System Committee, analog television standard body
10G-PON	10 Gb/s – PON, a system with downstream data rates of 10 Gb/s
PON	Passive Optical Network, a method of providing passive splitter and handling upstream congestion by auto-ranging used with FTTH systems

A-PON	ATM-PON, a passive system using the ATM signal format
B-PON	Broadband-PON, a passive system using ATM signals and an analog overlay
G-PON	Gb/s PON, a system with Gb/s data rates and ATM signal formats
GE-PON	Gigabit Ethernet-PON, a system using GbE for downstream signals
WDM-PON	Wavelength Division Multiplexed-PON, a system using multiple wavelengths, one for each subscriber
SBS	Stimulated Brillouin Scattering, due to an interaction of acoustic and optical waves
SCM	Sub-Carrier Multiplexed, a means to combine multiple analog signals by interleaving RF carriers, each of which are modulated with analog signals
TDM	Time Division Multiplexed, combines signals by interleaving digital streams
TDMA	Time Division Multiple Access, combines the signals from many users into one
Wi-Fi	Wireless-Fidelity, system for broadcasting internet signals inside homes

REFERENCES

[1] R. E. Wagner, J. I. Igel, R. Whitman et al., "Fiber-based broadband-access deployment in the United States," *J. Lightw. Technol.*, 24(12), 4526, December 2006.

[2] R. E. Wagner, "Broadband Access Network Options," OIDA Workshop on Broadband Access for the Home and Small Business, Palo Alto, April 22–23, 2003.

[3] P. W. Shumate and R. K. Snelling, "Evolution of fiber in the residential loop plant," *IEEE Commun. Mag.*, 29(3), 68–74, March 1991.

[4] US Census Bureau, "Manufacturing, Mining and Construction Statistics" Manufacturers' Shipments Historic Timeseries, http://www.census.gov/indicator/www/m3/hist/naicshist.htm, August 29, 2005.

[5] R. L. Whitman, "Global FTTH Risks and Rewards," FTTH Council Conference, Las Vegas, October 3–6, 2005.

[6] J. B. Bourgeouis, "FTTx in the US: Can we see the Light?" FTTH Council paper, http://www.ftthcouncil.org/documents/763530.pdf, November 2005.

[7] "Internet Users (ITU estimates)," United Nations Statistics Division, Millennium Indicators Database, http://millenniumindicators.un.org/unsd/mdg/Handlers/ExportHandler.ashx?Type=Excel&Series=608, March 10, 2005.

[8] "World Internet Usage and Population Statistics," Miniwatts Marketing Group, http://www.internetworldstats.com/stats.htm, June 30, 2007.

[9] Nielsons NetRatings, "NetView Usage Metrics," http://www.nielsen-netratings.com/news.jsp?section=dat_to&country=us, published monthly.

[10] C. Kehoe, J. Pitkow, K. Sutton et al., GVU Tenth World Wide Web User Survey, Georgia Institute of Technology, www.cc.gatech.edu/gvu/user_surveys, May 14, 1999.

[11] M. J. Flanigan, *"TIA's 2006 Telecommunications Market Review and Forecast,"* Telecommunications Industry Association, Arlington, VA, February 2006.

[12] M. D. Vaughn, "FTTH Cost Modeling: Impact of Split Location and Value of Added Splits, Extended Reach and Equipment Consolidation," 2003 FTTH Council Conference, Session B06, New Orleans, October 7–9, 2003.

[13] M. D. Vaughn, Member, IEEE, David Kozischek, David Meis, Aleksandra Boskovic, and Richard E. Wagner, "Value of reach and split ratio increase in FTTH Access networks," *J. Lightw. Technol.*, 22, 2617–2622, November 2004.

[14] M. D. Vaughn, M. C. Meow, and R. E. Wagner, "A bottom-up traffic demand model for LH and metro optical networks," *OFC 2003 Technical Digest*, Paper MF111, Atlanta, March 23–28, 2003.

[15] Telecommunications Act of 1996, Pub. L. No. 104–104, 110 Stat. 56.

[16] M. H. Dortch, Secretary FCC, "Review of the Section 251 Unbundling Obligations of Incumbent Local Exchange Carriers" FCC 03–36, *Report and Order and Order on Remand and Further Notice of Proposed Rulemaking* in CC Docket No. 01–338, adopted February 20, released August 21, 2003.

[17] M. Render, "FTTH/FTTP Update" 2005 FTTH Council Conference, Render, Vanderslice and Associates Research, www.ftthcouncil.org/documents/732751.pdf, Las Vegas, October 3–6, 2005.

[18] M. Render, "Fiber to the Home: Advanced Broadband 2007 Volume One," Render, Vanderslice and Associates Research, http://www.rvallc.com/ftth_reports.aspx, June 2007.

[19] J. George and S. Mettler, "FITP market and drivers in the USA," ITU-T Study Group 6, TD 227A1 (GEN/6), Geneva, May 14–18, 2007.

[20] G. Finnie, "Heavy Reading European FTTH Market report," FTTH Council Europe Annual Conference, Barcelona, February 7–8, 2007.

[21] D. van der Woude, "An overview of Fiber to the Home and Fiber backbone projects," FTTH Council Europe Annual Conference, Barcelona, February 7–8, 2007.

[22] R. Montagne, "FTTH deployment dynamics: Overview and French case," IDATE Consulting and Research, FTTH Council Europe Annual Conference, Barcelona, February 7–8, 2007.

[23] X. Zhuang, S. Huiping, Xuemengchi, and W. Hong, "FTTH in China," Ministry of Information Industry of China, ITU-T Study Group 6, TD 227A1 Rev.1 (GEN/6), Geneva, May 14–18, 2007.

[24] M. Coppet, J. F. Hillier, J. C. Hodulik et al., "Global Communications," UBS Investment Research Q-series, July 10, 2007.

[25] W. Ciciora, J. Farmer, D. Large, and M. Adams, "Chapter 18–Architectural Elements and Examples," *Modern Cable Television Technology*, Second Edition, Morgan Kufmann Publishers in an Imprint of Elsevier, San Francisco, pp. 733–766, 2004.

[26] M. D. Nava and C. Del-Toso, "A Short Overview of theVDSL System Requirements," *IEEE Commun. Mag.*, December 82–90, 2002.

[27] W. F. Caton, Acting Secretary FCC, "Advanced Television Systems and Their Impact Upon the Existing Television Broadcast Service," FCC 97–115, 6th Report And Order on MM Docket No. 87–268, adopted April 3, 1997.

[28] P. Eriksson and B. Odenhammar, "VDSL2: Next important broadband technology," http://www.ericsson.com/ericsson/corpinfo/publications/review/2006_01/06.shtml, Ericcson Review, Issue No. 1 2006.

[29] D. Meis et al., "Centralized vs Distributed Splitting in Passive Optical Networks," NFOEC'05, Paper NWI4, Anaheim, March 6–11, 2005.

[30] D. Faulkner, "Recent Developments in PON Systems Standards in ITU-T," SPIE Optics East 2005, Paper 6012–27, Tutorial IV, Boston, October 23–26, 2005.

[31] K. Nakajima, K. Hogari, J. Zhou et al., "Hole-assisted fiber design for small bending and splice losses," *IEEE Photon. Technol. Lett.*, 15, 1737–1739, 2003.

[32] See, for example, transport availability standards requirements listed in Bellcore GR-418-CORE (downtime minutes per year), ETSI standard EN 300 416 (outages per year), and ITU-T standard G.828 (errored seconds per year).

[33] C. Saravanos, "Reliability Performance of Optical Components in the Outside Plant Environment," 2004 FTTH Council Conference, www.ftthcouncil.org/documents/212550.pdf, Orlando, October 4–6, 2004.

[12] M. D. Vaughn, "FTTH Cost Modeling: Impact of Split Location and Value of Added Splits, Extended Reach and Equipment Consolidation," 2003 FTTH Council Conference, Session B06, New Orleans, October 7–9, 2003.

[13] M. D. Vaughn, Member, IEEE, David Kozischek, David Meis, Aleksandra Boskovic, and Richard E. Wagner, "Value of reach and split ratio increase in FTTH Access networks," *J. Lightw. Technol.*, 22, 2617–2622, November 2004.

[14] M. D. Vaughn, M. C. Meow, and R. E. Wagner, "A bottom-up traffic demand model for LH and metro optical networks," *OFC 2003 Technical Digest*, Paper MF111, Atlanta, March 23–28, 2003.

[15] Telecommunications Act of 1996, Pub. L. No. 104–104, 110 Stat. 56.

[16] M. H. Dortch, Secretary FCC, "Review of the Section 251 Unbundling Obligations of Incumbent Local Exchange Carriers" FCC 03–36, *Report and Order and Order on Remand and Further Notice of Proposed Rulemaking* in CC Docket No. 01–338, adopted February 20, released August 21, 2003.

[17] M. Render, "FTTH/FTTP Update" 2005 FTTH Council Conference, Render, Vanderslice and Associates Research, www.ftthcouncil.org/documents/732751.pdf. Las Vegas, October 3–6, 2005.

[18] M. Render, "Fiber to the Home: Advanced Broadband 2007 Volume One," Render, Vanderslice and Associates Research, http://www.rvallc.com/ftth_reports.aspx, June 2007.

[19] J. George and S. Mettler, "FTTP market and drivers in the USA," ITU-T Study Group 6, TD 227A1 (GEN/6), Geneva, May 14–18, 2007.

[20] G. Finnie, "Heavy Reading European FTTH Market report," FTTH Council Europe Annual Conference, Barcelona, February 7–8, 2007.

[21] D. van der Woude, "An overview of Fiber to the Home and Fiber backbone projects," FTTH Council Europe Annual Conference, Barcelona, February 7–8, 2007.

[22] R. Montagne, "FTTH deployment dynamics: Overview and French case," IDATE Consulting and Research, FTTH Council Europe Annual Conference, Barcelona, February 7–8, 2007.

[23] X. Zhuang, S. Huiping, Xuemengchi, and W. Hong, "FTTH in China," Ministry of Information Industry of China, ITU-T Study Group 6, TD 227A1 Rev.1 (GEN/6), Geneva, May 14–18, 2007.

[24] M. Coppet, J. F. Hillier, J. C. Hodulik et al., "Global Communications," UBS Investment Research Q-series, July 10, 2007.

[25] W. Ciciora, J. Farmer, D. Large, and M. Adams, "Chapter 18–Architectural Elements and Examples," *Modern Cable Television Technology*, Second Edition, Morgan Kufmann Publishers in an Imprint of Elsevier, San Francisco, pp. 733–766, 2004.

[26] M. D. Nava and C. Del-Toso, "A Short Overview of theVDSL System Requirements," *IEEE Commun. Mag.*, December 82–90, 2002.

[27] W. F. Caton, Acting Secretary FCC, "Advanced Television Systems and Their Impact Upon the Existing Television Broadcast Service," FCC 97–115, 6th Report And Order on MM Docket No. 87–268, adopted April 3, 1997.

[28] P. Eriksson and B. Odenhammar, "VDSL2: Next important broadband technology," http://www.ericsson.com/ericsson/corpinfo/publications/review/2006_01/06.shtml, Ericcson Review, Issue No. 1 2006.

[29] D. Meis et al., "Centralized vs Distributed Splitting in Passive Optical Networks," NFOEC'05, Paper NWI4, Anaheim, March 6–11, 2005.

[30] D. Faulkner, "Recent Developments in PON Systems Standards in ITU-T," SPIE Optics East 2005, Paper 6012–27, Tutorial IV, Boston, October 23–26, 2005.

[31] K. Nakajima, K. Hogari, J. Zhou et al., "Hole-assisted fiber design for small bending and splice losses," *IEEE Photon. Technol. Lett.*, 15, 1737–1739, 2003.

[32] See, for example, transport availability standards requirements listed in Bellcore GR-418-CORE (downtime minutes per year), ETSI standard EN 300 416 (outages per year), and ITU-T standard G.828 (errored seconds per year).

[33] C. Saravanos, "Reliability Performance of Optical Components in the Outside Plant Environment," 2004 FTTH Council Conference, www.ftthcouncil.org/documents/212550.pdf, Orlando, October 4–6, 2004.

[34] J. Halpern, G. Garceau, and S. Thomas, "Fiber-to-the-Premises: Revolutionizing the Bell's Telecom Networks," Bernstein Research & Telcordia Research Study, http://www.telcordia.com/products/fttp/bernstein-telcordia.pdf, May 20, 2004.
[35] P. Garvey, "Making Cents of it All – Costs/Key Drivers for Profitable FTTH," FTTH Council Conference, Las Vegas, October 3–6, 2005.
[36] M. Pearson, S. Bidnyk, A. Balakrishnan, and M. Gao, "PLC Platform for Low-Cost Optical Access Components," *Technical Digest IEEE LEOS 2005*, Paper WX1, Sidney, October 23–27, 2005.
[37] Y. Nakanishi, H. Hirota, K. Watanabe et al., "PLC-based WDM Transceiver with Modular Structure using Chip-Scale-Packaged OE-Devices," *Technical Digest IEEE LEOS 2005*, Paper WX2, Sidney, October 23–27, 2005.
[38] W. Chen, K. B. Little, W. Chen et al. "Compact, Low cost Chip Scale Triplexer WDM Filters," *OFC 2006 Technical Digest*, Paper PDP12, Anaheim, March 5–10, 2006.
[39] H. Sasaki, M. Uekawa, Y. Maeno et al., "A low-cost micro-BOSA using Si microlens integrated on Si optical bench for PON application," *OFC 2006 Technical Digest*, Paper OWL6, Anaheim, March 5–10, 2006.
[40] D. W. Vernooy, J. S. Paslaski, H. A. Blauvelt et al., "Alignment-Insensitive Coupling for PLC-Based Surface Mount Photonics," *IEEE PTL*, 16(1), 269–271, 2004.
[41] Wenhua Lin, and T. Smith, "Silicon Opto-Electronic Integrated Circuits: Bringing the Excellence of Silicon into Optical Communications – The Key to Large Scale Integration" http://www.kotura.com/SOEIC_Technology.pdf, February 2004.
[42] M. Turner, "Moving backwards in the OSP: Improving the FTTH distribution segment helps advanced services move forward," OutsidePlant Magazine, Article EVO-634, January 2006.
[43] D. Meis, "Get big savings in time and money," Broadband Properties Magazine, p. 20, August 2005.
[44] V. O'Byrne, D. Kokkinos, D. Meis et al., "UPC vs APC Connector Performance in Passive Optical Networks," NFOEC'05, Paper NTuF3, Anaheim, March 6–11, 2005.
[45] A. Woodfin, "Access makes the Parts Grow Stronger," FTTH Council Conference, Las Vegas, October 3–6, 2005.
[46] K. Himeno, S. Matsuo, N. Guan, and A. Wada, "Low-Bending-Loss Single-Mode Fibers for Fiber-to-the-Home," *JLT*, 23(11), November 3494–3499, 2005.
[47] G.S. Glaesemann, M.J. Winningham, and S.R. Bickham, "Single-Mode Fiber for High Power Applications with Small Bend Radii," in Proc. *SPIE Photonics Europe*, paper 6193–23, Strasbourg, April 3–7, 2006.
[48] A. Woodfin, A. B. Ruffin, and J. Painter, "Advances in optical fiber technology for analog transport: technical advantages and recent deployment experience," National Cable and Telecommunications Association, NCTA Technical Papers, 44th Edition, pp. 93–100, Chicago, June 8–11, 2003.
[49] A. Woodfin, J. Painter, and A. Boh Ruffin, "Stimulated Brillouin Scattering Suppression with Alternate Fiber Types – Analysis and Deployment Experience," NFOEC'03, *Technical Digest*, 4, (1265), Orlando, September 7–11, 2003.
[50] A. B. Ruffin, F. Annunziata, S. Bickham et al., "Passive Stimulated Brillouin Scattering Suppression for Broadband Passive Optical Networks," *Technical Digest LEOS'04*, Paper ThE2, Puerto Rico, November 7–11, 2004.
[51] A. Kobyakov, S. Kumar, D. Q. Chowdhury et al., "Design concept for optical fibers with enhanced SBS threshold," *Opt. Express*, 13, 5338–5346, July 2005.
[52] M.-J. Li, X. Chen, J. Wang et al., "Fiber Designs for Reducing Stimulated Brillouin Scattering," *OFC 2006 Technical Digest*, Paper OTuA4, Anaheim, March 5–10, 2006.
[53] A. B. Ruffin, Ming-Jun Li, Xin Chen et al., "Brillouin gain analysis for fibers with different refractive indices," *Opt. Lett.*, 30(23), 3123–3125, December 1, 2005.
[54] M. D. Vaughn, A. B. Ruffin, A. Kobyakov et al., "Techno-economic study of the value of high stimulated Brillouin scattering threshold single-mode fiber utilization in fiber-to-the-home access networks," *JON*, 5(1), January 4, 2006.

[55] J. Lepley, M. Thakur, I. Tsalamanis et al., "VDSL transmission over a fiber extended-access network," *J. Opt. Netw.*, 4, 517–523, 2005.
[56] R. Hunt, "Reforming Telecom Policy for the Big Broadband Era," New America Foundation article, http://www.newamerica.net/index.cfm?pg=publications&SecID=30&AlOf=2003, December 19, 2003.
[57] IEEE-USA Committee on Communications and Information Policy, "FTTH Policy: The Case for Ubiquitous Gigabit Open-Access Nets," *Broadband Properties Magazine*, p. 12, August 2005.
[58] R. Heron, "FTTx Architecture Transition Strategies," *NFOEC'2007*, Paper NThF, Anaheim, March 25–29, 2007.

Dr. Ruiz Library

12

Metro networks: Services and technologies

Loukas Paraschis[*], Ori Gerstel[*], and Michael Y. Frankel[†]

[*]*Cisco Systems, San Jose, CA, USA*
[†]*CTO Office, Ciena, Linthicum, MD, USA*

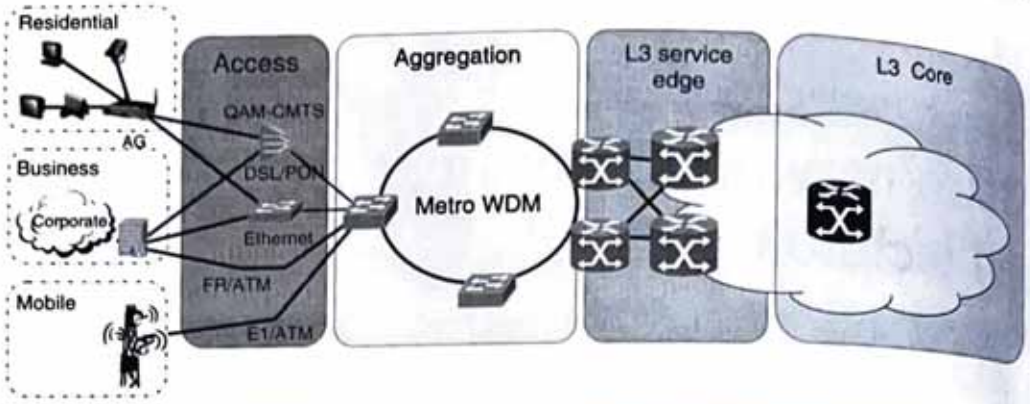
Abstract

This chapter summarizes the innovation in network architectures and optical transport that has enabled metropolitan networks to meet the diverse service needs of enterprise and residential applications, and cost-effectively scale to hundreds of Gb/s of capacity, and hundreds of kilometers of reach. A converged metro network, where IP/Ethernet services and traditional time-division multiplexed (TDM) traffic operate over a common intelligent wavelength-division multiplexed (WDM) transport layer, has become the most appropriate architecture for significantly reduced network operational cost. At the same time, advanced technology, and system-level intelligence have improved the deployment and manageability of WDM transport. The most important application drivers, system advancements, and associated technology innovations in metropolitan optical networks are being reviewed.

12.1 INTRODUCTION AND DEFINITIONS

This chapter discusses the evolution of optical metropolitan networks. We start from the evolution of services over the past several years and next few years, and drill down into increasing details about the implementation of the solution. To understand why the network is evolving the way it is and how it will continue to

Optical Fiber Telecommunications V B: Systems and Networks
Copyright © 2008, Elsevier Inc. All rights reserved.
ISBN: 978-0-12-374172-1



L0–L1	SONET, CWDM	CWDM, DWDM, OTN	N/A	DWDM, OTN
L2–L3	Ethernet	Ethernet, IP, MPLS	IP, MPLS	IP, MPLS

Figure 12.1 Context for metro networking within the entire network and technologies typically deployed (this figure may be seen in color on the included CD-ROM).

evolve, one has to first understand how services are evolving from simple point-to-point transport services to sophisticated packet services for video and other applications. This is covered in Section 12.2. The services, in turn, drive the architecture of the entire network, and this is covered in the Section 12.3. Once the architecture is defined, we are ready to delve into the implications of the architecture on the physical layer as described in the Section 12.4, while Section 12.5 discusses network automation tools required for successful design, deployment, and operation. We summarize the chapter in Section 12.6 and provide an outlook into the future of the network in Section 12.7.

Figure 12.1 depicts several network layers based on their packet functionality: access to customers, aggregation of traffic from various access points into larger central offices, the edge of the packet layer, and the core of the network. The table below the figure represents the most common technologies per layer, from a transport perspective (L0–L1 typically) and a packet perspective (L2–L3). A different segmentation is mostly based on geographical reach: access, metro, regional, and long-haul networks. While aggregation networks often correspond to metro/regional networks and core networks are often long-haul, this is not always the case: regional service providers (SPs) often run a metro core network, and access networks in sparsely populated areas often cover regional distances. In the rest of the section, we focus on the geographic segmentation as we find it more meaningful for dense wavelength-division multiplexing (DWDM) technology.

Access networks are typically classified by reaches below 50 km. Access networks are commonly deployed in a ring-based architecture to provide protection against fiber cuts, and are historically SONET/SDH (Synchronous Optical NETwork/Synchronous digital hierarchy), and more recently coarse

wavelength-division multiplexed (CWDM) systems. A 10-nm spacing of channel wavelengths in a CWDM system drives down the cost of pluggable transceivers by eliminating component cooling requirements, and simplifies optical filter design and manufacture. These systems also tend not to have optical amplification. Optical channel data rates in the access network today are predominantly at or below 2.48 Gb/s, with 4 and 10 Gb/s gaining some recent deployments.

Metro systems may be classified by reaches typically below ~300 km, with currently typical node traffic capacities of one to several 10's Gb/s. Given the reach, number of nodes and optical add/drop granularity, Metro networks are typically equipped with optical amplifiers and support DWDM with wavelength channel spacing of 0.8 nm (100 GHz). The lasers now require cooling, but can still be operated without active wavelength locking. The majority of channels in deployment operate at 2.48 Gb/s, but 10-Gb/s data rates are gaining in market share, and higher data rates are starting to see some spot deployments.

Regional systems are classified by reaches on the order of 600 km and below, and long haul is anything above 600 km. Traffic capacities are on the order of multiple 100's of Gb/s. These networks are always equipped with optical amplifiers. Data rates of 10 Gb/s are prevalent for these systems, with channel spacing as low as 0.2 nm (25 GHz), though 0.4 nm (50 GHz) is much more widely deployed. Active wavelength locking is mandatory for such tight channel spacing, with data rates of 40 Gb/s seeing deployments, and 100 Gb/s being pursued within standards bodies and industry development groups.

12.2 METRO NETWORK APPLICATIONS AND SERVICES

SPs have traditionally relied on different networks to address different consumer and enterprise market needs. POTS, TDM/PSTN, and to some extend ISDN, have typically served the voice-dominated consumer applications, while Frame Relay, asynchronous transfer mode (ATM) and TDM leased-line networks have served the more data-intensive enterprise applications. At the same time, video has been mostly distributed on separate, extensively analog, networks. Internet access, while representing a significant departure from the 64-kB/s voice access lines, has been relatively lightly used—mainly carrying content limited by human participation—such as e-mail and web access. In recent years, the paradigm has undergone a dramatic shift toward streaming and peer-to-peer applications, driving significant growth in the utilization of access lines as well as the core, as it is less dependent on the presence of a human being at the computer to drive the utilization of the network.

Whether the cause is, the spread of cell phones or voice over IP (VoIP), SPs have seen a steady decline in revenues from traditional telephony and a steep increase in voice-originated and other packet traffic. As a result, SPs are trying to find new revenue streams from residential customers via a strategy commonly referred to as

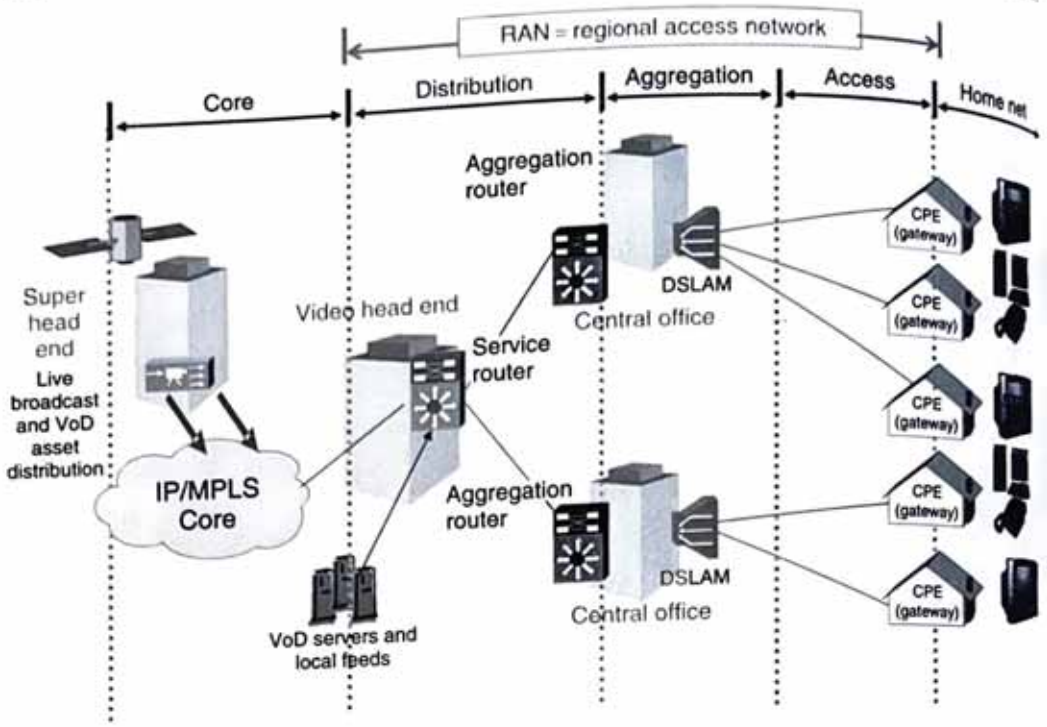


Figure 12.2 Video over broadband architecture overview (this figure may be seen in color on the included CD-ROM).

"triple (or even quadruple) play": providing voice, video, and Internet over a common packet infrastructure. This infrastructure requires significant investment in upgrading residential access, as well as back-end systems to create functions and generate content that will increase customer loyalty. The primary example for such applications is video—including high-definition broadcast and on-demand content. In an ideal SP world, customers will receive all their video needs from a network that is engineered around video delivery. Such a network is shown in Figure 12.2.

However in many geographic locations, particularly in the USA, SPs are facing tough competition from cable providers, who are much more experienced in delivering video content and are also building their own triple play networks over a coax cable infrastructure that is less bandwidth-constrained than the twisted-pairs SPs own. This, in turn, drives SPs to revamp the actual access medium to fiber (to the home, curb, or neighborhood). Perhaps the most serious competition, for both SPs and cable providers, comes from companies who are providing more innovation over the Internet. These "over-the-top" providers use the high-speed Internet access that is part of triple play to deliver video, music, photo sharing, peer-to-peer, virtual communities, multiparty gaming, and many other services without facing the access infrastructure costs. This competition for the hearts and pockets of consumers is driven by application-level innovation delivered over the Internet Protocol (IP), and therefore the transport network should be optimized for either IP or for Ethernet, which is its closely related lower-layer packet transport mechanism.

Enterprise services are experiencing an equally phenomenal growth. The wide-scale adoption of e-commerce, data warehousing, business continuance, server consolidation, application hosting, and supply chain management applications, has fueled significant bandwidth growth in the enterprise network connectivity and storage needs. The business critical nature of most of these applications also calls for uninterrupted and unconstrained connectivity of employees and customers. To best support this, most enterprises have upgraded their networks, replacing ATM, Frame Relay, and TDM private lines, with a ubiquitous Ethernet (GE and 10 GE) transport. In addition, regulatory requirements in the financial and insurance industries increased significantly the bandwidth needed to support disaster recovery. The large financial firms became the early adopters of enterprise WDM metro networks, driven by the need to support very high-bandwidth storage applications for disaster recovery such as asynchronous and synchronous data replication over metro/regional distances. An overview of storage-related services can be found in Figure 12.3.

To increase the value of the service and the resulting revenue per bit, carriers are looking for ways to provide higher level connectivity—beyond simple point-to-point connections between a pair of Ethernet ports. This includes multiplexed services—in which multiple services may be delivered over the same port—distinguished by a "virtual LAN" (VLAN) tag, and handled differently inside the network. It also includes point-to-multipoint and even multipoint-to-multipoint services, in which the network appears to the user switches and routers as a distributed Ethernet switch. These services are realized via various Layer 2 and even Layer 3 mechanisms. A summary of the various Metro Ethernet services can be found in Figure 12.4.

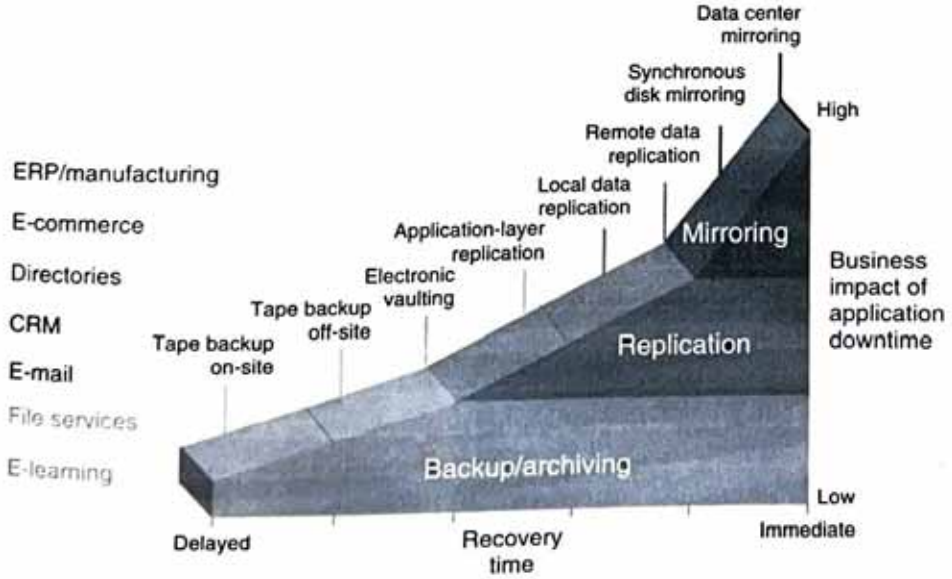


Figure 12.3 Mapping business continuance solutions (this figure may be seen in color on the included CD-ROM).

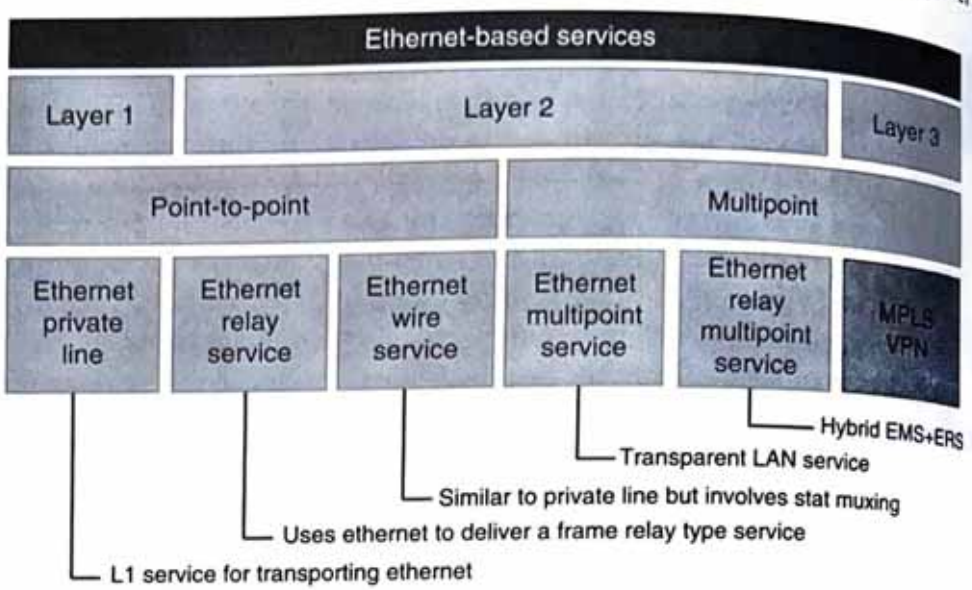


Figure 12.4 Overview of Ethernet-based services (this figure may be seen in color on the include CD-ROM).

12.3 EVOLUTION OF METRO NETWORK ARCHITECTURES

12.3.1 Network Architecture Drivers

In an environment of an ever increasing richness of services, at progressively higher bit rates, but with a price point that must grow slower (or even decline) than their required bandwidth, service providers must build very efficient networks—with the lowest possible Capital expenditure (CAPEX), as well as low operational costs (OPEX).

CAPEX can be optimized by the following means:

- Allowing for as much bandwidth oversubscription as possible, while still respecting the quality of service (QoS) customers are expecting. This cannot be achieved by connections with fixed preallocated bandwidth such as SONET private lines, and drives toward the adoption of packet technologies in the access and aggregation layers,
- Convergence of multiple per-service layers into a unified network. This allows for a better utilization of the network and for better economies of scale, as bandwidth can freely move from one application to another,
- Service flexibility. As new services are introduced, the existing infrastructure must be able to support them, even when the service deviates from the original design of the network.

- Hardware modularity. Hardware must be designed and deployed in a manner that can accept new technologies as they become available, without requiring complete new overbuilds.

OPEX can be optimized by the following means:

- Convergence of layers also helps reduce OPEX as operators do not have to be trained on diverse network elements and distinctive network management capabilities, but rather on one technology. This also allows for more efficient use of the resources as the same operators can work across the entire network instead of working in a "silo" dedicated to one service.
- Increasing the level of automation. Yesterday's transport systems required manual intervention for any change in connection bandwidth and endpoints. Moving to a network that adapts automatically to changes in traffic pattern and bandwidth allow reduction in manual work and cost.

Another related consideration is a barrier to entry and speed of deployment of a new service. Clearly new services will be introduced at an ever increasing rate, sometimes without a clear understanding of their commercial viability. Therefore, it is critical that they can be introduced with minor changes to the existing gear, without requiring a large capital and operational investment that goes with a new infrastructure. This can only be achieved with a converged network that is flexible enough.

So how does a carrier meet these requirements? By converging on a small number of very flexible and cost effective network technologies. Specifically, the following technologies have a good track record of meeting these needs:

- DWDM transport: since photonic systems are less sensitive to protocol, format, and even bit rate, they are able to meet diverse needs over the same fiber infrastructure with tremendous scaling properties,
- Optical transport network (OTN) standard describes a digital wrapper technology for providing a unified way to transport both synchronous (i.e., SONET, SDH) and asynchronous (i.e., Ethernet) protocols, and provide a unified way to manage a diverse services infrastructure [1, 2].
- Ethernet: this technology has morphed over a large number of years to support Enterprise services as well as effective transport for TCP/IP, predominantly in metro aggregation networks,
- TCP/IP: has proven resilient to the mind-boggling changes that the Internet has gone through, and is the basis for much of the application level innovation.

Metropolitan area networks (MANs) have been the most appropriate initial convergence points for multiservice architectures. The significant growth of applications with extensive metropolitan networking requirements has placed increased

emphasis on the scalability of multiservice MANs [3]. MAN architecture has thus evolved; Internetworking multiple "access" traffic-collector fiber rings in "logical" star or mesh, through a larger "regional" network [4, 5]. A converged metro network architecture of Ethernet/IP and high-speed OTN services operating over a common intelligent WDM layer, reduces CAPEX, and even more importantly OPEX, by enabling easier deployment and manageability of services.

12.3.2 Metro Optical Transport Convergence

WDM has been acknowledged early as the most promising technology for scalable metro networks [5, 6]. Its initial deployment, however, remained rather limited, addressing mostly fiber exhaust applications. Metro optical transport is particularly sensitive to the initial cost of the deployed systems, and the CAPEX cost of WDM technologies had been prohibitively high. Attempts to control CAPEX through the use of systems with coarse WDM channel spacing (i.e., 10 or 20 nm) has met with only limited success, as system costs are still set by the transponders, and total unregenerated optical reach and total system capacity is limited. Attempts to increase system capacity by introducing denser channel spacing (i.e., DWDM with 100 GHz spacing) while still avoiding expensive optical amplifiers, resulted in systems with a severe limitation on the number of accessible nodes and total reach.

However, metro networks have critical requirements for service flexibility, and operational simplicity. Moreover, metro WDM cost has to account not only for a fully deployed network, but also for its ability to scale with the amount of deployed bandwidth; as most metro networks do not employ all (or most) WDM channels at the initial deployment phase, but rather "light" unused channels only when actually needed to serve (often unpredictable) network growth. As a result, network operators delayed WDM deployment in their metro networks until these problems were solved with mature technologies from a stable supply base.

At the same time, the evolution of the SONET/SDH transport standards, enabled a successful generation of systems that supported efficient bandwidth provisioning, addressing most of the initial MAN needs, leveraging the advancements in electronics, and 2.5-Gb/s (STM-16) and 10-Gb/s transport (STM-64) [7]. These "next-generation" SONET/SDH systems further allowed improved packet-based transport over the existing time-division multiplexed (TDM) infrastructure, based on data encapsulation and transport protocols (GFP, VCAT, LCAS). Packet-aware service provisioning enabled Ethernet "virtual" private network (VPN) over a common service provider MAN. The initial rate-limited best-effort Ethernet service architectures, evolved to offer QoS guarantees for Ethernet, as well as IP services (like VOIP), and packet-aware ring architecture, like the resilient packet ring (RPR) IEEE 802.17 standard, provided bandwidth spatial reuse. Such Layer-2, and eventually Layer-3, intelligent multiservice provisioning has enabled significant statistical multiplexing gains, enhancing network scalability. The table

illustrates an example of a network with VC-4 granularity that serves a VPN with 4 gigabit Ethernet (GE) sites, six additional point-to-point GE connections, and a storage area network with 2 GE and two fiber channel (FC) services. A "purely" optical solution would require at least six STM-64 rings that, even after leveraging VCAT, would be at least 75% full. An advanced multilayer implementation that employs packet level aggregation and QoS in conjunction with VCAT (ML + VCAT) could be based on just four STM-64 rings, each with less than 40% of capacity utilization, saving more than 50% in network capacity. This new generation of multiservice platforms allowed, for the first time, different services to be deployed over a common network infrastructure, instead of separate networks, improving network operations.

Required VC4:		
VCAT Only		ML + VCAT
91	Ring 1	30
84	Ring 2	30
49	Ring 3	26
21	Ring 4	8
245	Total VC4	94
(6) 75% full	STM64 Rings	(4) 40% full

As traffic needs grew beyond 10 Gb/s per fiber, however, WDM transport became the best alternative for network scalability. To this end, multi-service systems evolved to "incorporate" WDM interfaces that connect them directly onto metro fibers, thus eliminating client optoelectronic (OEO) conversions and costs. The integration of WDM interfaces in the service platforms also changed the traditional "service demarcation point" in the network architecture. This seemingly straightforward convergence of the transport and service layers has introduced additional requirements for improved manageability in the WDM transport. The introduction of many different "wavelength services" amplified the value for "open" WDM architectures that provide robust and flexible transport. In this sense, a converged, flexible WDM metro transport architecture that supports all the different services with the lowest possible OPEX, leveraging elaborate planning and operational tools, and enabling standards-based interoperability, has become increasingly important.

A related but somewhat opposite trend is the integration of increased service layer functionality into the DWDM layer: as packet processors and other service handling mechanisms have become more compact and less expensive, transponders in the DWDM system are no longer restricted to converting client signals to WDM, but have taken on the task of multiplexing services into a wavelength, switching these services to their destination—potentially adding new services along the path, and the related management and control functions. Thus, ADM,

MSPP and Ethernet switch platforms "on a blade" have been introduced, replacing small ADMs and small Ethernet switches that would be managed separately from the DWDM layer by devices that are fully integrated into the DWDM layer. These devices typically have only a few WDM interfaces and are limited in size, and they can be logically interconnected in rings over the WDM layer. An example of such a device and its usage in the network can be found in Figure 12.5.

Figure 12.5(a)shows a conceptual drawing of a DWDM shelf with 3 ADM on a blade cards, each terminating a number of client interfaces and a single WDM wavelength each. These concepts hold for Ethernet-based cards as well.

Figure 12.5(b)shows a typical use of these cards on a physical ring topology. Each rectangle represents an ADM on a blade card and the color codes represent rings of ADMs on a blade. In some cases, these cards are concatenated in the same site to terminate a higher amount of traffic.

Figure 12.5(c) shows how these cards can be deployed over a physical mesh topology.

Another example is fiber channel (FC) "port extenders" which adapt FC over long distances by "spoofing" acknowledgements from the remote device toward the local device, thereby allowing the local device to increase its throughput without waiting a round trip delay for the remote device to acknowledge the

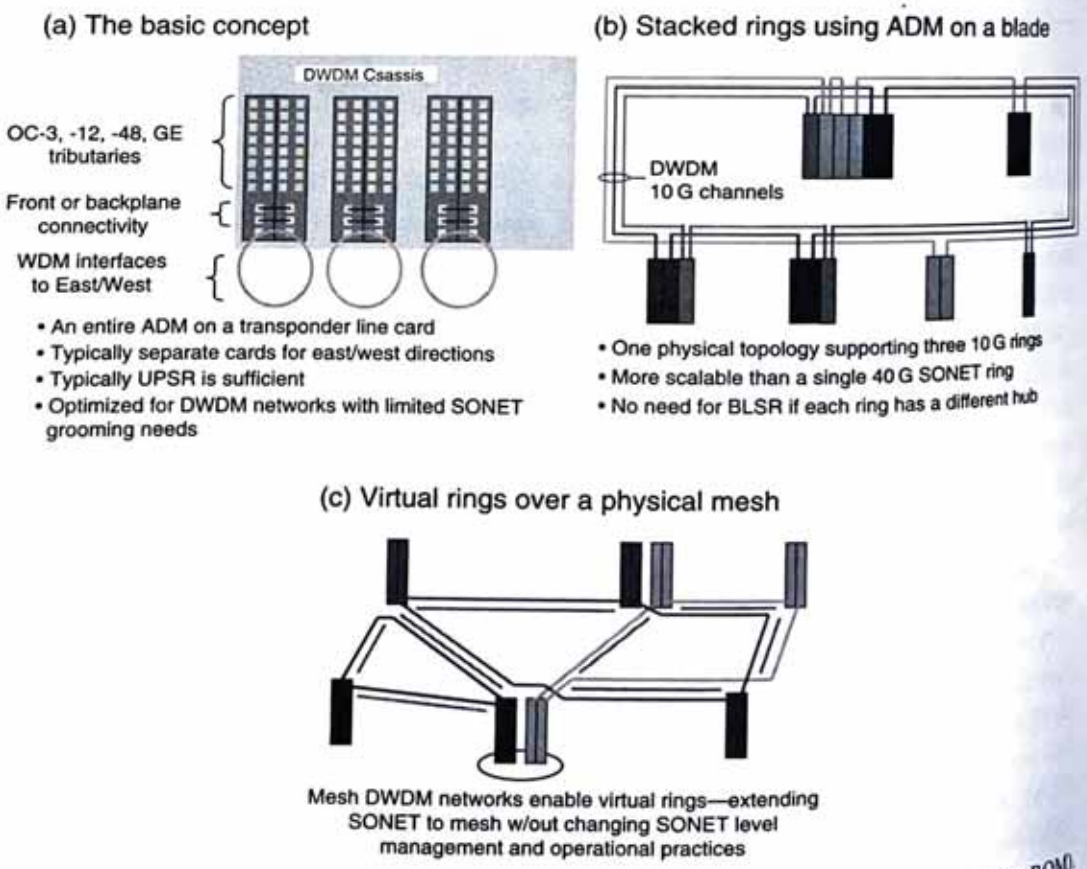


Figure 12.5 The ADM on a blade concept (this figure may be seen in color on the included CD-ROM).

messages. Another value such devices provide is that they give the SP visibility into application level issues and therefore enhance the SPs ability to troubleshoot the system.

Eventually, a new generation of metro-optimized WDM transport (often referred to as multiservice transport platforms or MSTPs) has contributed significantly to the recent progress in MAN WDM deployments. This WDM transport enables elaborate optical add–drop multiplexing (OADM) architectures that transparently interconnect the different MAN nodes. Moreover, such MSTP WDM systems have scaled cost-effectively to hundreds of Gb/s, and to hundreds of kilometers, and have significantly enhanced ease of deployment and operation, by automated control and integrated management of the optical transport layer [8].

12.3.3 Network Survivability

Due to the critical nature and volume of information carried over the network, carriers must ensure that network failures do not result in a loss of customer data. There are a number of schemes that may be implemented, with a general characteristic of providing redundancy in physical transmission route and equipment. At a high level there are two approaches to survivability: (1) protection in the optical layer and (2) protection in the client layer. While optical layer protection provides lower cost, it does not protect against all failures and cannot differentiate between traffic that requires protection and traffic that does not. Therefore, typically, intelligent clients such as routers are in charge of protecting their own traffic over unprotected wavelengths, while less intelligent clients rely on optical layer protection. In the rest of this section, we first focus on optical layer protection and then move to client layer protection.

A common optical layer protection scheme arranges nodes into a physical ring topology, such that a connection between any pair of nodes can take one of two possible physical routes. Should one physical routes experience a breakdown in fiber or equipment, an automatic protection switch (APS) is executed.

Figure 12.6 shows a four-node ring arrangement that is used to clarify various protection schemes encountered in metro networks, i.e., ULSR, UPSR, BLSR, BPSR. First letter indicates data flow direction around the ring, such that Unidirectional (U) implies that Node 1 communicates to Node 2 in a clockwise (CW) direction, and Node 2 also communicates to Node 1 in the same CW direction passing via Nodes 3 and 4. In this case, the counter-clockwise (CCW) direction serves a protection function. Bidirectional (B) implies that Node 1 communicates to Node 2 in a CW direction, while Node 2 communicates to Node 1 in a CCW direction. In this case, the other ring portion serves a protection function. Second letter refers to the whether protection is done at a Line (L) or Path (P) level. SR refers to the basic switched-ring network architecture.

While the terminology varies, the above schemes are generally applicable to both SONET/SDH, DWDM, and optical transport network (OTN) protection (see

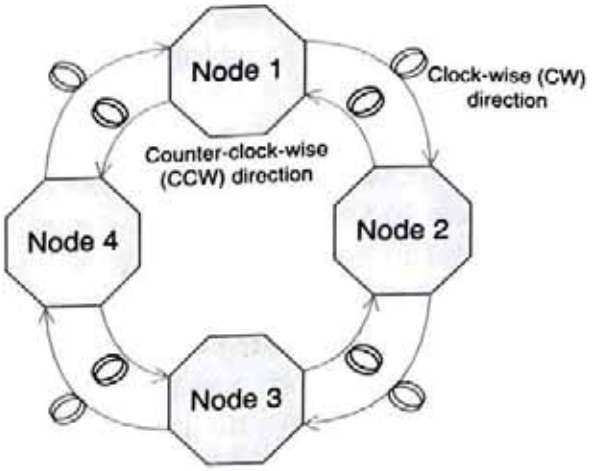


Figure 12.6 Two-fiber (2F) optical ring architecture (this figure may be seen in color on the included CD-ROM).

G.872). The scheme deployed vary based on ease of implementation and changes in demand patterns: in SONET/SDH networks, the most common protection is UPSR followed by BLSR for some core networks. In the DWDM layer, simple $1+1$ protection is typically implemented, which is equivalent to BPSR in the above notation.

It should be noted that the outlined protection approach inherently doubles the overall network bandwidth requirements relative to actual demand load. Unidirectional case allocates one fiber fully to work data, and one to protection data; bidirectional case allocates each fiber's bandwidth to work/protect in a 50%/50% split.

Figure 12.6 provides a very high-level view of the metro network ring architecture. Actual protection switching within a node can also be done in many different ways. For example, Figure 12.7 shows a few that see common implementation, in a general order of increasing cost. Figure 12.7(b) shows an implementation that protects against fiber cuts only, but minimizes hardware requirements. Figure 12.7(c) shows an implementation that both protects against fiber cuts and provides transport hardware redundancy, but requires only a single connection to the client equipment. Finally, Figure 12.7(d) shows that protection may be implemented at the electronic router/switch level, while increasing the required size of the electronic fabrics. All of these approaches fall into a general category of $1+1$ protection schemes, i.e., each work demand has a dedicated corresponding protection demand through a geographically disjoint route, and all approaches are capable of providing protection within a 50-ms SONET/SDH requirement.

All of the above approaches require an effective doubling of the network capacity, while providing protection only against a single route failure. The demands placed on the networks continue to grow in geographic extent, number of interconnected nodes, and in an overall network demand load [9]. As the size grows, a single ring implementation may not be practical from reliability and bandwidth capacity perspectives. The network may still be partitioned into a set of

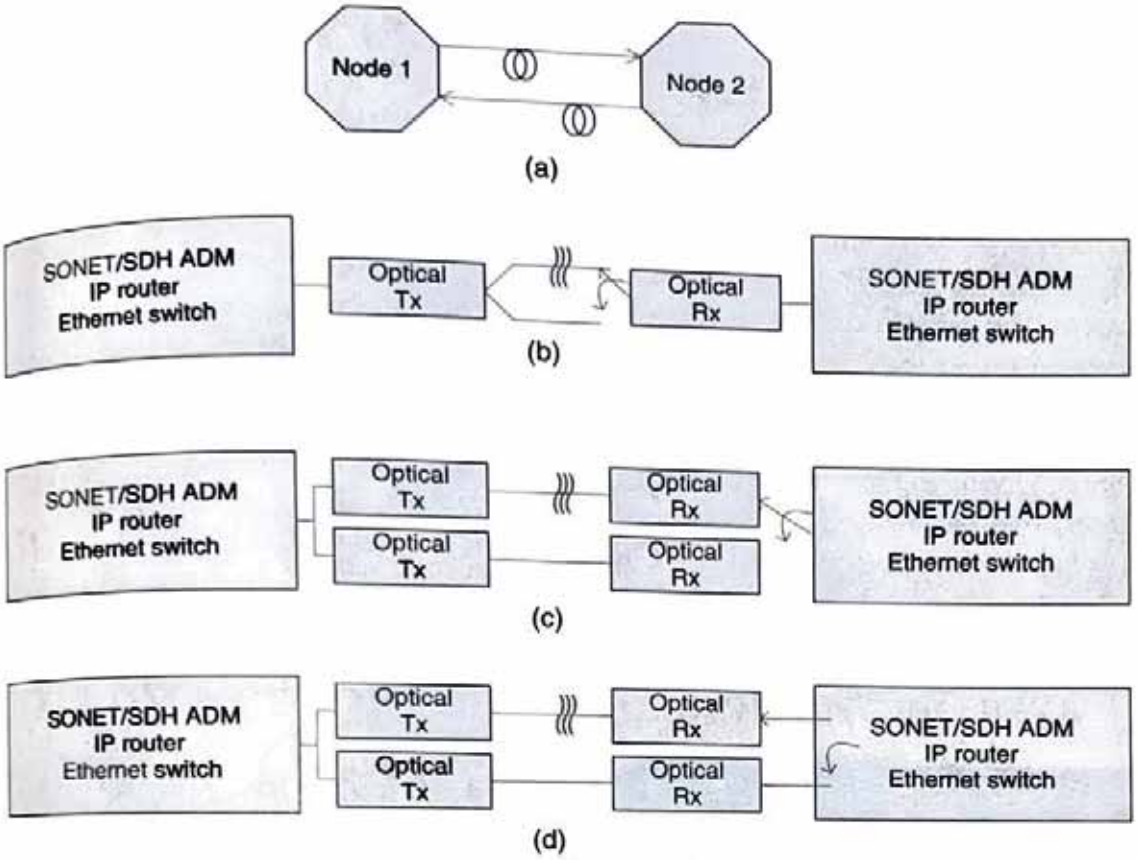


Figure 12.7 Protection switch options within a Node: (a) Generic node 2F connection, (b) Line-side optical switch, (c) client-side optical switch, (d) Line Terminating equipment switch (this figure may be seen in color on the included CD-ROM).

ring connections, with ring-to-ring interconnections. A single ring-to-ring connection will look like a single point of failure in a network, and rings may need to be joined at multiple points: physically the network starts to look like a mesh arrangement of nodes, as shown in Figure 12.8.

The rich physical connectivity of a mesh network is obvious as a node-to-node demand connection may take many diverse routes through the network. This diverse connectivity offers an opportunity to significantly improve network's utilization efficiency. Recall that a fully protected ring-based network required dedicated doubling of its capacity relative to the actual bandwidth. A shared protection scheme allocates protection only after a failure has occurred. Thus, assuming that a network suffers only a small number of simultaneous failures, and that a rich physical network connectivity affords several route choices for the protection capacity, each optical route needs to carry only an incremental amount of excess protection bandwidth [10, 11] reduced by a factor of $1/(d\text{-}1)$, where d is the average number of diverse routes connected to nodes. In addition to reduced cost, another benefit of this approach is an ability to gracefully handle multiple network failures. The actual capacity is consumed only after a failure is detected. However, the trade-off is a substantially more complicated restoration algorithm

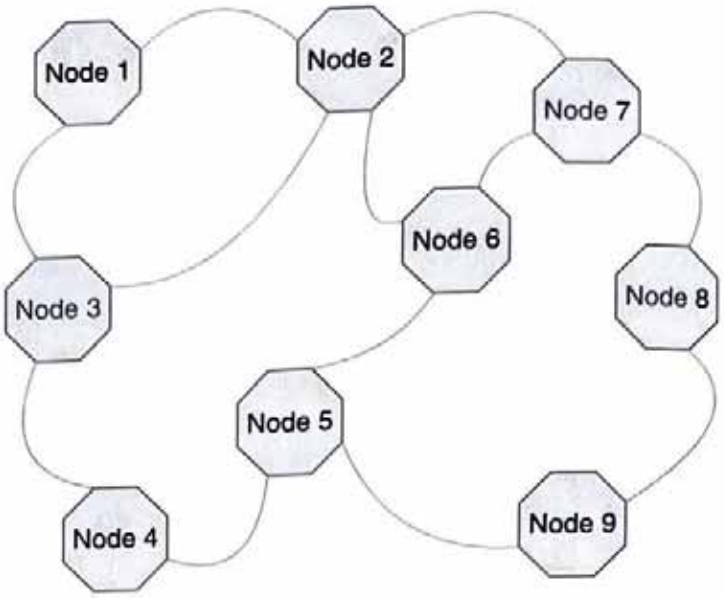


Figure 12.8 Network with a large number of interconnected nodes takes on a "mesh" appearance (this figure may be seen in color on the included CD-ROM).

that now requires both network resources and time to compute and configure a protection route [12]. Further, re-routing is done via electronic layer, not optical one, given today's status of optical technology.

As the service layer moves to packet-based devices, new protection mechanisms are being considered. Examples include MPLS fast reroute (FRR), RPR, as well as Ethernet convergence. Out of these mechanisms, RPR is the closest to SONET protection, in that it is mainly confined to rings and loops around the ring in the event of a failure. However, since RPR uses statistical multiplexing, it is able to drop traffic that has low priority depending on the actual amount of high priority traffic currently in the network, whereas the optical layer was limited to protecting the total working bandwidth irrespective of the actual usage of the bandwidth. This added flexibility allows SPs to offer a large number of services, while with SONET the only service that was available was a fully protected service (99.999%). It is worth noting that some SPs tried to also offer a preemptible service using protection bandwidth in SONET, but since this bandwidth was frequently preempted, the service was only useful for niche applications.

MPLS offers a more flexible mechanism that is not restricted to rings, based on a working path and a predefined protection path. Again, thanks to statistical multiplexing, the bandwidth along the protection path does not have to be reserved to a particular working connection, but rather is used by the traffic that requires protection based on priorities. While the packet level mechanism is simple, this scheme does require planning to ensure protection bandwidth is not oversubscribed to a point that the service level agreement cannot be guaranteed.

Finally, Ethernet also offers a convergence mechanism that ensures packet can be forwarded along a new spanning tree should the original spanning tree fail. However, this mechanism is typically slow and does not scale to larger networks.

Given these protection mechanisms, what is the role of the WDM layer in protecting traffic? Quite a bit of research has been performed on how WDM protection can coexist with service layer protection and how the layers coordinate and benefit from the respective protection mechanisms [13, 14]. However in reality, most SPs prefer to keep protection to one layer for simplicity. Naturally, when protection does exist in the service layer, it is more beneficial to use it, as it covers failure modes that are unrecoverable in the optical layer (such as an interface failing on the service box). As discussed, often times service level protection is more efficient driving to an overall lower cost of protection even as service layer equipment is more expensive than WDM equipment [15]. This leaves narrow room for protection at the WDM layer: typically for point-to-point applications that do not have their own protection mechanism, such as SAN applications.

12.4 WDM NETWORK PHYSICAL BUILDING BLOCKS

12.4.1 Client Service Interfaces

Metro networks generally see traffic that has already gone through several levels of aggregation multiplexing, and equipment interconnections are done at 1-Gb/s data rate and above. By definition, these are meant to connect equipment from a wide variety of manufacturers, and several international standards (as well as industry-wide Multi-Sources Agreements) have been developed [16, 17] that cover optical, mechanical, electrical, thermal, etc. aspects. Given the required high data rate and the physical connectivity length, metro equipment client interconnections are almost exclusively optical.

Early systems had interface hardware built from discrete components. A highly beneficial aspect of developing and adhering to standards is an ability of multiple vendors to provide competitive interchangeable solutions. Over the last several years, the optoelectronics industry has seen a tremendous amount of client interface development, and a near complete transition from custom-made interfaces to multi-sourced, hot-pluggable modules. The interfaces have evolved [18] from GBICs offered at 1-Gb/s data rate [19] to SFP for multirate application up to 2.5-Gb/s data rate [20] to XFP at 10 Gb/s [21]. The economics of manufacturing are such that it is frequently simpler to produce a more sophisticated component and use it across multiple applications. Further, providing a single electrical socket on the equipment allows the interfaces to be reconfigured simply by plugging in a different client module, with same interfaces being able to support SONET/SDH, Ethernet, FC, etc. applications. Figure 12.9 shows a comparison of relative physical form factors for a variety of pluggable interfaces targeting 10-Gb/s data rate, and the rapid relative size reduction over the course of a few years.

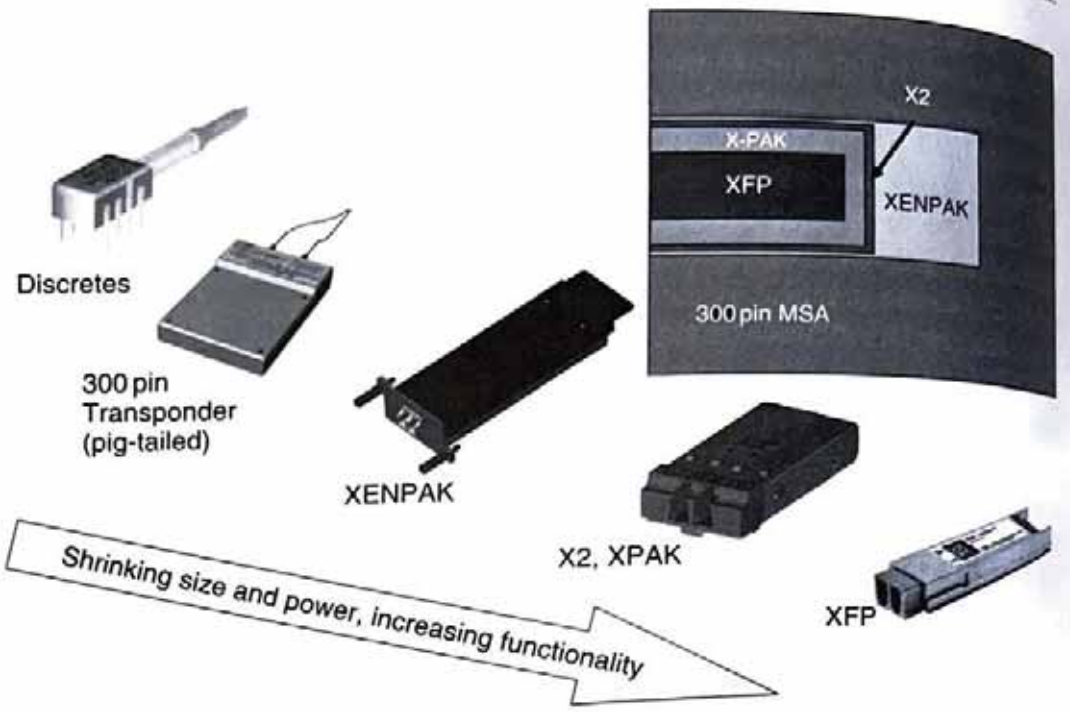


Figure 12.9 Evolution of client size interface form factors (this figure may be seen in color on the included CD-ROM).

12.4.2 WDM Network-Side Optical Interfaces

The requirements on the WDM (network-side) of the optical system are more stringent than on the client interfaces. While client interfaces need connectivity over a relatively short distances, with 90% falling within 10-km distance, network side interfaces need to cover distances of hundreds of km and many demands are multiplexed on the same fiber using WDM.

The optical wavelength of client side interfaces has a wide tolerance range and uncooled lasers are most often used. The WDM side, due to multichannel requirement, has lagged the client side interface in size development. Initial network side pluggable modules were developed in GBIC form-factor for 2.5-Gb/s application, and used uncooled lasers for CWDM with relaxed 10 nm wave separation. More recently, 2.5-Gb/s interfaces with DWDM (100-GHz) channel spacing were implemented in SFP form factor, too. As networks evolved to support higher bandwidth services, 10 Gb/s network side interfaces were implemented in 300-pin MSA form factors [22]. Subsequently, MSA modules evolved to support 50-GHz channel spacing with full tunability across all C-band wavelengths. Tunable laser technologies have being increasingly employed in Metro WDM systems to reduce inventory cost, and improve operations [16]. The choice of the appropriate transmitter technology is particularly important, as its cost usually dominates the total cost of a fully deployed transport system [23]. At the same time, much smaller XFP

packages supporting 10-Gb/s WDM interface at 100-GHz channel spacing were developed using lasers without wavelength locking. Such next-generation pluggable transmitters are very important, not only for their enhanced performance or lower cost, but also for more easily integrating into the different service platforms further simplifying the network architecture and thus reducing the overall network cost. At such high data rates, however, optical performance, predominantly dispersion-tolerant (chirp-minimized) modulation, becomes also important. Current development efforts are pursing fully C-band tunable 50-GHz WDM interfaces in XFP form factor. Higher data rate 40-Gb/s interfaces currently require a larger package, but are following the same general trajectory of rapidly decreasing size and increasing capability.

In addition to the extremely rapid advances in the optical technology developments, electronics technology is also providing increased performance, and reduced size and power consumption. Of particular interest are the field programmable gate array (FPGA) technology and the FEC technology. Metro networks are generally called on to support a rich variety of services, such as SONET/SDH, Ethernet, and FC. The protocols and framing formats, overhead, and performance monitoring parameters are very different, while intrinsic data rates may be quite close. Since client interfaces are pluggable, it is highly desirable to provide a software-configurable, flexible electrical processing interface such that a single hardware circuit pack can be field-reconfigured to support different services. FPGA elegantly fulfills such a role, providing high gate count, and low-power and high-speed capability with 65-nm CMOS geometries available in 2007, and 45-nm CMOS geometries expected to be available in 2009–2010 time frame. At the same time, FEC and increasingly enhanced FEC enable much more flexible transmission performance.

12.4.3 Modulation Formats for Metro Networks

Non-return-to-zero on-off keying (NRZ-OOK) is arguably the simplest modulation format to implement for WDM network signal transmission. NRZ-OOK is the format with the widest deployed base of commercial systems, given that excellent propagation characteristics can be achieved with quality implementations having good control over rise/fall times, limited waveform distortions, and high optical extinction ratio. The longer (300 km) demand reach requirements imply that both fiber dispersion and loss become quite important. The intrinsic dispersion tolerance is determined by the modulation format and data rate. For example, NRZ-OOK at 2.5 Gb/s has an intrinsic dispersion tolerance of ~17,000 ps/nm, corresponding to ~1000 km of NDSF fiber. The fact that this is well above reach "sufficiency" for Metro networks permits an engineering trade for a lower-cost, lower-quality implementation. Relaxing transmitter chirp control can lower costs substantially, while still allowing for a dispersion tolerance in 1600-ps/nm to 2400-ps/nm range (i.e., 140 km of NDSF fiber).

Intrinsic optical transceiver characteristics are set by launch power, dispersion tolerance, receiver sensitivity and ASE tolerance. A network must satisfy this multi-dimensional demand simultaneously and preferably with a single modular implementation. Metro networks geographic characteristics and traffic demands cover a broad range from a few tens of kilometers up to 200 km. Such wide range of networks implies that there is no generic target characteristics: networks may be limited by any combination of the above mechanisms, depending on demand reach length, number of nodes, fiber characteristics, etc. Even within the same network some demands may be power-limited, while others may be limited by dispersion, while others may be limited by ASE noise. A desire to minimized network hardware costs, while still supporting the demands, poses a challenging optimization problem.

Unamplified links have two dominant limiting characteristics, and performance is typically expressed in terms of receiver power penalty as a function of dispersion. Noise determined by receiver electronics is independent of input signal power, and affects both "0" and "1" signal levels equally. System performance must be kept above a threshold line defined by a minimum receiver power required to achieve desired bit error rate (BER) performance at a set dispersion. Optically amplified system noise is primarily determined by the beating between signal and ASE components, and impacts primarily "1" level for OOK modulation formats [24, 25]. Its functional form is different from a direct power penalty, and systems need to consider three characteristics: received power, dispersion and OSNR. Amplified systems performance can be expressed in terms of two-dimensional power and dispersion surface, and shown as a target OSNR required to achieve a certain BER, with an example shown in Figure 12.10.

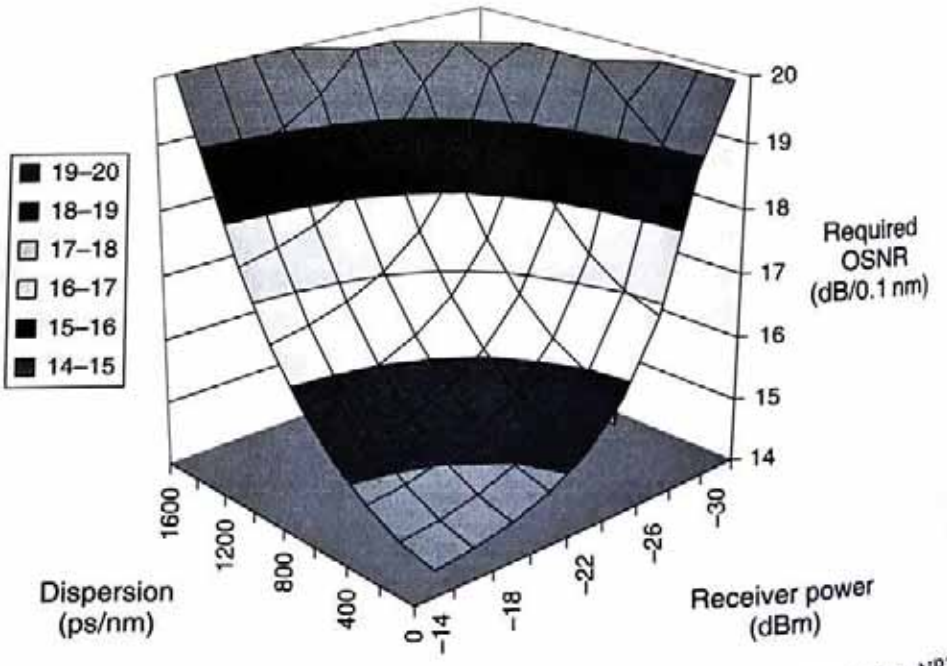


Figure 12.10 Target OSNR required to achieve a BER $\sim 10^{-7}$ for a typical 10-Gb/s NRZ-OOK implementation, which assumes a receiver sensitivity of –28 dBm at 10^{-9} BER, and an OSNR sensitivity of 10 dB/0.1 nm at 10^{-3} BER (this figure may be seen in color on the included CD-ROM).

Optically amplified, ASE-limited systems exhibit an inverse relationship between required target OSNR and unregenerated optical reach. Thus, a 1-dB increase in required OSNR produces a corresponding reduction in unregenerated reach, and can easily cross the network performance threshold.

12.4.4 Handling Group Velocity Dispersion

As service demands pushed network transport rates to 10 Gb/s, the requirement to cover an identical network geographic extent remained unchanged. Current 10-Gb/s transmission is still most frequently done with NRZ-OOK format, which intrinsic dispersion tolerance is 16 times smaller than equivalent 2.5 Gb/s (i.e., ~1200 ps/nm for unchirped versions). 10-Gb/s data rate crosses the threshold of dispersion tolerance for many Metro networks, and some form of dispersion compensation is required. In-line dispersion compensating fiber (DCF) is the most commonly deployed technology. DCF is very reliable and completely passive, has a spectrally transparent pass band compatible with any format and channel spacing, and can be made to compensate dispersion across the full spectral range for most deployed transmission fiber types. The disadvantages are added insertion loss that is most conveniently "hidden" in the optical amplifier midstage, nonlinear effects requiring controlled optical input power, both of which degrade overall link noise figure. Existing networks cannot be easily retro-fitted with DCFs without significant common equipment disruption.

New ways of handling dispersion based on transponder-based technologies are advantageous for seamless network upgrades. Particularly, transmitter-side modulation formats [26] or receiver-side electronic distortion compensation (EDC) are attractive, if they can be made economically viable. Detailed characteristics and implementations for a variety of modulation formats are described in Chapter 2, and electronic distortion compensation is also described in Chapter 18 of Volume A [27, 28]. However, it is instructive to note some options that have been specifically applied to metro networks: prechirped NRZ-OOK modulation, duobinary modulation, and receiver-side EDC. With both prechirped and duobinary modulation formats optimal dispersion is shifted to higher values, but performance may be degraded for other parameters, such as ASE tolerance or overall dispersion window. Receiver side equalization, whether electronic or optical, is a technology that introduces additional cost, power consumption, and board space. A decision to use transceiver-based dispersion compensation is not simple or universal, though EDC is finding good adoption especially on the client interfaces.

Modulation formats may mitigate some of the dispersion problems. However, the desire to upgrade existing field-deployed networks to support higher channel data rates transport is impeded by several considerations: (1) higher rate signals require OSNR increase of 3 dB for each rate doubling (assuming constant format and FEC), (2) receiver optical power sensitivity increasing by 3 dB for each data

rate doubling, (3) dispersion tolerance decreasing by 6 dB for each data rate doubling, (5) in-line optical filtering, and (4) system software upgrades.

12.4.5 Optical Amplifiers

Optical signal loss in metro networks accumulates from three components: transmission fiber, optical components embedded in the frequent add/drop nodes, and passive fiber segment connections. Metro networks are generally deployed in very active environments, and while not very long, they are subject to frequent mechanical disturbances and breaks. Metro network fiber accumulates passive loss due to frequent repairs and splicing.

There are three main optical amplifier choices available for overcoming the loss: erbium-doped fiber amplifier (EDFA), semiconductor optical amplifier (SOA), and Raman amplifier. Of these, EDFA provides a cost-effective solution to overcome loss in the network with good performance. SOA amplifiers have the advantage of wide amplification bandwidth, but have more limited output power, susceptibility to interchannel crosstalk issues, and high noise figure. Distributed Raman amplifiers' primary benefit is in reducing effective amplifier spacing which lowers the overall link noise figure. Metro nodes are already closely spaced and "distributed" benefit is small. Distributed Raman also relies on transmission fiber quality, and passive losses and reflections have a substantially deleterious effect [29], making use of Raman amplifiers in metro quite rare.

Amplifier response to optical transients is as critical a characteristic as gain, noise figure, output power, and spectral flatness. Rich traffic connectivity patterns require a high level of immunity to possible optical breaks. Figure 12.11 shows an

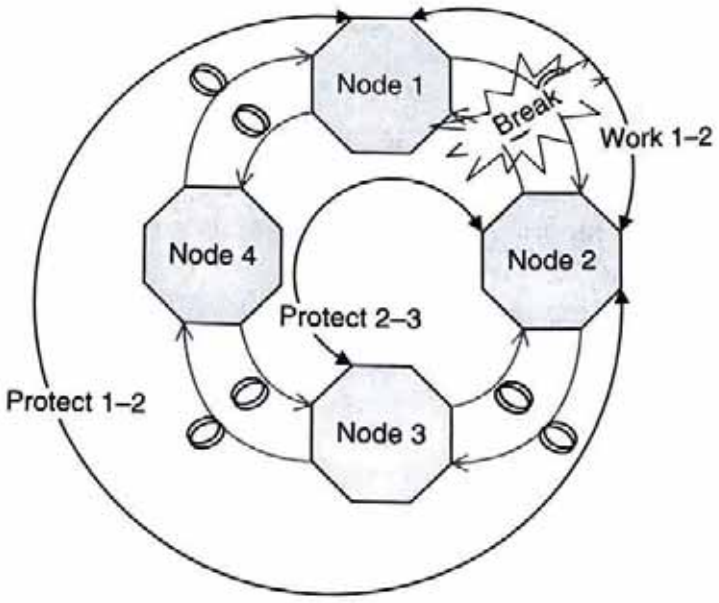


Figure 12.11 Optical interaction between optical work and protect routes (this figure may be seen in color on the included CD-ROM).

example whereby a break between Node 1 and Node 2 causes a loss of optical Work 1–2 and Protect 2–3 channels. Protect 2–3 channels are optically coupled to Protect 1–2 channels, and any disruption on Protect 1–2 channels will cause a corresponding loss of connectivity between Node 1 and Node 2.

Table 12.1 below shows a table of typical effects that can lead to EDFA transients and associated time constants. The same physical properties that make EDFA excellent for multichannel transmission with negligible crosstalk also make it hard to suppress optical transients. Techniques based on optical reservoir channels suffer from having to add extra optical hardware, and the fact that EDFA dynamics are spectrally dependent, i.e., loosing channels at short wavelengths does not have the same effect as adding a reservoir channel at long wavelengths. The same is true with gain clamping via lasing [30]. The most cost-effective strategy, and one that has seen actual field deployment, is using optical tap power monitors on the amplifier input and output ports, with electronic feedback to the pump lasers. The electronic feedback loop can be made quite fast, but pump laser electrical bandwidth and intrinsic EDFA gain medium dynamics limit possible control speed. A simple addition of a controlled attenuator is insufficient to guarantee flat spectral output under different channel load.

Figure 12.12 shows an example measurement of an electronically stabilized EDFA amplifier transient response to an optical step function, with step function fall time as the parameter. Several temporal regions can be identified. First, there is an optical power increase due to a redistribution of optical power into surviving channels, with a rise time corresponding to the optical channel power loss fall time. Second, the pump power is rapidly reduced by the electronic feedback control, and the channel power recovers with a time constant set by the control loop dynamics. Third, depending on the parameters of the control loop and their interplay with optical dynamics, there may be some amount of transient undershoot and possibly ringing. Finally, a steady state is reached that is likely to have some finite error in the channel power due to electronic errors, due to finite broadband ASE power, etc.

The transient are shown for a single amplifier, while real systems employ many cascaded amplifiers in a route. Each subsequent amplifier will see not only the

Table 12.1

Table of possible optical transient time constants, in order of increasing speed.

3 mm jacketed fiber slow bending	500 ms
3 mm jacketed fiber caught in shelf cover	100 ms
3 mm jacketed fiber fast bending	7 ms
3 mm jacketed fiber wire stripper cut	5 ms
Connector E2000 fast unplug	2 ms
Bare fiber wire stripper cut	200 µs
Bare fiber knife chop cut	100 µs
Bare fiber (break at splice using tensile force)	2 µs

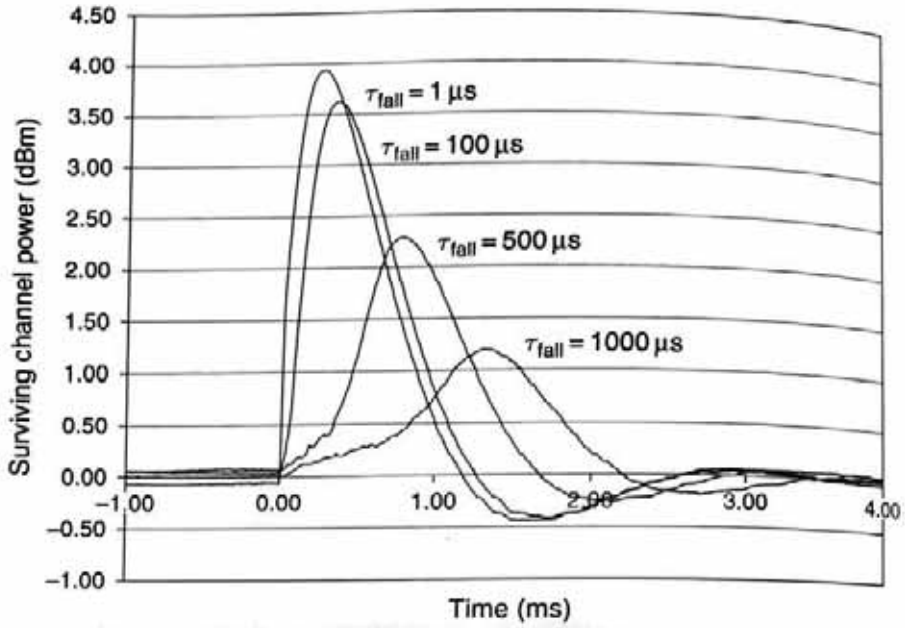


Figure 12.12 Example amplifier optical transient response, with optical channel loss fall time as a parameter.

original step-like loss of the optical signal, but will also experience the accumulated effects of all of the preceding amplifiers. Thus, control loop parameters have to be developed and verified on amplifier cascades, in addition to individual modules [31].

It is fundamentally not possible to completely prevent and eliminate optical transients, and these impact optical channel performance in several ways. Increasing optical power may lead to nonlinear fiber effects and corresponding distortions in the received signal. Low power leads to a decrease in optical SNR. Further, power may exceed the dynamic range of the receivers, and may interact with the dynamic response of the receiver electrical amplification and decision threshold mechanisms. These combined effects may lead to burst errors in the optical channel. Protection switching mechanisms need to be designed with corresponding hold-off times to prevent such events from triggering unnecessary switching.

12.4.6 Optical Add/Drop Nodes

WDM multiplexing and connectivity provides some of the functionality previously supported at the SONET layer, such as multiplexing and circuit provisioning. As discussed previously and shown in Figure 12.11, networks nodes may be arranged on an optical fiber ring. However, demand connections may be such that an optical channel bypasses a node and stays in the optical domain. OADMs accomplish this function. Most of the metro networks were deployed with OADMs implemented with fixed optical spectral filters. The filters are positioned within the optical path

along the ring to both drop and add signals of pre-determined wavelengths. This provides a very cost-effective access to the optical spectrum, minimizes signal insertion loss, and allows for possible wavelength reuse in different segments of the optical ring or mesh network. In a sense, predeployed optical filters associate a specific destination wavelength range address with each node. Optical filters themselves may be based on a wide variety of technologies, and are beyond the scope of this chapter.

Fixed filter based nodes are cost-effective, have low optical loss, and simple to design and deploy. However, they pose an operational challenge due to the intrinsic lack of reconfigurability. Each node must have a predesigned amount of capacity (i.e., wavelength address space) associated with it and cannot be changed without significant traffic interruptions. For example, some portions of the network may experience more than expected capacity growth and may require additional spectral allocations, which would be impossible to achieve without inserting additional filters into the common signal path, thereby interrupting traffic. Others portion of the network may lag expected growth, and will thus strand the bandwidth by removing unused spectrum from being accessible to express paths. Network deployment with fixed filters require significant foresight into the expected capacity growth, and several studies have addressed the question of what happens when actual demands deviate from the expectations [32], with as much as half of the overall network capacity possibly being inaccessible. One of the frequently proposed techniques to overcome such wavelength blocking limits is the use of strategically placed wavelength converters in the network [33, 34], but is quite expensive in terms of additional hardware that must be either predeployed or require as-needed service field trips.

An alternative to deploying wavelength converters is to provide dynamic reconfigurability that is generally associated with electrical switching directly at the optical layer. Developments in optical technology have allowed a new level of functionality to be brought to the OADMs, and fall under a general term of reconfigurable OADM (ROADM). It should be pointed out that while ROADMs substantially reduce wavelength blocking probability [35, 36], they cannot completely eliminate it, especially if all wavelengths remain static after assignment. Some amount of wavelength conversion or dynamic wavelength retuning may be required (see Chapter 8).

A variety of optical ROADM node architectures can be considered, depending on the particular goals of the network designer [37, 38]. These architectures can be subdivided into three broad categories. The first category can be described as a space-switch-based architecture surrounded by MUX/DEMUX elements. An example for a Degree 2 node is shown in Figure 12.13(a). With the current state of the art, the implementation is done with integrated MUX/DEMUX, add/drop switch, and direction-switch elements. The channelized aspect of the architecture introduces optical filtering effects into the express path, thereby limiting bit rate and channel spacing transparency. Increasing the connectivity degree of the node requires a change in the configuration by adding either a new level of integration

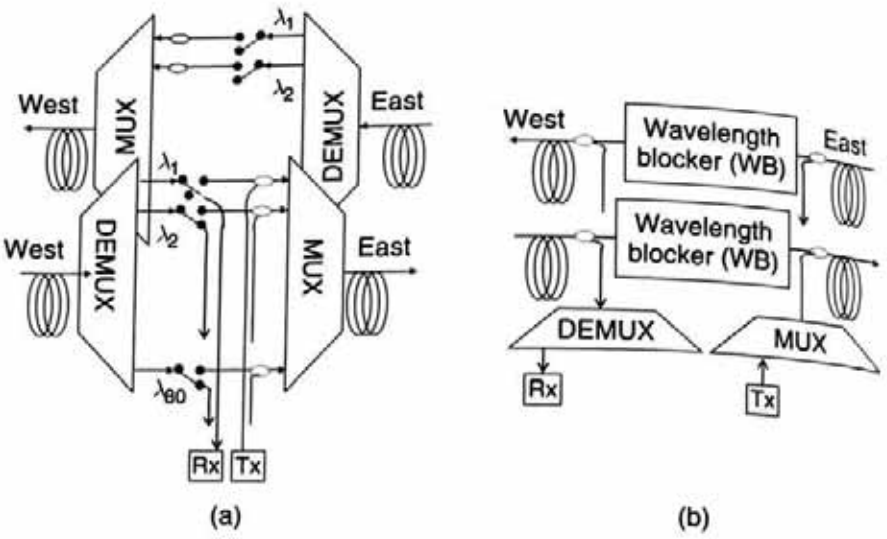


Figure 12.13 (a) Degree 2 space-based OADM; (b) Degree 2 broadcast-and-select OADM.

of additional MUX/DEMUX and switch elements, or externally interconnecting smaller building blocks of Figure 12.13(a) with additional external switching.

The second category can be described as broadcast and select architecture, and is shown in Figure 12.13(b). The architecture relies on an integrated wavelength blocker (WB), which can provide arbitrarily selectable pass and stop bands with continuous spectrum and single-channel resolution. The express path continuous-spectrum attribute reduces channel filtering effects, improves cascadability, and permits a high level of bit rate and channel spacing transparency. One disadvantage of the broadcast and select architecture is its requirement of a separate MUX/DEMUX structure to handle local add/drop traffic, which adds cost and complexity. A second disadvantage is that scaling to a higher Degree N node interconnect requires $N \times (N-1)$ WB blocks, with a Degree 4 node requiring 12 WB blocks.

More recently, a third ROADM architecture has been introduced that attempts to combine a high level of integration and express transparency associated with a broadcast and select architecture, with an integrated MUX/DEMUX functionality [39]. The architecture, shown in Figure 12.14, is based on an integrated multi-wavelength, multiport switch (MWS) and is particularly attractive for metro-type applications that are susceptible to frequent traffic and node churn, but have moderate bandwidth requirements. MWS elements have several output ports (in the 4–9 range) and can selectably direct any combination of wavelengths to any output ports. A receiver can be connected directly to the MWS if only a small number of add/drop wavelengths is required, or a second level of DEMUX can be implemented to increase the add/drop capacity. The second level itself can be based on a low-cost fixed architecture, or a more complex wavelength-tunable one.

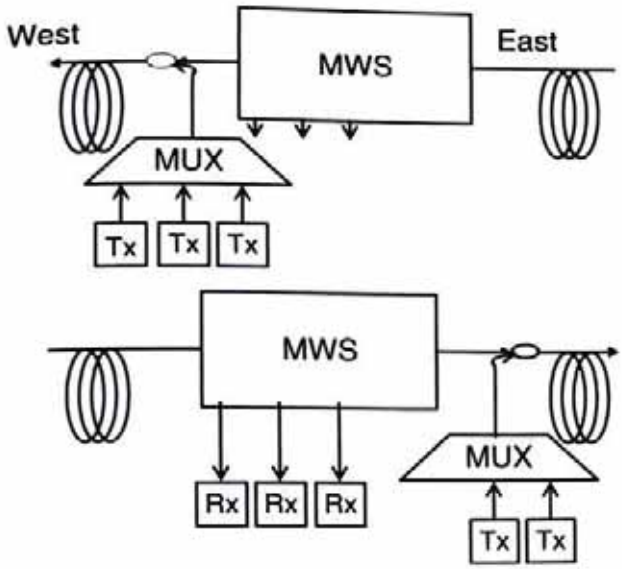


Figure 12.14 Reconfigurable OADM Architecture.

Table 12.2

Comparison of three optical ROADM architectures.

Parameter	Space switched	MWS + MWS DEMUX	MWS + fixed DEMUX
Cost	Independent of # A/D channel	Same up to 8 ch, 5 × at 40 ch	Same up to 8 ch, 2 × at 40 ch
Optical channel monitor	Built-in	Usually external	Usually external
Channel power equalization	Yes	Yes	Yes
Multidegree	Highly integration dependent. Difficult in-service growth	Up to Degree 8 (in service)	Up to Degree 8 (in service)
0.5 dB passband	~40 GHz	~50 GHz	~50 GHz
Express loss	<12 dB	<13 dB	<13 dB
Flexibility	Add/drop ports have fixed wavelengths	Drop ports are fully tunable. Adds are wavelength independent	Drop ports are commonly fixed wavelengths. Adds are wavelength independent

Table 12.2 shows a comparison table of possible space switch and MWS-based architectures, and associate trade-offs.

Recently, the optoelectronics industry has evolved a new direction toward higher levels of functional integration, which has the potential to reduce the cost

of OE and EO conversion (Chapters 6 and 10 of Volume A). Possible availability of such low-cost optical interface components has prompted a re-evaluation of the existing trade-offs between optical and electrical switching, and proposals for a "Digital ROADM" have been presented (see Chapter 10 of Volume A). A Digital ROADM is more accurately termed a reconfigurable electrical add/drop multiplexer. It is an electrical switching fabric very similar in functionality to the legacy SONET/SDH add/drop Multiplexers or to an Ethernet switch, but updated to operate on signals compliant to the recently developed OTN framing standard. Highly integrated photonics indeed hold a promise of reduced costs; unfortunately, optical interfaces comprise only a small fraction of the overall node cost. The impact of electronically processing every digital bit stream holds both advantages of being able to completely regenerate wavelength signals, monitor bit stream quality, and improve their grooming efficiency, as well as detrimental aspects of substantially increased electronic power consumption and mechanical footprint. Table 12.3 captures the more salient comparison points, and argues that metro networks still preferentially benefit from optical OADM rather than electrical one.

In summary, ROADM has introduced network design flexibility, and automated and scalable link engineering, both critical to the success of metro WDM architectures. Wavelength-level add/drop and pass-through, with automated reconfigurability (ROADM) at each service node, is also only operationally robust solution for WDM deployments that support the uncertain (often unpredictable) future traffic patterns in MANs that scale to hundred of Gb/s. ROADM network flexibility also provides the ability to set up a wavelength connection without visiting any intermediate sites, thus minimizing the risk of erroneous service disruptions during network upgrades. In the context of the present analysis, it is also useful to identify and distinguish between two main functional characteristics of ROADMs at the network level: (1) ROADM solutions allow for switching of each individual wavelength between the WDM ingress and egress, potentially among more that one fiber facility. (2) More elaborate solutions could also allow extraction or insertion of any client interfaces to any wavelength of any fiber. This latter solution has been often proposed, in combination with predeployed tunable transmitters and/or receivers, to realize advanced network automation. Such a network, with the addition of a GMPLS control plane, enables dynamic bandwidth provisioning, and fast shared optical layer mesh protection. Current network deployments, however, are primarily interested in the ROADM functionality and the most cost-effective-related technologies (rather than in the most advanced solution that would meet any conceivable future need, irrespective of price). In this sense, the technologies captured in the above table, are currently the main focus of network deployments, as they have sufficient functionality to meet most customer needs, and are the most mature and thus cost-effective technologies [40].

Table12.3

Comparisons of reconfigurable electrical and optical add/drop nodes.

Parameter	Electrical ADM	Optical ADM
Optical mux/demux	First level external optical Second level integrated within Rx	First level MWS switch Second level fixed filter or MWS
Channel monitor	Full electrical PM	Analog optical wavelength power
Channel regeneration	Yes, but not critical for metro	Only power equalization, OK for metro
Multidegree	Requires highly sophisticated switching fabric with full nonblocking interconnect capability	MWS provide direct degree interconnects, but wavelength blocking may exist
Optical line amps	Requires Rx-side and Tx-side OLA to deal with MUX/DEMUX loss.	Requires Rx-side and Tx-side OLA to deal with MUX/DEMUX loss
Per-channel power consumption	~50–100 W per 10 Gbps data stream (estimated average from published Ethernet and OTN switch fabrics)	~2.5 W per 10 Gbps wavelength (<100 W for OLA + MWS)
Granularity	Provides subwavelength switching capability	Wavelength-level switching capability
Subrate channel flexibility	Subrate channels electrically multiplexed to wavelength	Subrate channels electrically multiplexed to wavelength.
Super-rate channels	Must be inverse multiplexed across several wavelengths. For example, 40 Gbps services occupy 4 × 10 Gbps λ's, and may cross integrate part boundary	May be inverse multiplexed. Or may used new modulation format technology for direct transport over existing line system
Total system capacity	Fixed on Day 1 install	Can grow as new XCVR technology is introduced to populate unfilled spectrum, i.e., improved FEC, modulation format, equalization
Relative cost of a high capacity ADM node	Assuming OEO interfaces are low-cost, high-capacity electrical FEC, framing and switching fabric expected to dominate costs	OEO interfaces maybe relatively higher, but optical switch fabric is much lower cost than comparable electronic one

12.5 NETWORK AUTOMATION

Network design tools are mandatory for proper design, deployment, and operation of an optical WDM network, especially when large degree of reconfigurability is deployed. Network planning process focuses an optimized network design for efficient capacity utilization and optimum network performance for a given service load. Such design and planning tools must combine a set of functions that span multiple layers of the network operation, from a definition of user demands, to service

aggregation and demand routing, and finally to the physical transport layer. Higher (logical) network layers need not worry about the physics of light when calculating paths across a network, with a possible exception of physical latency associated with the connection. However, the physical layer must include important optical propagation effects when deciding demand routing and capacity load attributes.

A typical network design process, graphically illustrated in Figure 12.15, includes several steps. First, a set of "Input" parameters is formed by a combination of logical user service demands and by an abstracted layer describing known physical connectivity and limitations. The inputs may also include a description of existing network configuration, which may be automatically uploaded from field-deployed hardware, or may be a completely new installation. Second, service demands are aggregated into optical wavelengths, considering actual electronic hardware limitations, user-defined constraints, and required network performance. Third, aggregated wavelength demands and a refined definition of the physical layer (i.e., definitions of specific fiber types, lengths, optical losses) are provided as an input to drive physical network design process. The result of this multi-step process is a complete description of the network hardware, with a detailed description of the Bill of Material, deployment process documents and drawings, as well an estimate of the network performance characteristics.

The actual software implementation may partition the overall process into relatively independent modules, with each step performed sequentially. An alternative is to provide coupling between each step to allow a higher level of network optimization and refinement. The software itself may be an off-line design/planning tool driven by a user interested in targeting substantial network configuration

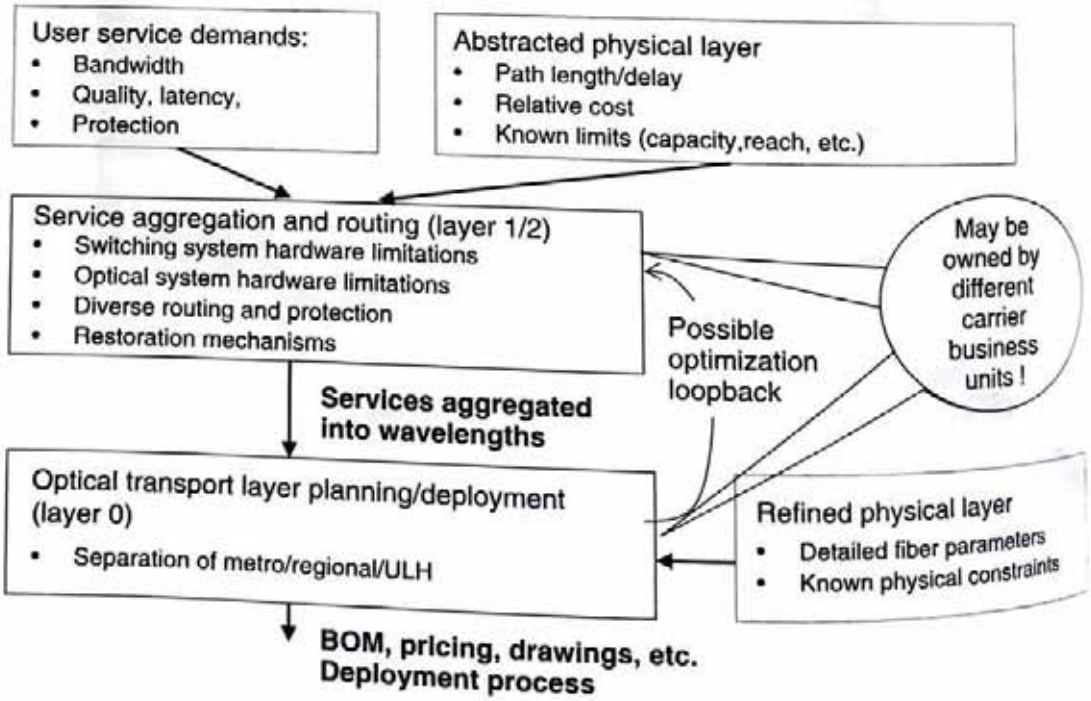


Figure 12.15 Network planning and design process.

changes or upgrades. A similar process and software may also be used for handling new service demand requests that may arrive to the automatically switched optical networks, with the process triggered when a single new service demand arrives to the network and must be satisfied within the shortest amount of time and within constraints of null or minimal hardware change.

The geography of the metro networks is frequently characterized by a wide variability in traffic-generating node separation, which may range from sub-km range for nearly co-located customers to 100 km, and possibly longer. Premise space and power availability, traffic add/drop capacity requirements are also quite variable. Networks with such diversity benefit from a highly modular system transmission design, whereby components such as add/drop filters, optical amplifiers, dispersion compensation modules, and other signal conditioning elements are independent of each other and are deployed on as needed basis. Especially considering the case of closely spaced nodes, the decision to deploy optical amplifiers and dispersion compensation modules cannot be made based on a purely "local, nearest node" basis. The presence of optically transparent degree 3 and higher nodes complicate the configuration process even further by optically coupling multiple network segments and even coupling directions. For example, a purely linear bidirectional system has independent optical propagation for the individual directions. However, a T-branch network geometry couples East-to-West and West-to-East directions, since both share a common Southbound path. The configuration and optimization of such networks cannot be made based on simplified engineering rules in networks that are mostly focused on the cost of the solution. Fortunately, the field of optimization algorithms is very advanced [41] and can be leveraged directly to solving the network problem. The algorithms may be additionally fine-tuned to target specific carrier requirements, such as preferentially focusing on lowest initial cost, highest network flexibility, best capacity scalability, or some other parameters defined by the carrier.

The operational aspects of reconfigurable optical networks require that same algorithms used for network design and planning be applied in-service. Whenever a network receives a new demand request, it must rapidly assess its current operational state and equipment availability, determine if the requested demand can be satisfied either directly or with some dynamic reconfiguration, and consider both logical connectivity and physical layer impairments with the newly proposed wavelength assignment and routing. This is a nontrivial problem requiring a large number of complex computations in near real time, and is currently seeing increasing research interest [42].

12.6 SUMMARY

In this chapter, we discussed innovations in network architectures and optical transport that enable metropolitan networks to cost-effectively scale to hundreds Gb/s of capacity, and to hundreds of kilometers of reach, meeting the diverse

service needs of enterprise and residential applications. A converged metro network, where Ethernet/IP services, along with the traditional TDM traffic, operate over an intelligent WDM transport layer is increasingly becoming the most attractive architecture addressing the primary need of network operators for significantly improved capital and operational network cost. At the same time, the optical layer of this converged network has to introduce intelligence, and leverage advanced technology in order to significantly improve the deployment and manageability of WDM transport. We reviewed the most important operational advancements, and the technologies that cost-effectively enhance the network flexibility, and advance the proliferation of WDM transport in multiservice metro networks.

This chapter has identified the two main trends in the transport layer of the Metro optical networks. First, there is a preference to use standards-based approaches at all layers of the network. Second, high network flexibility as demanded by the extremely rapid evolving market dynamics. Standards-based approaches allow carriers and equipment manufacturers to leverage a wide base of industry-wide efforts. Flexibility, when provided at low incremental costs, allows carriers to be "wrong" at initial network deployment, but still rapidly adapt to the ever changing and evolving markets. Software definable, flexible client interfaces allow a single network to support a rich variety of existing services, as well as to allow real-time changes as older protocols are removed and new ones are added to the same hardware. At the same time, new unanticipated services and protocol developments can be readily added without affecting installed hardware base. A different approach to the same end is a protocol-agnostic open WDM layer that allows alien wavelengths that carry traffic directly from service-optimized and integrated WDM interfaces on client platforms to operate over a common WDM infrastructure. Manufacturing advances and high volumes have made wide-band wavelength tunability a reality for the Metro space, where a single part number could be produced in extremely high quantities to cover all applications. In addition, wavelength tunability offers the promise of reconfigurability with little or no capital outlay for separate optical switch components. Fixed wavelength filters may serve as an effective "optical address" to which a tunable transmitter may be tuned to establish a particular traffic pattern. Tunable optical filters would have a similar application. Reconfigurable optical add/drop multiplexers bring a level of flexibility and optical transparency to optical switching and routing. Again, unanticipated network growth patterns, new services, and evolving channel data rates can be easily added. As new Transceiver technology is developed, overall network capacity may even be increased substantially beyond initial design parameters without requiring new overbuilds. Finally, evolving sophistication of the network planning and design tools lets carriers minimize their CAPEX costs, while at the same time retaining highly desirable flexible OPEX characteristics.

12.7 FUTURE OUTLOOK

From a transport perspective, DWDM is the main growth technology for metro networks as required bitrates on a fiber—in particular in the presence of video services—outpace the ability of single-wavelength transmission technologies to deliver the bandwidth cost-effectively. Moreover, photonic switching technologies allow for the electrical layer to scale more moderately, as much of the traffic can bypass electronics in most sites.

At the same time, Ethernet has become much more mature and robust, and is incorporating various mechanisms that will allow it to scale more gracefully—including fault detection mechanisms, hierarchical addressing schemes (such as 802.1ah and 802.1ad), as well as carrier class switch implementations. In fact, Ethernet, in its various incarnations, is gaining popularity as a replacement of SONET/SDH for sub-wavelength transport.

Finally, the IP layer has clearly become the convergence layer of most services—both for the residential and enterprise markets. IP—in particular MPLS Pseudowires, has also been used increasingly as a mechanism to converge legacy technologies such as FR and ATM over IP and as a back-haul mechanism to aggregate them into the router. Different approaches exist in terms of role of Ethernet vs IP in the metro. Some carriers see the value of Layer 3 intelligence as close to the customer as possible. This allows for efficient multicast, deep packet inspection and security mechanisms that guarantee that a malicious user will have a minimal impact on the network. Other carriers are trying to centralize Layer 3 functions as much as possible, claiming that this reduces cost and allows for more efficient management. Whatever the right mix of Ethernet and IP may be, it is clear that the network will benefit from tighter integration of the packet layer and the optical layer.

Efficient packet transport based on WDM modules in routers and switches have recently enabled the first such "converged" deployment in the emerging IP-video MAN and WAN architectures [43]. The advents in optical switching and transmission technologies discussed in Section 12.3, further allow a flexible optical infrastructure that efficiently transmits and manages the "optical" bandwidth, enabling advanced network architectures to leverage the IP-WDM convergence, and to realize the associated CAPEX and OPEX savings. These architectures would ideally be based on open "WDM" solutions (like the ones described above), so that the same WDM infrastructure can also continue to support traditional TDM traffic, as well as wavelength services or other emerging applications that may not converge over the IP network, e.g., high-speed (4 or 10 Gb/s) fiber channel. Such open WDM architectures further need to offer performance guarantees for the different types of wavelength services, including "alien" wavelengths, support mesh network configurations, and also benefit from coordinated management, and network control.

ACKNOWLEDGMENTS

The authors would like to acknowledge many colleagues at Cisco and Ciena.

REFERENCES

[1] ITU G.709 standard
[2] A. McCormick, "Service-enabled networks aid convergence," *Lightwave*, May 2007
[3] Special Issue on "Metro & access networks," *J. Lightw. Technol.*, 22(11), November 2004.
[4] L. Paraschis, "Advancements in metro optical network architectures," *SPIE*, 5626(45), 2004
[5] N. Ghani et al., "Metropolitan optical networks," in *Optical Fiber Telecommunications IV B: Systems and Impairments*, (I. P. Kaminow and T. Li, eds), Academic Press, 2002, pp. 329–403, Chapter 14.
[6] A. Saleh et al., "Architectural principles of . . . ," *J. Lightw. Technol.*, 17(10), 2431–2448, December 1999.
[7] D. Cavendish, "Evolution of optical transport technologies . . . ," *IEEE Comm.*, 38(6), 2002
[8] M. Macchi, et al., "Design and Demonstration of a Reconfigurable Metro-Regional WDM 32 × 10 Gb/s System Scaling beyond 500 km of G.652 Fiber," in *Optical Fiber Communication Technical Digest Series, Conference Edition, 2004*, Paper OTuP2, *Optical Society of America*.
[9] M. J. O'Mahony, et al., "Future optical networks," *J. Lightw. Technol.*, 24(12), 4684–4696, December 2006.
[10] S. Ramamurthy, et al., "Survivable WDM mesh networks," *J. Lihgtw. Technol.*, 21(4), 870–883, April 2003.
[11] N. Mallick, "Protection Capacity Savings due to an End-To-End Shared Backup Path Restoration in Optical Mesh Networks," in Proc. *OFC 2006*, paper NThB2, March 2006.
[12] M. Bhardwaj et al., "Simulation and modeling of the restoration performance of path based restoration schemes in planar mesh networks," *J. Opt. Netw.*, 5(12), 967–984, December 2006.
[13] P. Demeester et al., "Resilience in multilayer networks," *IEEE Communications Magazine*, 37(8), 70–76, August 1999.
[14] J. Manchester, P. Bonenfant, and C. Newton, "The evolution of network survivability," *IEEE Communications Magazine*, 37(8), 44–51, August 1999.
[15] R. Batchellor and O. Gerstel, "Protection in Core Packet Networks: Layer 1 or Layer 3?," proceedings *ECOC Conference*, September 2006, Paper Tu3.6.4.
[16] L. Paraschis, O. Gerstel, and R. Ramaswami, "Tunable Lasers Applications in Metropolitan Networks," in *Optical Fiber Communication Technical Digest Series, Conference Edition, 2004*, Paper MF 105. *Optical Society of America.*
[17] ITU-T Rec. G.957, "Optical interfaces for equipment and systems relating to the Synchronous Digital Hierarchy," 1995.
[18] http://www.schelto.com/
[19] ftp://ftp.seagate.com/sff/INF-8053.PDF
[20] ftp://ftp.seagate.com/sff/INF-8074.PDF
[21] http://www.xfpmsa.org/cgi-bin/home.cgi
[22] 300 pin Multi Source Agreement web site: http://300pinmsa.org/
[23] L. Paraschis, "Innovations for Cost-effective 10 Gb/S Optical Transport in Metro Networks," Invited Presentation in *LEOS 2004: The 17th Annual Meeting of the IEEE Lasers and Electro-Optics Society. November 2004*, Paper WU1.
[24] S. Norimatsu et al., "Accurate Q-factor estimation of optically amplified systems in the presence of waveform distortion," *J. Lightw. Technol.*, 20(1), 19–27, January 2002.
[25] J. D. Downie, "Relationship of Q penalty to eye-closure penalty for NRZ and RZ signals with signal-dependent noise," *J. Lightw. Technol.*, 23(6), 2031–2038, June 2005.

[26] A. Tzanakaki, "Performance study of modulation formats for 10-Gb/s WDM metropolitan area networks," *IEEE Photon. Technol. Lett.*, 16(7), 1769–1771, July 2004.
[27] Q. Yu and A. Shanbhag, "Electronic data processing for error and dispersion compensation," *J. Lightwave Techn.*, 24(12), 4514–4525, December 2006.
[28] T. Nielsen and S. Chandresekhar, *J. Lightw. Technol.*, 23(1), 131, January 2005.
[29] S. K. Kim et al., "Distributed fiber Raman amplifiers with localized loss," *J. Lightw. Technol*, 21(5), 1286–1293, May 2003.
[30] M. Karasek and J. A. Valles, "Analysis of channel addition/removal response in all-optical gain-controlled cascade of Erbium-doped fiber amplifiers," *J. Lightw. Technol.*, 16(10), 1795–1803, October 1998.
[31] S. Pachnicke, et al., "Electronic EDFA Gain Control for the Suppression of Transient Gain Dynamics in Long-Haul Transmission Systems," *OFC*, March 2007, paper JWA15.
[32] A. Sridharan, "Blocking in all-optical networks," *IEEE/ACM Trans. On Netw.*, 12(2), 384–397, April 2004.
[33] C. C. Sue, "Wavelength routing with spare reconfiguration for all-optical WDM networks," *J. Lightwave Technol.*, 23(6), 1991–2000, June 2005.
[34] O. Gerstel, et al., "Worst-case analysis of dynamic wavelength allocation in optical networks," *IEEE/ACM Trans. On Netw.*, 7(6), 833–845, December 1999.
[35] J. Wagener, et al., "Characterization of the economic impact of stranded bandwidth in fixed OADM relative to ROADM networks," *OFC*, March 2006, paper OThM6
[36] H. Zhu, "Online connection provisioning in metro optical WDM networks using reconfigurable OADMs," *J. Lightw. Technol.*, 23(10), 2893–2901, October 2005.
[37] S. Okamoto, A. Watanabe, and K. Sato, "Optical path cross-connect node architectures for photonic transport networks," *J. Lightw. Technol.*, 14(6), 1410–1422, June 1996.
[38] E. Iannone and R. Sabella, "Optical path technologies: a comparison among different cross-connect architectures," *J. Lightw. Technol.*, 14(10), 2184–2196, October 1996.
[39] M. Fuller, "RFP activity boosts ROADM development," *Lightwave*, April 2004. http://lw.pennnet.com/Articles/Article_Display.cfm?Section=ARTCL8ARTICLE_ID=203231&VERSION_NUM=1
[40] C. Ferrari, et al., "Flexible and Reconfigurable Metro-Regional WDM Scaling beyond 32x10Gb/s and 1000km of G.652 Fiber," in *LEOS 2005: The 18th Annual Meeting of the IEEE Lasers and Electro-Optics Society*, November 2005, TuG5.
[41] Optimization Software Guide by Jorge J. Moré and Stephen J. Wright (SIAM Publications, 1993).
[42] G. Markidis, "Impairment-constraint-based routing in ultralong-haul optical networks with 2R regeneration," *Photon. Techn. Lett.*, 19(6), 420–422, March 2007.
[43] O. Gerstel, M. Tatipamula, K. Ahuja et al., "Optimizing core networks for IP Transport," *IEC Annual Review of Communications*, 59, 2006.

13

Commercial optical networks, overlay networks, and services

Robert Doverspike* **and Peter Magill†**

*Transport Network Evolution Research AT&T Labs Research, Middletown, NJ, USA
†Optical Systems Research, AT&T Labs, Middletown, NJ, USA

13.1 THE TERRESTRIAL NETWORK MODEL

We will describe the architecture of today's service and transport networks for large terrestrial commercial carriers. It is important that those who study optical networks have a comprehensive understanding of what generates the demand for optical networks and the type of network architectures they comprise. A prerequisite of such an understanding is knowledge of the classes of commercial services and how the networks to provide them are constructed, maintained, and engineered.

Figure 13.1 shows a pictorial view of a useful network model for understanding today's commercial telecommunications network. This model breaks the "network" visually according to *horizontal* and *vertical* characterizations. The horizontal axis represents areas of the network divided into territorial and structural segments. The US network can be roughly categorized into three segments: *access*, *metro* [also sometimes called Metropolitan Area Network (MAN)], and *core* (also called *long distance*). Each of these segments consists of a complex interlacing of network layers (the ***vertical*** axis). We also note that European, Asian, North American, and other continental networks will look similar, but will have critical differences that vary with the deployment of different technologies, geographical characteristics, and politico-economic telecommunications organizational structures.

To define this more precisely, a *network layer* (or *overlay network*) consists of *nodes*, *edges* (or links), and *connections*. The nodes represent a particular set of switches or cross-connect equipment that exchange data (in either digital or analog form) among one another via the edges that connect them. Edges can be modeled as *directed* (unidirectional) or *undirected* (bidirectional) communication paths.

Optical Fiber Telecommunications V B: Systems and Networks
Copyright © 2008, Elsevier Inc. All rights reserved.
ISBN: 978-0-12-374172-1

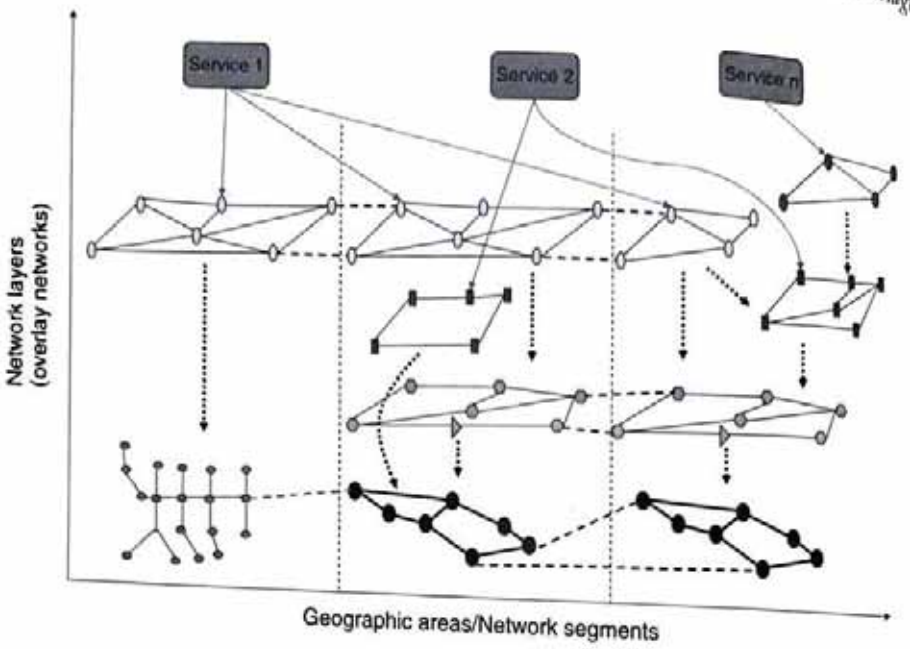


Figure 13.1 Graphical network model (this figure may be seen in color on the included CD-ROM).

The combination of nodes and edges transports *connections* from/to sources and destinations. Connections can be *point-to-point* (unidirectional or bidirectional), *point-to-multipoint*, or (more rarely) *multi-point-to-multipoint*. Connections serve two purposes. First, telecommunications services (depicted as large rounded rectangles in Figure 13.1) are transported by connections at various network layers in particular segments. The traffic for a given layer is carried by the connections at that layer. Second, edges of a given network layer are transported by the connections of one or more lower-layer networks. In this way, each layer is providing a "service" for the layer immediately above it—to provide connectivity. Additionally, for each layer this connectivity "service" may or may not be guaranteed against failures (fiber cuts, equipment outages, etc.). Some layers provide *restoration* to maintain connectivity, while others do not. This will be explained more thoroughly in Section 13.2.3. Note that in this chapter we use the term "restoration" to include other commonly used terms, such as *protection*, *resiliency*, and *robustness*. An important observation (especially for the student of packet and optical networks) regarding the nodes and edges in these network models is that they should all be considered as *logical*, even for the bottommost layers.

For example, the bottom layer of principal interest to this chapter usually represents some form of fiber transport media. However, an edge or link between two fiber nodes most likely represents multiple physical entities and/or layers. It in fact, can consist of multiple cable segments and splicing technologies, which are inside of ducts or subducts, which are often inside substructures (e.g., concrete conduit or plastic pipe). Thus, such an edge (or link) indeed represents a simplification (usually for routing or planning and engineering purposes) of multiple detailed layers and may consist of hundreds of individual physical components.

Another example is given by the Open Systems Interconnection (OSI) model developed by the ISO standards organization [1] and the colloquial classification of packet layering (e.g., "Layer 1, 2, 3," etc.) which has subsequently emerged in the industry. This relegates everything below Layer 2 [e.g., frame relay, multi-protocol label switching (MPLS), etc.] to the label of *Layer 1* or *physical layer* (PHY). This could include SONET, SDH, or other protocols/signals, which in fact can be one or more layers above the fiber layer and, as such, quite logical in nature. Finally, network technologies like Ethernet encompass multiple layers. Even though it started as a Layer 2 protocol, Ethernet has been standardized with multiple specific Layer 1 (or PHY) transmission technologies (e.g., 10BaseT, 100BaseFX, GigE, etc.). And the various Ethernet PHY definitions are used solely to transport Ethernet frames (Layer 2), and nothing else. The standardization of this tight integration of Layers 1 and 2 is a large factor in the very low cost of Ethernet equipment.

13.1.1 Network Segments

We provide a more specific (but still pictorially suggestive) picture of the three horizontal lower-layer network segments (access, metro, and core) in Figure 13.2. Note that for simplification, these three segments are laid out in partitioned, planar graphs. However, note that the reality is not as clean as depicted, with the segments often geographically intersecting one another.

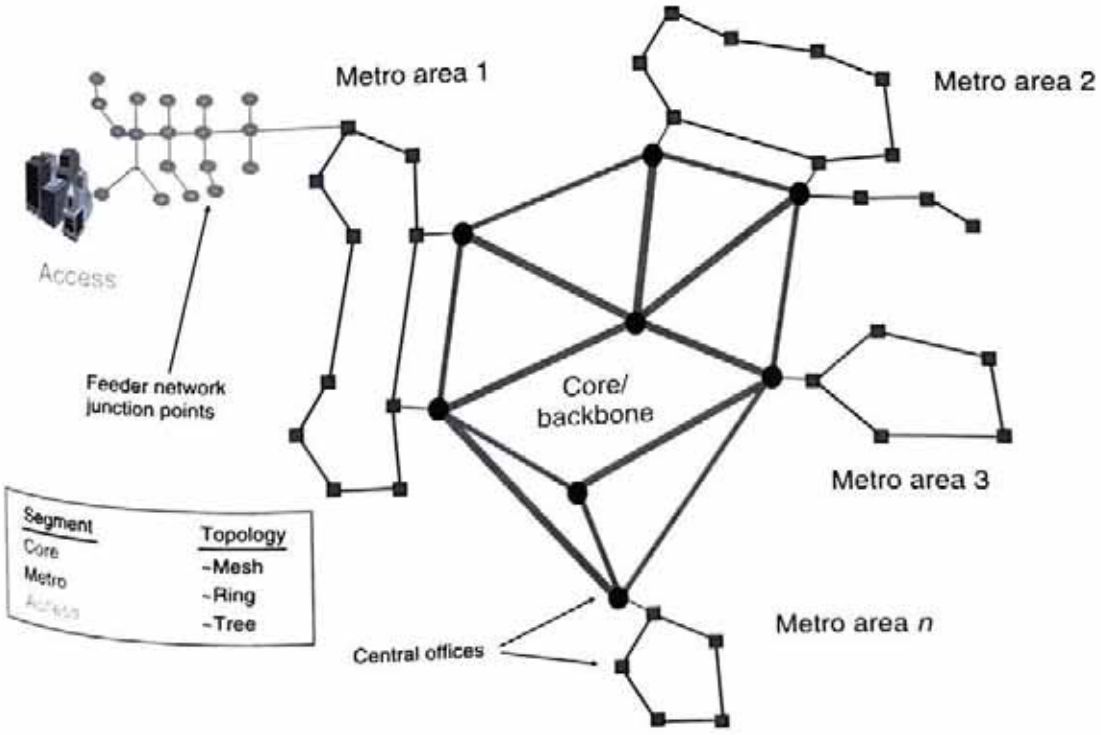


Figure 13.2 Example of geographical (horizontal) network segments (this figure may be seen in color on the included CD-ROM).

The lowest layer of the *access* network consists of copper, coaxial, or fiber media and generally has a tree-like graphical structure, especially in residential neighborhoods. The access segment is the "last-mile" to the customer, who is usually categorized as either *residential* (sometimes called *consumer*) or *business* (sometimes called *enterprise*). Note that wireless is also a predominant access-segment lower-layer technology, but we will not explore wireless technology architectures here.

A *metro* network uses time division multiplexing (TDM), packet, and optical multiplexing technologies [like dense wavelength division multiplexing (DWDM)] at the lowest layers. The metro network segment is generally defined as the nodes [central offices (COs)] and edges (connecting COs) within a metropolitan area. The *core* network generally consists of the nodes (each sometimes called a point of presence or *POP*) that are connected by intercity edges. Although various forms of metro packet transport and a *convergence* of TDM and packet technologies have emerged, the SONET/SDH ring continues to be the dominant transport-layer technology for metro networks since the early 1990s (the specific layering will be described in later sections). This is why the metro network segment is depicted as rings in Figure 13.2, although we note most metro networks can indeed support mesh-like topologies. A general description is that the metro network collects traffic from end customers and determines which traffic to route intrametro or intermetro; the intermetro traffic is then handed to the core network segment at the appropriate network layer.

The *core* or intercity network uses similar technologies as the metro, but has different traffic clustering and distance criteria and constraints. The core network tends to have a "mesh" structure at the lowest layer. This is generally justified because it is more economical to connect the many different cities by physically diverse routes, given the large amount of traffic that is aggregated to be carried on the core network. However, note that core networks differ significantly by country or continent. For example, because of smaller distance limitations, European core networks often have different technologies and network graphs than in the United States.

13.1.2 Access-Layer Networks and Technologies

Some of the residential access network layers for a large commercial local *Telco* carrier for the United States are illustrated in Figure 13.3 (a telecommunications carrier with a historical lineage to telephone companies will be referred to as a *Telco*). Note that for access networks there are many different architectural options, both existing and planned. A few of the principal Telco architectures are shown in Figure 13.3. The bottom layer is actually depicted as a composite of three separate layers, *copper loop*, *fiber loop*, and *fiber feeder*. Depending on region, these three layers can be disjoint or can be geographically coincident, as in Figure 13.3. As shown, the loop pairs can either route to a remote terminal (RT) or all the way to the serving CO. Feeder fiber connects the RTs to the CO.

Four architectural examples are illustrated in Figure 13.3, with a given combination of service offerings at the top layer. The reason for so many different

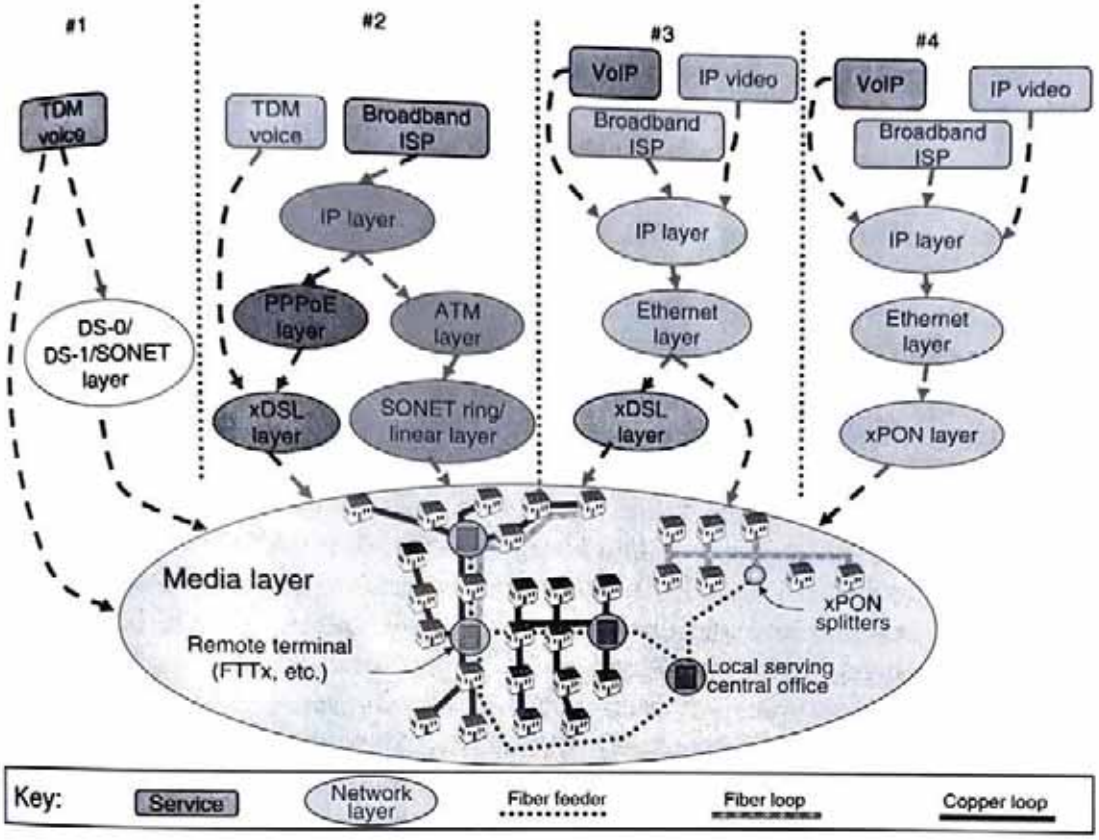


Figure 13.3 Some Telco access segment network layers (residential) (this figure may be seen in color on the included CD-ROM).

architectures is mostly due to the historical evolution of technology and services. This brings up an important aspect of real carrier networks: It is generally easier and more acceptable to introduce new architectures and technologies to a commercial network as a new overlay network or geographical segment than it is to remove an older architecture and technology base. This phenomenon is mostly dictated by economics and the reluctance/difficulty for customers to transition their services. As a result, there tends to be a "stacking up" of technologies and architectures, plus a sequence of product vintages and releases, over many years. Note that, as unattractive as it appears, there are some regions where all four of these architectures are geographically co-existent.

Architecture 1 is traditional TDM voice, the grandfather service for all Telcos. The analog voice signal is carried over copper pairs until it reaches an RT where it is digitized and then carried in the TDM channels of SONET systems (older DS-3/DS-1/copper feeder technology also still exists). The SONET systems can be restorable (e.g., UPSR ring) or unrestorable (linear chain).

Architecture 2 shows the evolution to digital subscriber line (DSL) technology, where TDM voice plus broadband ISP service are offered. There are several variations, but the traditional DSL broadband ISP service is shown, architected with point-to-point protocol over Ethernet (PPPoE) technology between the customer modem and the DSL access multiplexer (DSLAM) in the RT or CO.

Analog voice is multiplexed with the packet stream by the DSL technology over the copper wire pairs. Asynchronous transfer mode (ATM) switches are installed in the RT and route individual customer packets over ATM virtual circuits to the ISP edge switch (i.e., typically an *Internet Protocol (IP) router*, which is not shown). The ATM links are routed over SONET rings/chains at a lower layer, whose links in turn route over the feeder fiber layer. In our layered model, the ATM links are implemented as connections in the SONET layer [in units of synchronous transport signal (STS)-n, where an STS-1 signal is 51.84 Mb/s and an STS-n signal has a data rate of $n \times 51.84$ Mb/s]. There is consideration for using coarse wavelength division multiplexing (CWDM) or other WDM technology to carry the SONET OC-n [where the fundamental optical carrier (OC-1) carries an STS-1] links, but this is rare in access networks today.

Architecture 3 shows the more recent IP-over-xDSL technology, which has been installed mostly to carry video. Here, voice service is digitized [Voice over Internet Protocol (VoIP)] and all three services are co-transported over the IP layer. These services and applications per customer are carried via higher-layer protocols [such as Real-Time Protocol (RTP) or video-framing protocols] and separated/differentiated by various *virtual circuit* or *tunneling* methods [such as MPLS Layer 2 pseudo-wire [2] or Ethernet Virtual Local Area Networks (VLANs)]. The IP layer is carried over the Ethernet layer. Ethernet links are routed over xDSL (principally VDSL) to the RT (containing an *IP-DSLAM*) and then routed directly over fiber feeder to the router in the CO. WDM technology can also be used to multiplex the use of access fiber in the feeder network (not shown). Note that there exist hybrid variations where TDM voice service is carried over xDSL up to the RT and then from that point routed to the CO. This is analogous to how TDM voice is handled in architecture 2.

Finally, architecture 4 shows the use of xPON (e.g., BPON, GPON, or EPON) over the fiber loop layer that routes all the way to the customer premises. The upper layers are similar to architecture 3. Note that there are hybrids of architectures 4 and 3 where the xPON stops short of the customer premises (e.g., fiber-to-the-pedestal, -curb, -node, etc.) and xDSL is employed for the last few hundred feet.

The access network layers for business customers include basically the same layers as residential, but there are key differences. The media layers of business access networks have a higher penetration of fiber loop than residential networks, although over the period from 2005 to 2015 large carriers plan to install much more fiber to the residence. However, most business locations, especially the smaller ones, are still connected by copper pairs. For data services, carriers use DS-1 over copper or various xDSL technologies to route to the RT or CO. TDM voice is still a very prevalent service. Most business locations with fiber access (often called *fiber-on-net*) use a SONET ring layer. Ethernet access is growing in penetration and popularity, and there are a variety of Ethernet-over-SONET, Ethernet-over-fiber, and Ethernet-over-copper architectures being deployed.

For access to business customers, the last portion of the network (between the carrier equipment and the customer premise) varies in complexity much more than

for residential access. On the one hand, much of small-business access, especially away from the downtown areas, looks very similar to the residential access in Figure 13.3. In contrast, the configuration becomes more complex in central business districts. For the fiber-on-net locations, there is a mixture of technologies and architectures. One of the key factors in determining the restoration capabilities for the access network is the type or layout of the customers' buildings. In multitenant buildings (where the bulk of demand for bandwidth for Telco business services originates), this is influenced by whether a Telco has *common space* in that building. Common space is a rented area inside the building where the Telco can install its equipment. Today, that consists mostly of fiber or DSX cross-connect patch panels, DS-1/DS-3/STS-1 multiplexing equipment, and SONET ADMs to support voice trunks and private-line services. Ethernet switches are becoming more common, as well.

We illustrate an example multitenant building layout in Figure 13.4. The common space is often in the basement or in a maintenance closet. Fiber, coax, and copper cables are wired to the common space and travel up cable *risers* of the building, which usually are along plumbing or electrical conduits or in elevator shafts. Figure 13.4 shows some DS-1/DS-3 multiplexers and SONET ADMs (OC-12 and OC-48). The customers have equipment (called an M13) on their premises that multiplexes DS-1s into a DS-3 or which terminates an OC-3 or an OC-12, which is then cabled to the basement. Smaller customers simply use

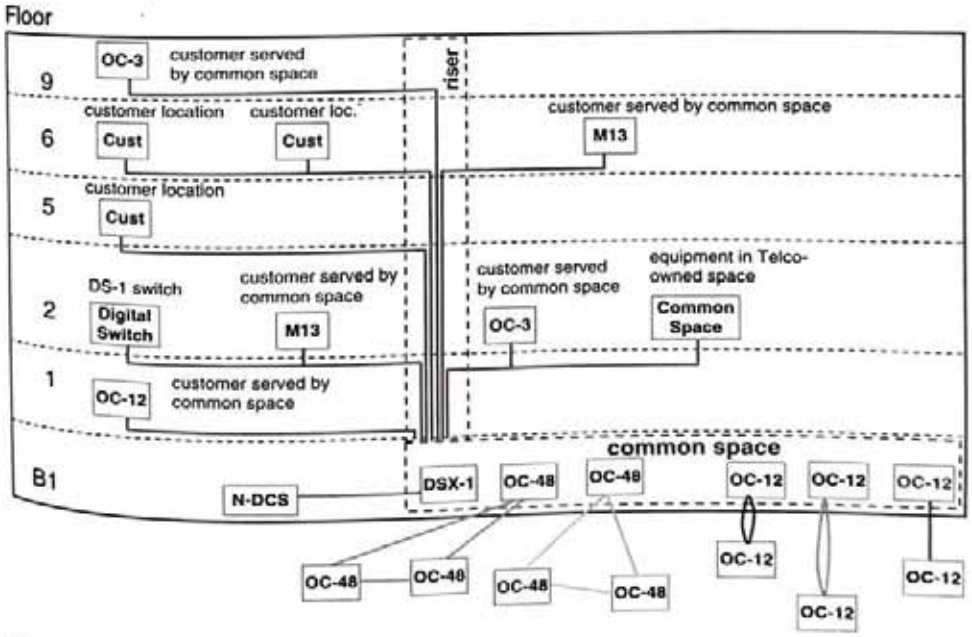


Figure 13.4 Example of building with common space (this figure may be seen in color on the included CD-ROM).

copper lines connected directly to the common space. Ethernet customers run their Fast Ethernet lines (copper or fiber) or GigE (fiber) directly to the common space. The common space often resembles a mini CO, where DS-0s and DS-1s are packed into STS-1s (called *grooming*) to ride on SONET rings to their next location, which might be either another customer location (for multinode rings) or a CO. At the CO (in the metro network segment—see Section 13.1.3 below), the lower-rate connections (DS-1s) are usually groomed to pack into STS-1 links. (Grooming is better illustrated by the example of Figure 13.7, covered later in the discussion of the metro segment.) Thus, even within the building there can be a complex network. Indeed, intrabuilding networks can be very convoluted because of such factors as constraints on physical access to common space (which usually involves landlord-tenant small-business contracts), the intersecting interests of various carriers, mergers and acquisitions of carriers, building riser restrictions, and, perhaps most of all, different vintages of equipment for both customer and carrier.

Except for voice, the principal business services are different from residential services, although small-business customers may use residential DSL or coaxial-based broadband ISP (cable modem) services. Virtual Private Ethernet (VPE), Virtual Private Network (VPN) (an IP-based service), Virtual Private LAN Service (VPLS), TDM Private Line, and Business ISP are the most typical examples of business data services. Business VoIP (often called *B-VoIP*, in contrast to residential/consumer VoIP, often called *C-VoIP*) is a growing service. This is because most businesses tend to need more bandwidth for data services and thus have more opportunities to use their data networks for their voice services as well. B-VoIP services tend to mostly use the *Session Initiation Protocol (SIP)* [3] to signal to the VoIP network to set up and control their calls and call features. The traditional, bulky Private Branch Exchange (PBX) is slowly being replaced by the more agile and flexible SIP- or IP-PBX. The *IP multimedia subsystem (IMS)* architecture, which is based on SIP, is a higher-layer server and protocol architecture that many carriers are pursuing to provide wired and wireless VoIP and data services [4].

Some of the other significant access-segment architectures not shown are coax and hybrids of coax and fiber feeder and non-IP-based video over fiber.

13.1.3 Metro-Layer Networks and Technologies

The network layers for a typical US metro network are depicted in Figure 13.5. Of the three major network segments, the metro network is particularly complicated because all the access architectures discussed previously (many of which were not shown) have to be transported over the metro segment. Since the access segment is the "last mile," it does not evolve as a unit and some parts have interfaces and technologies that are decades-old vintage. Thus, as newer transport technologies are introduced into the metro segment, it must still provide aggregation and transport for the older access-segment overlays and technologies. Once these services are mapped onto the metro segment, further aggregation can occur before

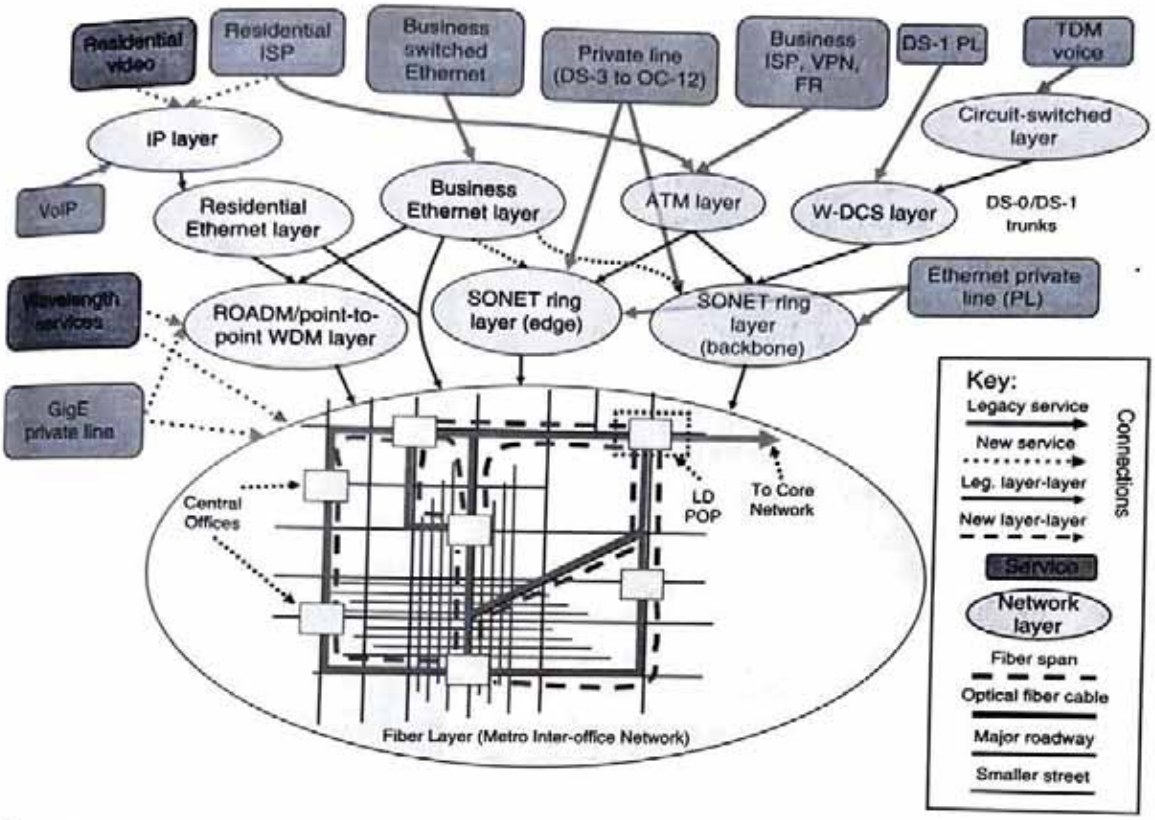


Figure 13.5 Example of Metro-Segment Network Layers (this figure may be seen in color on the included CD-ROM).

handing off to the core, which tends to simplify core-segment architectures. Furthermore, the large majority of service connections are intra-metro (i.e., both ends are in the same metropolitan area) and, in fact, some services are only offered or targeted as intra-metro service.

The bottom layer of Figure 13.5 is the fiber network. Virtually all Telco COs today are connected via fiber. There are usually one or more COs in each metro area that are each a point-of-presence (POP) where services are handed off to a long-distance entity (which can be another carrier, the same carrier, or a separate business unit of the same carrier). The idea of a POP arose in the years leading up to the breakup of the Bell System in 1984 and became entrenched via its Consent Decree and Federal Communications Commision (FCC) guidelines for its enforcement. That FCC document laid out the boundaries of each Telco metro network more rigorously than the original 161 LATAs (Local Access and Transport Area). This metro-to-core POP handoff can be complex, a mixture of virtual and physical in nature which varies by network layer and service. We illustrate this with the handoff for IP services later.

As we observe in Figure 13.5, most fiber cables run along streets and other public conduits, such as transportation lines (e.g., subways) or sewer/water lines, and bridges. The cables in cities are usually inside hard conduit and ducts or subducts and, in fact, may run in parallel in the same conduit with older copper cables that provide access-segment connectivity. Fiber cables in the city outskirts are often aerial

(on telephone poles or power lines), especially in the Northeastern United States. Furthermore, there is usually no clean geographical boundary between core and metro. For example, core-segment fiber cables often run in parallel with metro-segment cables over the same conduit structures. The factors determining how cables among the various segments are routed are very complex and depend on business factors such as negotiated rights-of-way, or Indefeasible Right of Use agreements (IRUs), capitalized long-term leases, and carrier corporate lineage (mergers, spin-offs, acquisitions, mutual agreements). However, although much industry "buzz" has ensued about the various long-distance overlay networks and services and legislation to govern US Telecommunications policy, the answer to the question of "who owns the cables and rights-of-way for our metro and access segments" is the key economic and business factor that has molded the structure of both our major and minor US carriers. This also ultimately has an impact on the structure of the equipment suppliers who supply those carriers.

As can be seen in Figure 13.5, since most fiber cables run along highways, the fiber network has a noticeable "grid" pattern, and as such, there are typically two and usually no more than four physical cable routes out of a CO. Often, multiple fiber routes exist that share part of their route on the same conduit section. Sometimes, fibers share the same cable and then split into different directions at cross streets or at the entrance to customer buildings. Where fiber cables enter a CO, they are usually wired from the cable vault entrance to some form of fiber patch-panel (often called an *LGX* or Lightguide Cross Connect) which is a physical device with manual or automated (remotely controlled) cross-connects. Some carriers have begun to deploy automated cross-connects based on mechanical fiber switches or MEMS devices at large COs. *Fiber spans* are defined between the patch panels. This is illustrated by the large dashed edges in the fiber layer of Figure 13.5. Mathematically speaking, these edges form another overlay network (nodes = patch panels and links = fiber spans) over the network layer of fiber cables (nodes = buildings and links = cable runs, depicted by large solid lines).

One can now understand why the introduction (Section 13.1) stressed the idea of "logical networks." The "fiber layer" in Figure 13.5 is, in fact, itself composed of multiple layers. If a network provisioner/planner wants to route a higher-layer network link or customer service over fiber, then he/she routes along the network graph corresponding to these fiber spans. The major benefit is that most of the physical splicing of the cable is done when the cable is installed, especially for major carriers. Normally, whenever a connection (higher layer link or service) has to be provisioned over fiber, the only physical installation (splicing and/or installation of fiber patch cords) occurs on the ends of the connection to connect the ports of the equipment interface cards into the patch panel [e.g., the interface card of a router, Ethernet switch, reconfigurable optical add/drop multiplexer (ROADM), Optical Transponder, etc.]. The connection through the intermediate nodes of the metro network can then be set up via remote cross-connect commands to the patch panels. An exception would be for needed intermediate equipment, such as O-E-O optical regenerators at intermediate nodes which, due to their high cost, are

not usually pre-installed and pre-connected into the patch panel. Even if the patch panel is manually cross-connected, no splicing is usually required end-to-end. The difficulty, as one can observe in Figure 13.5, is that the fiber spans (edges) can sometimes run together on portions of the route, called a *shared-risk-link group (SRLG)* originally defined in Ref. [5]. Thus, the planner must be knowledgeable about the actual fiber paths of these fiber spans to ensure that restoration objectives are met for whatever higher-layer network link (or service) is being provisioned. SRLGs are addressed again later in the discussion about network restoration.

Figure 13.5 illustrates only a simplified picture of the common metro layers. The dotted lines signify newer service or network layer connections vs. more traditional (legacy) transport architectures. Four network layers are shown that route over the fiber layer: edge SONET ring, backbone SONET ring, ROADM/point-to-point WDM, and IP. Furthermore, two services are shown that often route directly onto the fiber layer, Gigabit Ethernet Private Line and *wavelength services*. These services are also shown routing directly over the ROADM layer. This is because, in contrast to the core network, WDM is not yet pervasive in the metro segment. Thus, even in metros that have ROADM layers, many customers' connections may have to route directly on fiber for part of their route where the ROADM network has not been installed. Note that in reality, although not pictured at that level of detail, most carriers prefer some sort of network equipment between the customer premises equipment (CPE) and the fiber, a form of "demarcation point" so that they can monitor performance and isolate performance issues.

It is useful to clarify the term "Wavelength Services." For technical readers of an optical technology book, this may be confusing, because the term is, in fact, more marketing-oriented than technically accurate. Wavelength services today usually refer to private-line services corresponding to 2.5, 10, or 40 Gb/s SONET or SDH signals, their emerging ITU (ODU-1,2,3) containers, or 1-, 10-, or 100-Gigabit Ethernet signals. These signals are, in fact, electrical but are often transported over WDM equipment; therefore, they have colloquially been mislabeled as "Wavelength Services."

The metro WDM layer is a mixture of point-to-point and newer ROADM technologies; hence our label "ROADM/Pt-Pt WDM Layer." See Chapter 8 for a full description of the technologies in the ROADM layer. The ROADM overlays tend to be installed in ring-like (two-connected) topologies and, therefore, tend to have a more "network" appearance (for graphs-theorists, the ROADM network is comprised of just one connected subgraph). In contrast, point-to-point WDM was (and is) mostly installed to relieve fiber exhaust and therefore has a more scattered topology (consisting of many disconnected subgraphs). A determining factor for the architecture of the metro segment is minimizing network cost; consequently most carriers must cost-justify every individual installation of WDM technology. Where spare fiber is abundant, this is harder to justify economically. There is also a common industry misconception that installation of ROADM network layers can be justified in terms of reduced operations or provisioning costs. Given the discussion of patch panels above and the observation that the number of daily

connection requests for the ROADM layer (from higher layer links or services) is very small, it is clear the cost savings is insignificant in comparison with the capital cost of the WDM equipment.

We continue the explanation of the configuration of Figure 13.5 by providing some historical perspective. Let us examine the right-most stack of network layers in Figure 13.5. In the 1950s and 1960s, the Bell System (including Western Electric, its equipment supplier division at the time) developed digital encoding (called *pulse code modulation*) of analog voice and its resulting metro multiplexing and transmission technology, the DS-1 (1.54 Mb/s), subsequently, in the 1970s and 1980s the DS-1, with its ability to carry 24 DS-0 (64 Kb/s) voice-bearing trunks, became the mainstay transmission unit in metro network carriers. A natural evolution of the digital DS-0-based switch, the node of the circuit-switched layer in Figure 13.5, the manual DSX-0 and DSX-1 cross-connect frame, and finally the *narrowband digital cross-connect system (N-DCS)* and *wideband digital cross-connect system (W-DCS)* followed. An N-DCS cross-connects DS-0 channels from among its DS-1s interfaces and a W-DCS cross-connects DS-1 or SONET VT-1.5 channels from among its higher rate interfaces. Telcos later expanded their W-DCS layer beyond its original purpose (namely, to transport links of the circuit-switched layer) to establish DS-1 private-line services. This included the wholesaling of DS-1s to other carriers, the price of which was governed by tariffs.

Then packet switches began to offer the ability to encapsulate their data (payloads) inside clear-signal (nonchannelized) DS-1s, which could be readily provisioned by the W-DCS. With the advent of fiber optics and high-speed fiber optic terminals (multiplexers), the impact of single network component failures also grew. Bellcore decided to standardize the collection of fiber optic transmission rates with its SONET standard, as well as take the opportunity to standardize the collection of transmission systems with various restoration (protection) schemes that had emerged; consequently, SONET self-healing ring standards were established. Given the ability of SONET rings and chains to transport DS-3s links between W-DCSs or between DS-1/DS-3 multiplexers at smaller COs, the DS-3 private-line market emerged. The initial customers of the DS-3 private-line service were predominantly other carriers. The STS-*n*c market naturally emerged as the demand for higher rates grew and as the W-DCS was upgraded to have SONET optical interfaces. Thus, even though a revolution was occurring from voice to packet services, a major upheaval of metro network architectures (the rightmost network layer stack) was avoided by the combination of (1) a sustained, multiyear sequence of small advancements in interface rates, (2) gradual reduction of the per-unit costs, (3) marketing/wholesaling of associated private-line services, and (4) encapsulation of packets within TDM signals.

Looking at the emergence of packet networks in more detail, given the provisioning, planning, operations, and service marketing methodologies that evolved around the rightmost stack of Figure 13.5, packet networks had to find a place to fit in a mature stack of overlay networks. In particular, some of the earliest large packet network overlays arose from long-distance frame-relay service. The frame-relay

protocol was a simple layer-2 protocol to encapsulate IP packets and, in particular, through its concept of the permanent virtual circuit (PVC) enabled the setting up of connections in a more flexible and economic packet environment than private line. This provided (and still provides) customers with multiple (say, n) sites in different cities a more economical alternative to the "n-squared" problem; that is, they would require $n(n-1)/2$ point-to-point private-line connections to fully interconnect their sites, whereas with frame relay they only require n interfaces into the long-distance carrier's frame-relay network. The frame-relay customer can then set up a fully-connected mesh of virtual PVCs between his sites. The cost (and associated pricing) is such that customers with high values of n benefit the most. However, as ATM technology emerged, its concept of a *virtual channel* overlapped with that of frame-relay PVC. Thus, today almost all frame-relay services are carried over ATM networks. Frame-relay PVC/DLCIs are mapped to ATM VCs using the concept of a virtual channel identifier (VCI). The DS-1 private-line service continues to be the primary metro service to transport frame-relay signals from customer locations to the ATM network. In fact, as soon as the DS-1 encounters the first ATM switch, all the frame-relay encapsulation is discarded and replaced within the ATM adaptation layer (AAL). The frame-relay encapsulation is recreated at the far end so that it can interface to the customer port at that end. In fact, many customer interfaces on their switches completely bypass the frame-relay stage and originate as ATM cells.

The links of the ATM switches have to be transported at a lower layer, so ATM encapsulation within STS-*nc* signals was developed and thus could readily use the SONET infrastructure that evolved for private-line services. Furthermore, since ATM standards were not developed with a comprehensive restoration methodology (a major reason why some think SONET succeeded to a far great level than ATM), SONET rings also provided a ready restoration mechanism against lower-layer failures, which helped ATM networks to meet the higher level of QoS often expected for the services it transports. We note ATM vendors did provide various restoration methods, but they were never universally adopted by carriers. These relationships among layers for restoration are discussed below in the sections on restoration.

Given that metro ATM networks were deployed primarily for transport of local frame relay service and some early IP transport, they were the first standardized packet networks the Telcos deployed. Therefore, they naturally evolved to transport the emerging consumer DSL-based ISP services. All this discussion is captured in Figure 13.5, which shows residential (and smaller business) ISP services plus frame relay and business IP services transported over the metro ATM network layer.

Given this history, we observe that the "king" of metro transport for the past 15 years has been the SONET/SDH ring (which, for simplicity, we will confine ourselves to SONET for the remainder of this chapter). We have broken this into two layers to better illustrate the common architecture of the SONET ring layer. In most Telcos, the edge SONET ring layer mostly consists of *unidirectional path-switched rings (UPSRs)* or two-node, 1 + 1 or 1:1 protected systems (see Section 13.2.3 below) and the backbone SONET ring layer consisting of *bidirectional*

line-switched rings (BLSRs). The links of these rings are routed over the fiber although some links are routed onto point-to-point WDM systems on some spans that have reached fiber exhaust, as discussed above. The reason for the U BLSR demarcation is mostly due to network evolution and economics. Gener the UPSR is more economical for a traffic matrix whose nonzero demands have one endpoint, while the BLSR is better suited to a more distributed traffic matri reality, since the SONET ring has built-in restoration characteristics, its use es lished a QoS expectation for most private-line services or overlay networks wh links use the SONET layer. As a result, when Telcos provisioned private- services, they had to install a ring close to the customer location. For buildi that are on-net, this results in the installation of a (usually) smaller ADM in building, as illustrated in Figure 13.4. The UPSR predated the emergence of BLSR by many years. Most initial SONET metro networks were covered w multinode UPSRs. However, as demand grew and higher rate SONET interfac were offered, the architecture with edge and backbone SONET rings evolved. T typical SONET edge ring deployment migrated toward two-node rings: one at t customer premise and the other at the CO. Smaller customers received lower-ra rings. Furthermore, the end of the two-node ring at the CO is often a port on a card a larger ADM. The result has evolved into a massive collection of SONET-laye edge rings. Private-line connections (usually called *circuits* in Telco terminology with ends on two different edge rings route over the backbone rings or share a node the CO. Alternately, many private-line connections are access links to packet net- works or the metro segment of a long-distance private-line circuit, so only have one end on an edge ring of a given metro network. However, note that backbone SONET rings are mostly deployed to transport the links of the other higher layer networks such as ATM, DCS, or IP layers.

Technically speaking, the SONET layer is not as "pure" a routing layer as in packet networks or a WDM network with multidegree ROADMs. This is because most SONET rings consist of individual Add/Drop Multiplexers (ADMs) that are connected by OC-n (typically $n = 12$, 48, or 192) signals. Because network demand has become enormous compared to the (now) relatively small size of a ring, each metro network consists of hundreds (even thousands) of rings and consequently, there are enormous numbers of ADMs in large COs. Graphically, each CO consists of many (sometimes hundreds) of individual ring nodes. We illustrate this with a picture of the rings for a large metro network in Figure 13.6. A geographically-correct diagram is virtually impossible to show for large num- bers of rings, therefore Figure 13.6 represents a *connectivity graph*, where the nodes and edges of the rings are optimally placed to avoid edge crossings and hence be more visually useful. The heavily shaded areas are in fact a large collection of edges from enormous numbers of rings. As the reader can see, there is no "typical" topological pattern or architecture. This is even more pointed when we note that this figure depicts a medium-sized *competitive local exchange carrier (CLEC)* of a large city. The picture for an incumbent carrier (essentially a Telco) would be an order of magnitude more complex.

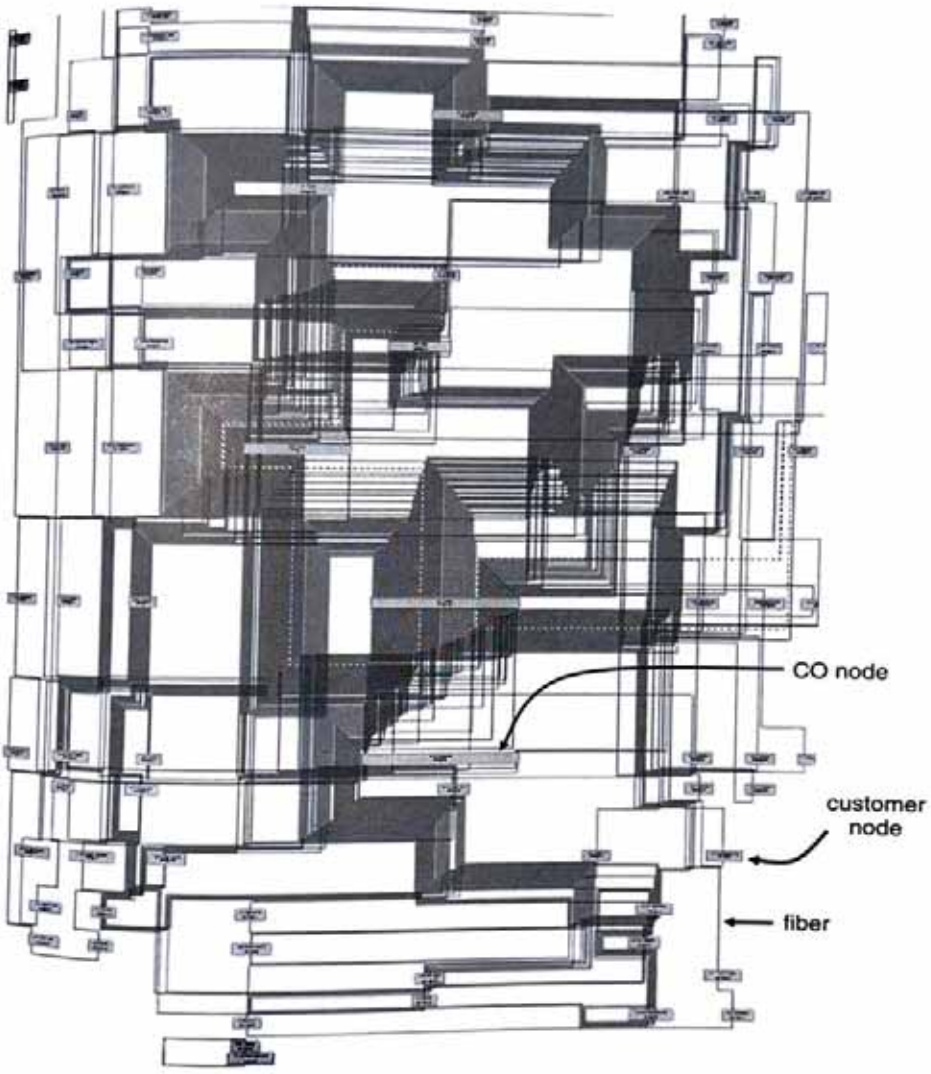


Figure 13.6 Connectivity graph of rings in large metro (this figure may be seen in color on the included CD-ROM).

Connections are generally routed across rings by a method called *ring hopping*. The usual approach is to provision tie links among the ADMs. For example, two different OC-192 ring nodes may have installed channelized OC-48 links between them on their *drop-side* or *tributary-side* ports. SONET STS-*nc* connections that have to route between the two rings can be assigned to the spare channels on the tie-links and cross-connected by the cross-connect fabrics of the two ADMs by remote command. The *broadband digital cross-connect system (B-DCS)* is a DCS

that cross-connects at DS-3/STS-1 or higher rates and was developed to mitiga the need for many pair-wise tie links needed for ring-hopping among ma ADMs in a given office. In fact, the value of a single B-DCS with multip interface cards with ports that act as single SONET ring nodes (called *subtend* rings) was studied and recommended many years ago [6]. Basically, this reduce multiple ADMs in a CO to just one equipment platform. However, the B-DC was not widely adopted in the metro network segment; consequently, although exists, we do not depict it in Figure 13.5. The reason for its limited deploymen was mostly because of overlap issues with regard to the W-DCS, which wa widely deployed prior to the advent of the B-DCS, and further coupled with th economics of multiplexing/demultiplexing TDM connections (grooming). How ever, in contrast, we note that in AT&T's core network segment, an early B-DC layer (with electrical DS-3 interfaces) was widely deployed in the early 1990s an adapted with a DCS restoration method called FASTAR to provide networ restoration [7]; however, this has been superseded by a more modern, restorable TDM cross-connect layer with optical interfaces and distributed intelligence although the original FASTAR network is still operational. This is described i the next section, as well as later sections on restoration.

In addition to links of higher-layer networks, Ethernet private-line services are routed over the SONET ring layers. Ethernet private line services provide inter faces on ADMs that receive GigE or Fast Ethernet signals and then encapsulate the Ethernet frames into standard SONET payloads in units of STS-1 capacity and then are transported similar to any SONET private line service. *Next-Gen* SONET features, such as Virtual Concatenation (VCAT) (described in other chapters), provide more flexible connection sizing than the historical SONET concatenated STS-1/3c/12c offerings. Even though the interface can be GigE, the interface card usually polices the rate down to the private line rate ($n \times$ STS-1). Full rate GigE private lines are also sometimes provided over SONET rings (e.g., inside STS-24c or VCAT STS-22vc signals), but generally Telcos are choosing to deploy them directly over the fiber layer or ROADM layer (or mixtures of the two) because of the large amount of capacity they use up on SONET rings.

Next the hand-off from the metro to core segments will be illustrated in more detail. Originally, the concept of a "POP" was a simpler concept that was narrowly defined around the Public-Switched Telephone Network (PSTN), mostly for transporting voice service [8]. The POP was a location where the Inter-LATA carrier put a circuit switch that had trunks to a metro (Intra-LATA) carrier *tandem* switch. The term Point of Interface (POI) was further defined, wherein the trans port network boundary was established between the two carriers. However, as private line services and packet networks evolved, this has become a far more complex concept that varies by service and the network layers. An example is shown in Figure 13.7. This diagrams the connection of a business customer for long-distance ISP service. Each gray vertical box represents a different piece of equipment (as labeled above it). The logical and physical links between these elements are shown by line segments along their corresponding layer as

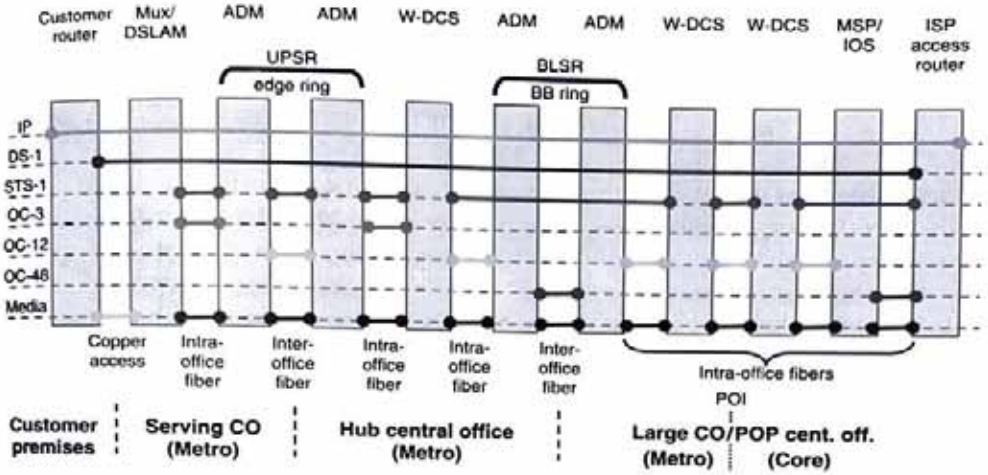


Figure 13.7 ISP POP example—TDM backhaul to access router (this figure may be seen in color on the included CD-ROM).

connection rate. The lowest horizontal line indicates the actual media used for that link in this example. The customer router has a DS-1 interface. From the point of view of the customer, he/she has a direct DS-1 to a port on the access router (AR) of the Inter-LATA carrier. From the point of view of the metro carrier, this is simply a DS-1 TDM private-line connection between the customer and a port/channel on the metro W-DCS: the metro carrier has no idea what is contained in the DS-1. This sort of encapsulation of packets in a TDM private-line connection to a packet-network interface is often called TDM *backhaul*.

The AR resides in the POP location, along with a lot of other equipment associated with the lower-layer networks that support this connection. As one can see, although the IP protocol stack views the DS-1 as a "physical" connection/layer, it is quite logical in nature. The dotted lines indicate the interface level of the connections between the network equipment, mostly determined by the multiplexing and demultiplexing that the DS-1 experiences as it traverses the metro network and mixes with other connections with different endpoints. Out of the customer premise, the DS-1 is carried over copper (the prevalent media to most customer locations) to a DSLAM or DS-1 multiplexer. The DS-1 multiplexer multiplexes DS-1s into STS-1s (or DS-3s for older equipment) and hands off via an OC-3 containing other STS-1s as well, to a SONET ADM. The signal is then routed on a leg of a UPSR OC-12 ring containing yet more traffic to a larger office with a W-DCS, which is the hub location associated with the serving CO of the customer location. The DS-1 is de-multiplexed into an STS-1 channel of an OC-3 link to the W-DCS. The links between W-DCSs (that form a mesh network) are channelized STS-1s, generally transported by OC-48/192 backbone rings. Figure 13.7 shows the DS-1 riding one link of the W-DCS network, which ends at the POP CO, which is usually co-located in the same building as a large metro CO.

For simplicity, this example assumes that this circuit routes on one leg (a single hop) of each ring.

Once the circuit enters the POP, it hands off from the metro W-DCS to the core W-DCS across the POI. This figure illustrates the interface as an OC-12, although the DS-3 interface is still a common handoff rate. One end can be owned by the metro carrier and the other by the core carrier. Note that as carriers have split up and subsequently merged since 1984, these two carriers might indeed be members of the same corporation. The FCC still has rules concerning separation of services, and some separation is likely to exist for some years following the writing of this chapter. After the DS-1 enters the core W-DCS, it is groomed with all the other DS-1s destined for the AR. The STS-1s destined for the AR are routed usually through a SONET cross-connect, which is a more intelligent and distributed version of a modern B-DCS with optical interfaces [which AT&T termed an *intelligent optical switch (IOS)*]. The purpose of routing through a platform like an IOS is that if there is a need to redirect the STS-1 toward another AR, this can be done easily by rerouting the STS-1 in the IOS network. This flexibility and grooming capability is the key motivation for routing through such cross-connect devices. Note that we depict the IOS along with an multi-service platform (MSP). An MSP multiplexes the lower rate signals to OC-48 or higher to hand off to the IOS. Its use is governed by the economics of pricing for interface cards on high capacity equipment.

The business access segment of the metro network is evolving to Ethernet for both interfaces and transport protocol. This is mainly since businesses prefer Ethernet interfaces into the metro *Wide Area Network (WAN)* because of the consistency with their LAN, simplicity, and low cost. Most Telcos have introduced Ethernet transport services. Initially, the Ethernet layer networks to carry these services have hub-and-spoke topologies that consist of a low number of switches or routers (sometimes just one) in the backbone portion and low cost Ethernet switches at the customer premise.

Figure 13.8 shows the service example of Figure 13.7 using the Business Ethernet layer. This is still basically the same ISP service, but here the customer has an Ethernet interface that is transported as a switched Ethernet service over the Ethernet layer. Currently, few of the links between Ethernet switches are carried over ROADM, but we illustrate a likely scenario for the coming years. It is unlikely that economics will support the use of ROADMs in the COs of the metro network that are currently occupied by the edge SONET ring layer, so we show the link from the *network premise equipment* (NPE), which is generally a low-cost Ethernet switch owned and controlled by the carrier but on or near the customer's premises, to the backbone Ethernet switch. The links between backbone Ethernet switches are shown routed over ROADMs since these are more likely to be in large COs. One major difference with Figure 13.7 is that, in contrast to SONET rings, the ROADM rings generally have no restoration capabilities; restoration has to be handled by the Ethernet layer. The handoff to the core router is typically GigE (or 10 GigE) over fiber. Note that this fiber might route through a fiber patch panel for flexibility. The customer's access link from his/her CPE to the AR is separated and identified by a

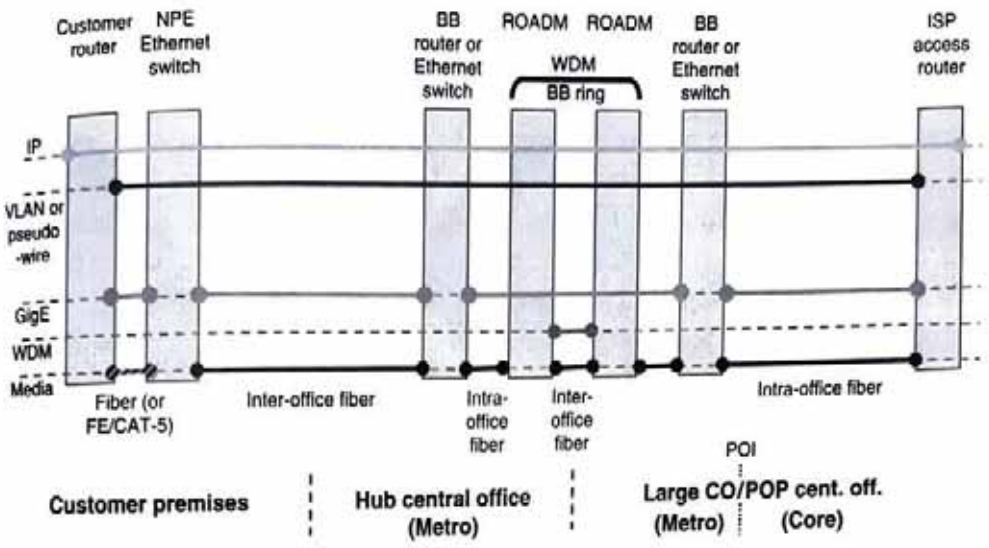


Figure 13.8 ISP POP Example—Ethernet transport to access router (this figure may be seen in color on the included CD-ROM).

VLAN ID or pseudo-wire ID, if layered with a PWE3-capable protocol, such as MPLS. Comparing Figures 13.7 and 13.8, one can readily see the improved simplicity of carrying packet services over packet networks.

Finally, we briefly mention the IP layer that has emerged to carry IPTV. In this discussion, we have segmented the Ethernet layer into two layers because this is how the Ethernet transport is being implemented in the near-term. Because of the order-of-magnitude size difference between the two Ethernet transport networks (residential/entertainment video vs. business Ethernet services), the residential Ethernet layer is being customized for the video application and most of the ROADM network layers have been economically justified to support the large numbers of GigE or 10 GigE links needed. Furthermore, the business Ethernet services have to support VPLS, spanning tree protocols and other virtual LAN services, which are not generally relevant for residential packet services. At some point, it is possible that the two Ethernet networks will merge, but that evolution is not clear.

13.1.4 Core-Layer Networks and Technologies

The typical network layers of the core network segment for a large US commercial carrier are depicted in Figure 13.9. Note that, as with the metro and access examples, this figure represents a significant simplification and does not describe all the network elements and technologies or all the services and will vary somewhat carrier by carrier; however, it does capture the most predominant layers and principal inter-layer relationships. As with the metro network segment, let us explore first the legacy stack of network layers on the right. Network layers that are essentially capped (not growing or growing very slowly) are hashed. As with the metro segment, this stack was engineered and developed over many decades to provide legacy voice services.

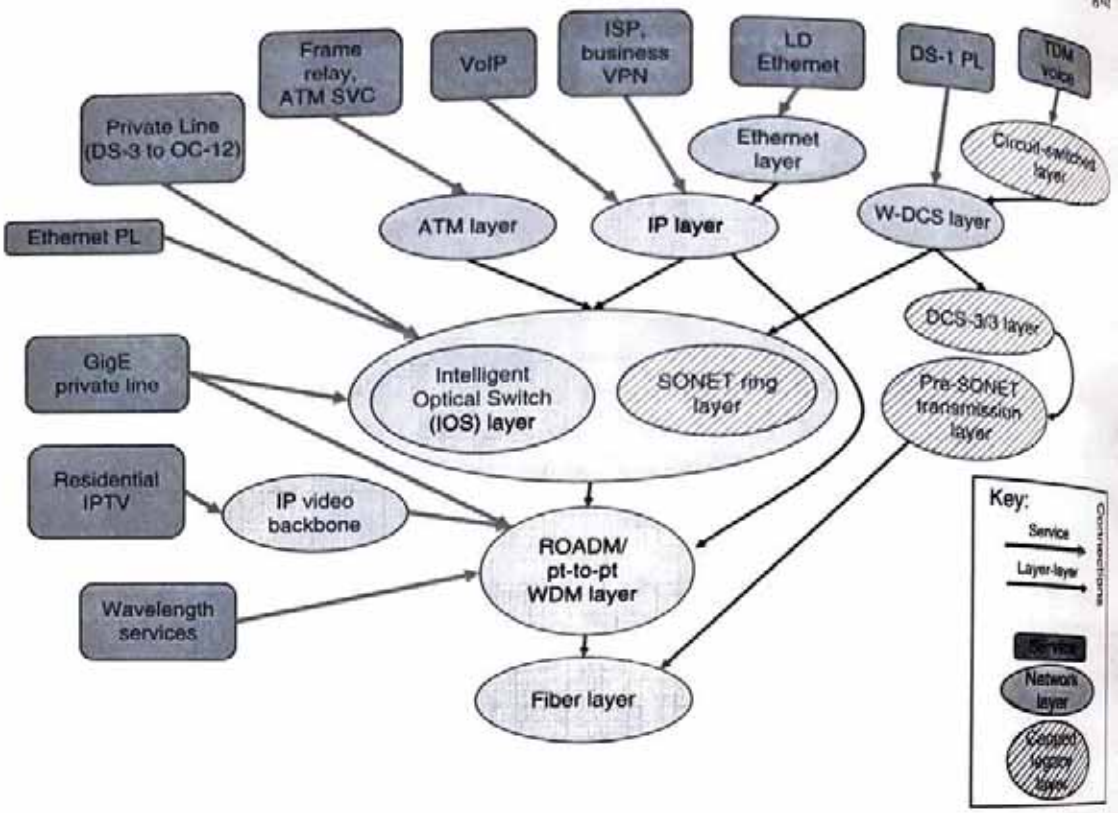


Figure 13.9 Example of core segment network layers (this figure may be seen in color on the included CD-ROM).

In the case of the core network segment, the original circuit-switched network was used for the wholesaling of telecommunications services (initially telephone calls). In fact, the wholesaling capability (among various other factors) led to the eventual breakup of the Bell System monopoly in the early 1980s as the AT&T Long Lines business unit was forced to wholesale call minutes to emerging, competitive long-distance companies, from which the US intra-LATA/inter-LATA competitive framework for long-distance carriers eventually emerged. Similar to the metro network segment, the DS-0 trunks (channels of DS-1s) that connect the nodes of the circuit-switched layer route over the W-DCS layer. Continuing down the "legacy" network stack, the DS-3 links of the W-DCS are transported over DCS-3/3s, a DCS that cross-connects DS-3s that interface it (note that the term B-DCS was coined by Bellcore to include both the pre-SONET DCS-3/3 and SONET DCS with its optical interfaces and ability to cross-connect at STS-1 or higher rates). The DS-3 links between the DCS-3/3s are routed over pre-SONET transmission systems. Originally, the DCS-3/3 was developed as a natural evolution of the manual DSX-3 cross-connect frame functionally analogous to the fiber "patch-panel" of today but with coax cables. That is, it was viewed as a remote-controlled DSX frame for cross-connecting the basic unit of transport at the time, the DS-3. The ability of the DCS-3/3 to rapidly (relatively to that era) and automatically reroute DS-3 connections motivated development of the first large-scale transport mesh restoration methods [7].

With the advent of SONET, growth of private line and IP services, the legacy transport stack became inadequate. During a transitional period, SONET rings were introduced into the core network. However, examining Figure 13.6, the impracticality of using SONET rings in a large core segment is obvious. Furthermore, early restoration studies showed SONET rings to be economically inferior to mesh restoration with DCS-type equipment, such as the Intelligent Optical Switch (IOS) [9]. Upon deployment of the IOS, the SONET ring layer was relegated to reaching smaller COs (i.e., edge of the core segment) and is essentially capped in most core network. The other main factor in this evolution is restoration, detailed in Section 13.2.3. The relationship of the other services, discussed for the most part in the metro segment, can be seen in Figure 13.9. Note that the separate IP video backbone layer can be functionally carried over the IP layer. However, full-featured entertainment video has very specialized performance and multicast properties and, hence, is usually carried over a dedicated layer. Perhaps as advanced QoS features of the core IP network are introduced, they will eventually enable the convergence of all IP services, including consumer entertainment video.

Comparing Figure 13.9 with Figure 13.5, the differences between the core network segment and the metro network segment are evident. While the SONET ring layers dominate in the metro, the SONET ring layer is minimal in the core. In the core, the fastest growing layer is the IP layer and almost all upper network layers route over a ubiquitous WDM layer. We thus focus on the IP/OL (read "IP over Optical Layer") architecture as the evolving "network of the future" and examine its migration in more detail in Figure 13.10. One can see that the IP layer is segmented into ARs and backbone routers (BRs). The reasons for this segmentation involve aggregation and restoration and, as such, are covered in the next section.

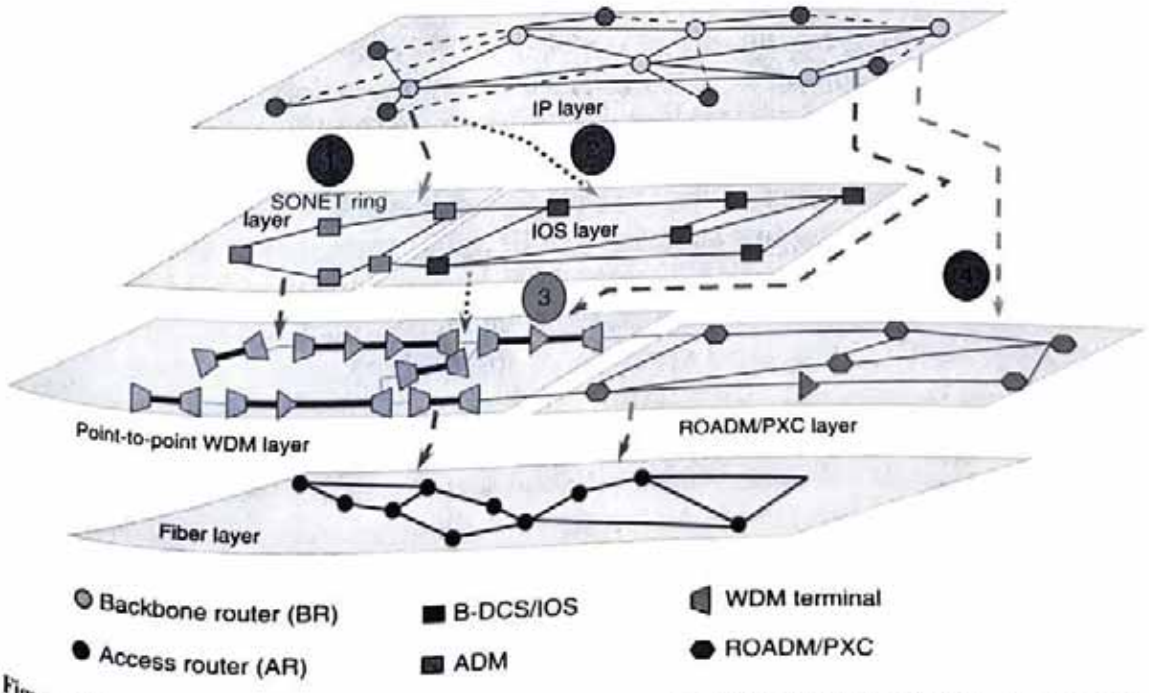


Figure 13.10 Example of core network layers (this figure may be seen in color on the included CD-ROM).

The core IP layer originally had its links (e.g., OC-12) route over the SONE ring layer and then the IOS layer (or combinations thereof). As the size of the links increased, they routed directly over the WDM layers, skipping all intermed ate SONET cross-connect layers. A chief reason for this is that the IP layer capable of restoration itself using the Internet Engineering Task For (IETF)-standardized suite of Interior Gateway Protocols (IGPs) [e.g., open shorte path first (OSPF) or intermediate-system-to-intermediate-system (IS-IS) reconve gence]. The economics and other tradeoffs of restoration at the IP layer vs. lowe layers are discussed in the next section.

The WDM layer is itself a study in network evolution. Originally, just as in the metro network, point-to-point WDM systems were deployed to relieve cables facing fiber exhaust and their shelf organization was specialized to interface links of SONET rings. However, as WDM became ubiquitous (which involved an evolution and then "de-evolution" of various fiber types and characteristics, as described in other chapters), the economics of core-segment transport became dominated by the short length of these point-to-point WDM systems (around 100 miles on average) and subsequent large use and high price of O-E-O regenerators. This led to the deployment of ultra-long-haul systems that connect ROADMs and photonic cross-connects (PXCs) which route photonic signals i.e., wavelengths. While the subject of ROADMs and PXCs is covered thoroughly in Chapter 8, we will give a brief overview for this discussion.

The simplest ROADMs allow for connections to be added to and dropped from wavelengths. A ROADM is connected to other nodes via links in two directions (often called East and West) and hence is termed a degree-2 node. More sophisticated ROADMs will need to be capable of multidegree routing, or to route wavelengths among more than two interfacing fiber pairs or directions. Multidegree capability is needed when the length of the WDM transmission systems (i.e., the distance between O-E-O regenerators) exceeds the distance between major network nodes. A multidegree node is connected by links to other nodes in more than two geographic directions. These are termed degree-3, degree-4, or higher. Thus, instead of adding/dropping traffic from the various directions in the electronic domain, a multidegree ROADM can add/drop in the photonic domain.

The evolutionary stages of the core network are marked with ovals, numbered 1–4 in Figure 13.10. The first stage had the IP traffic carried over SONET rings. In stage 2, that traffic was carried over the IOS network. In stage 3, IP signals were routed directly over the WDM layer, while in stage 4 that traffic is carried on a ROADM- and PXC-based network. As of this writing, this evolution is still in progress and currently all of the arrows of Figure 13.10 still exist, but the fourth stage is the target architecture and generates the most capacity growth. Today, the links of the IP layer between BRs are 10- and 40-Gb/s POS (Packet-over-SONET). However, 10-Gb/s Ethernet is emerging as well as 100-Gb/s Ethernet. We note that in contrast to enterprise networks, where Ethernet interfaces are significantly less expensive than their SONET counterparts, this relative difference narrows for very-high-speed router interfaces, whose costs are dominated by other factors related to packet switching and routing.

13.2 RELATIONSHIP OF SERVICES TO THE LAYERS

13.2.1 Service Requirements and How They Affect Technologies of the Layers

Network services are provided by network overlays at each layer. Each layer then places requirements on lower layers all the way down to the optical layer. Service requirements differ by the type of service and layer where they are provided. Service requirements can be divided into two basic classes: (1) expectations for provisioning new services or features of existing services, including bandwidth constraints/options, and (2) Quality-of-Service (QoS) parameters, such as Bit Error Rate (BER), packet loss, latency/delay, jitter. As one would expect, all of these requirements can differ by network segment, layer, and type of service. We give some high-level examples in Table 13.1. Note that these are for illustration purposes only and do not represent actual service requirements of any particular commercial carrier.

To achieve the stringent QoS requirements (such as network availability), a commercial carrier must provide various restoration methods in the network segments and layers. These will be described in a Section 13.2.3.

13.2.2 QoS, SLAs, and Network Availability

QoS is a complex function that varies layer by layer. For the TDM and optical layers, it is mostly a function of maximum BER tolerance and delay. Other chapters define and explain BER for optical layer networks in more detail. We will discuss QoS as it relates to the other layers of the network. However, note that all QoS examples we provide in this section are tutorial in nature. In particular, they do not constitute recommendations and are not meant to imply any commercial service guarantees. By modeling BER as a 0/1 function [i.e., either a network link exceeds a maximum and generates a *threshold crossing alert (TCA)* which generates a critical alarm or it does not] and then including whether a link is up or down [e.g., it generates *loss of signal (LOS)* or other critical alarm], one can formulate a *network availability* model for a particular network layer. This subject is complex, involving probability and networking, the details of which are outside the scope of this chapter, but the following will attempt a simple description.

Consider the WDM layer and index all of the n components of it plus its lower fiber layer for a particular specific network (e.g., optical amplifiers, WDM terminals, ROADMs, optical transponders, fiber spans) and then define a 0/1 variable, c_k, for each component k for $k = 1, \ldots, n$. $c_k = 0$ signifies that component k is *unavailable* or *down* and a $c_k = 1$ signifies that the component is *available* or *up*. A given *failure state*, s, of the network can be represented by the n-tuple $s = (c_1, c_2, \ldots, c_n)$. The failure state space is the set of all possible states. For a reasonable size network, n could be hundreds or thousands and thus

Table 13.1
Examples of service requirements for today's services.

Segment	Service	Example provisioning requirements	Example QoS
Access (Residential)	TDM Voice	New line in 7 days	Levels of static, buzz, outage. Repair in 3 days
	Broadband Residential ISP	New service in 3–14 days Various rates (down, up) = (512 Kb/s, 125 Kb/s), (1.5 Mb/s, 512 Kb/s), etc.	Refunds for outages ≥6 to 24 hours
	IP Video	Still under development	Must keep jitter and loss very low. Various forms of retransmit, FEC and buffering used to mitigate
Metro	Private Line	New service anywhere from 1 week to 6 months. Rates from DS-0 to OC-192/10 GigE	Jitter/delay meet Bellcore TR-253 Availability range from 0.9995 to 0.9999
	Business ISP	New service in weeks. If uses private-line backhaul for access, subject to above Bandwidth limited by access size or other policing SLA	For 95% of the time, no more than 1% packet loss. Latency <10–20 ms
	Voice	Residential: same as access; Wholesale business: varies by contract	≤0.005 blocking
Core	Private Line ≤622 Mb/s	Similar to metro; ends includes metro segment	Jitter/delay meet Bellcore TR-253 Latency varies by distance. Availability similar to metro except core availability is without the metro ends
	Private Line ≥622 Mb/s & Wavelength Services	Generally longer than private line if routed directly over WDM layer	SONET Private Line meets Bellcore TR-253. 1/10/100 GigE not well specified Availability not well specified since usually no restoration provided
	Business VPN	1 day. If uses private-line backhaul, access is subject to Private Line requirements	Packet availability QoS similar to metro Latency depends on distance and lower layer routing, sometimes ranging from 5 to 50 ms

the state space, although finite, is intractable in size (2^n). The rate at which components go down is governed by a stochastic process, often a Poisson, which is totally specified by a single parameter for each of them, the mean time between

failure (MTBF). The rate at which components come up is governed by a stochastic process, usually following an exponential distribution specified by the mean time to repair (MTTR). Similar to a massive Markov process, various failure states transition to other states by components going down and coming up (being repaired) with the given rates. Thus, the probability that the network is in a given failure state can be computed. However, the intractable size of the state space must be computationally addressed. The most significant observation is that the nonfailure state (fully operation state) of the network, $(1, 1, \ldots, 1)$, typically has a high probability (e.g., 0.8 or more). Failure states are usually rare. For example, if network availability reaches $1 - 10^{-5} = 0.99999$ (the popular "5 nines"), then it is generally considered very reliable. Thus, network states with probability less than a certain tolerance, ε, such as 10^{-5}, need not be explicitly evaluated. Let us index the states, s_i, in order of decreasing probability. Although the size of the state space is exponential, the maximum value k such that $\sum_{i=1}^{k} \Pr\{s_i\} \leq 1 - \varepsilon$ does NOT grow exponentially in the variable n. If we lump the states s_i for $i > k$ into an *unevaluated* set, then we only have to evaluate its complement (the *evaluated* set), which may be of tractable size. We know that we can bound the answer.

To make use of this, we define *performance functions* (PFs) for the QoS parameters of the services carried over the network under study, most of which involve some form of loss. Of course, not every service/customer/user of a network is affected by every failure state. For a network, this is generally defined in terms of the connections that use it. As an example, consider multichannel entertainment digital video (like IPTV) over a WDM backbone network from a central source (e.g., primary Head End) to destination nodes (e.g., secondary Head Ends). Define an availability PF, $a(d, s_i)$, to be 0 if failure state s_i causes destination d to be disconnected from the central source, and 1 otherwise. The expected availability of the video service to node d can be found by computing $\sum_{i=1}^{n} a(d, s_i) \cdot \Pr\{s_i\}$. To make this computation tractable, apply the technique above to only compute the QoS availability over the evaluated set. For example, suppose the availability PF over the evaluated set is found to be 0.999 (3 nines). Then we can bound the expected network availability, $E(a, d)$, is bounded by $0.999 \leq E(a, d) \leq 0.999 + \varepsilon$. The lower bound corresponds to disconnection [unavailable, $a(d, s_i) = 0$] over every state in the unevaluated set while the upper bound corresponds to service-without-loss [service available, $a(d, s_i) = 1$] over every state in the unevaluated set.

However, the catch is that *a priori* one does not know how to order the failure states. Generally, most analysts or network operators/engineers, even academicians, make simplifying assumptions about the evaluated set. For example, in most cases they assume it only consists of single-fiber failures and ignore all other failure states. But, often this only gets us to 0.99 or 0.999 probability of the state space. There are mathematical techniques and algorithms, such as statistical dominance rules, to build analytical performance tools to allow setting of a tight tolerance, ε, such as 10^{-5}. See Ref. [10].

Another issue with QoS parameters is how to define them over multiple connections. Drawing on our previous simple example of video service over WDM layer, we could define the *worst-case* network availability as $E_{WC}(a) = \min_d E(a, d)$. Or, we can define the average network availability $E_{AVG}(a) = \frac{1}{m}\sum_d E(a, d)$, where m is the number of connections (destinations). However, it should be noted that the average availability is a risky QoS measure to use as a design criterion because networks can be quite uneven in their design and traffic matrices. More sophisticated QoS measures can be defined on the distribution, such as finding the number x, such that the availability is more than x for 95% of the nodes.

Let us now turn our attention "up" the stack to the IP layer. There are many more potential and complex loss QoS PFs. For example, packet loss is the most common QoS for the IP layer. This is now a good place to introduce the notion of a *Service Level Agreement (SLA)*. An SLA is a set of guarantees that a carrier provides to a customer for the network service. Typical SLAs for the IP layer contain a packet loss PF based on the probability distribution. An example SLA could be that a customer should expect no more than 5% packet loss 95% of the time. Evaluating packet loss on the model we described above is more difficult than the much simpler network availability for the optical and circuit-switched layers. Because of the complex protocols that run over the IP layer, the somewhat unpredictable behavior of (IP) routers, the buffering and queuing disciplines in the routers, the routing aspect of traffic, and the bursty nature of packet traffic, this is truly challenging. That complexity will not be described in detail here, but refer the reader to Ref. [10] for more detail. Alternatively, for a private-line circuit (connection) over the DCS or SONET layers, an example SLA might be a simple availability QoS PF of network availability ≥ 0.99995. Again these are illustrative examples and do not represent actual SLAs of any particular carrier.

Let us now focus on a three-layer network: IP/WDM/fiber (read "IP over WDM over fiber") from the core segment of the overall network. Following the layered network model of Section 13.1, the IP links form connections transported by the WDM layer. One can then calculate the availability of those connections due to failures in the WDM layer just as done previously for the video example. By including the IP-layer components to the failure state space (e.g., router common cards, router fabric, router line cards, and optical transponders to connect the router links to the WDM terminals), one can then calculate the availability of the IP network. However, if the nonfailure state has a probability of about .8 or even .9 (as is often the case), the astute reader may ask, "How do we get the network closer to an objective in the range four to five nines (.9999–.99999)?" The greatest control that network administrators have to improve *network availability is restoration*. But recall that it has been found to be more cost-effective to limit restoration to certain layers. In this example, assuming that the core WDM layer offers no restoration (the usual situation in today's Telco core network), then restoration must be provided at the IP layer. Now, in addition to the downtime

from an IP-layer link (originating from a failure or planned take-down event, such as maintenance, in any one of the layers) and then subsequent return to the "up" state after repair, automatic restoration must be included, which does not return the IP link to the up state, but reroutes traffic around the down link. However, if indeed we were examining a multilayer network stack (such as the ATM/IOS layers in Figure 13.9) was being examined where restoration was provided at the lower layer, then the upper layer link would in fact experience a short down period while its connection was rerouted at the lower-layer network. These are factors that basically make the computation of the QoS PF, such as $a(d, s_i)$, more complex and dependent on the specific network layers, their interrelationships, and restoration methods at each layer. We examine specific restoration methods in the next section.

13.2.3 Network Restoration

The authors estimate that network operators and engineers spend at least 75% of their time attending to the maintenance of network performance. This involves avoiding, detecting, measuring, or planning for potential network component failures or traffic congestion, as well as maintenance and upgrade of network components to meet their network QoS and SLAs. After the QoS/SLA discussion of the previous section, the reader can now understand why network restoration is the main architectural tool that network operators have for achieving their QoS objectives.

In today's network segments, restoration against network failure can and does occur across several network layers. If one does not understand these layers and their relationship, then fully understanding restoration in large commercial networks is difficult. With the foregoing explanation of the network layers, there are some fundamental properties of network failure analysis that we will list here:

(1) Failures can originate at any layer. As a general rule, failures that originate in a given layer cannot be restored at a lower layer. For example, if an ATM switch fails, all the PVCs that route through that switch will be lost. The links which terminate on that switch cannot be restored at a lower layer since the switch itself has failed. The PVCs must be rerouted around the lost switch at the ATM layer.

(2) Since links at any layer are essentially logical, multiple links may route over the same connection or node at a lower layer. Thus, a network or node failure at a lower layer usually results in multiple link failures at higher layers. This phenomenon means operators must identify the shared-risk-link groups (SRLGs), mentioned earlier, which simply means it is important to identify the lower-layer links that each connection routes over. This is a critical fact often glossed over in some academic circles, where only single-link failures are examined in overlay networks.

(3) Given property 1, to meet even the weakest network availability or QoS objectives, all upper-layer networks must provide restoration in some form,

no matter how rudimentary. The crudest form of restoration is typica accomplished by reprovisioning connections through their connection se control mechanisms in response to network failure.

(4) It is important to understand how the behavior of a link, X, at a given lay depends on lower-layer restoration. If X is provided as a restorable connectio by the lower layer and a failure originates at the lower layer that causes X t go into an unavailable state (go down), then it is down for the (usually ver short) period until the connection is restored at the lower layer by rerouting i connection to a nonfailed path, assuming that sufficient restoration capacity i provided to reroute the connection. In contrast, if the lower layer provides n restoration or if link X is provisioned as a nonrestorable connection, then lin X stays down until the lower layer failure is repaired. If the failure originate at the same layer as X, then link X stays down until repaired.

Restoration Methods by Network Segment

The restoration methods for each network layer of the network segments discusse in previous sections are summarized below. As usual, there is no claim that thi listing is complete or fully represents the architectures or performance of an carrier. The restoration methods for the access-, metro-, and core-segment examp networks are summarized in Tables 13.2, 13.3, and 13.4 respectively.

We give brief and simplified descriptions of the most important restoratio methods.

UPSR SONET Ring (Figure 13.11a) In this architecture, the ring is confi ured over two unidirectional transport rings, one transmitting in a counterclock wise direction over the nodes of the ring (sometimes called the *outer ring*) and the other transmitting in the clockwise direction (sometimes called the *inner ring*). A

Table 13.2
Example of access-segment restoration methods.

Network layer	Restoration method(s) against network failures that originate at that layer or lower layers	Example Restoration time s
Copper/fiber	No automatic rerouting	hours-days
xPON, xDSL, PPPoE, ATM	No automatic rerouting	hours-days
Ethernet	No automatic rerouting	hours-days
IP	No automatic rerouting	hours-days
SONET chain (residential)	Hot-standbys for card failure. No automatic rerouting	10 ms (hot standby); hours-da (otherwise)
SONET ring (business)	UPSR or 1 + 1	10–20 ms

Table 13.3
Example of metro-segment restoration methods.

Network layer	Restoration method(s) against network failures that originate at that layer or lower layers	Example Restoration time scale
Fiber	No automatic rerouting	Hours
ROADM/Pt-Pt WDM layer	No automatic rerouting	Hours
ROADM layer with path-switched restoration	1 + 1 restoration	10–20ms
SONET ring (backbone)	BLSR	50 ms
SONET ring (edge)	UPSR or 1 + 1	10–20ms
ATM	(1) No automatic rerouting or	(1) Hours
	(2) P-NNI reprovisioning of PVCs or	(2) Seconds
	(3) centralized mesh restoration	(3) Seconds
W-DCS	No automatic rerouting	Hours
Ethernet or Layer 2 (e.g., MPLS)	(1) No automatic rerouting	(1) Hours
	(2) Spanning tree reconfiguration	(2) Seconds
	(3) Rapid spanning tree (RST)	(3) Subsecond
	(4) Link fast reroute (FRR)	(4) 50 ms
IP	IGP reconfiguration [reroutes new flows and generates new multicast trees (where used)]	10–60 s
Circuit-switched	Automatic alternate rerouting of new calls (existing calls dropped)	Seconds

STS-n connection enters and leaves the ring (denoted as nodes A and Z, respectively) via add/drop ports of the ADM at the two end nodes of the connection. For a ring in a nonfailed state, A transmits to Z and Z transmits to A in the same counterclockwise direction (on the outer ring). The connection transmits over n consecutive channels (or time slots), starting at a channel numbered $1 + kn$ (where k is an integer ≥ 0). It sends a signal from A to Z on one ring and a duplicate signal from A to Z in the other direction on the other ring. A port selector switch at the receiver chooses the normal service signal from one direction (usually counterclockwise). If this signal fails (i.e., the ADM receives an alarm), then the port selector switches to the alternate signal. It reverts to the original signal after receipt of a clear signal or under other control messages. Some versions of OC-3 or OC-12 rings are channelized to VT 1.5 channels.

BLSR SONET Ring (two-fiber) (Figure 13.11b) In this architecture, the first half of the channels (time slots) in each direction are used for transmission when the ring is in the nonfailed state. The second half of the channels is reserved for restoration. When a failure occurs, a *loop back* is done at the add/drop

Table 13.4

Example of core-segment restoration methods.

Network layer	Restoration method(s) against network failures that originate at that layer or lower layers	Example Restoration time scale
Fiber	No automatic rerouting	Hours
ROADM/Pt-Pt WDM layer	No automatic rerouting	Hours
ROADM layer with path-switched restoration	1 + 1 restoration (electrical)	10–20 ms
SONET ring	BLSR	50–100 ms
IOS (DCS) layer	Distributed path-based mesh restoration	Subsecond to seconds
ATM	P-NNI reprovisioning of PVCs	Seconds
DCS-3/3 layer	FASTAR (DCS mesh restoration under centralized control)	Minutes
W-DCS	No automatic rerouting	Hours
Ethernet or Layer 2 (e.g., MPLS)	TBD	
IP and IP video	(1) IGP reconfiguration [generates new multicast trees (where used)]	(1) 10–60s
	(2) Layer 2 fast reroute (FRR)	(2) 50 ms
Circuit-switched	Automatic alternate rerouting of new calls (existing calls dropped)	Seconds

port of the nodes surrounding the failed ring segment and the failed segment is patched with (rerouted over) the restoration channels over the links in the opposite direction of the failed segment around the ring. In the case of node failures and multiple failures, complicated procedures using *squelch tables* and other mechanisms are standardized to prevent mis-cross-connection for connections whose ends are in the failed segment. Connections must be assigned to the same channels on each link of the ring over which they route. A BLSR has the advantage over a USPR that the same channels (time slots) on different links can be assigned to different connections whose routes do not overlap over links of the ring and thus save some capacity.

BLSR SONET Ring (four-fiber) Four-fiber BLSR is not a commonly deployed technology and therefore we do not describe it here.

IP-Layer IGP Reconvergence This is the most common method of restoration in large commercial IP-layer networks today. It usually uses either OSPF or IS-IS signaling messages for general topology updates and then to recompute paths for routing IP flows or for MPLS forwarding. The IGP reconvergence mechanism is used whenever the IP topology changes by flooding messages (i.e., sending them to all neighboring nodes, and then recursively over the entire network) called *link state advertisements (LSAs)* in response to the change. Generally, the LSA is a *link-up*, *link-down*, or *weight change* message. If an IP-layer link changes status

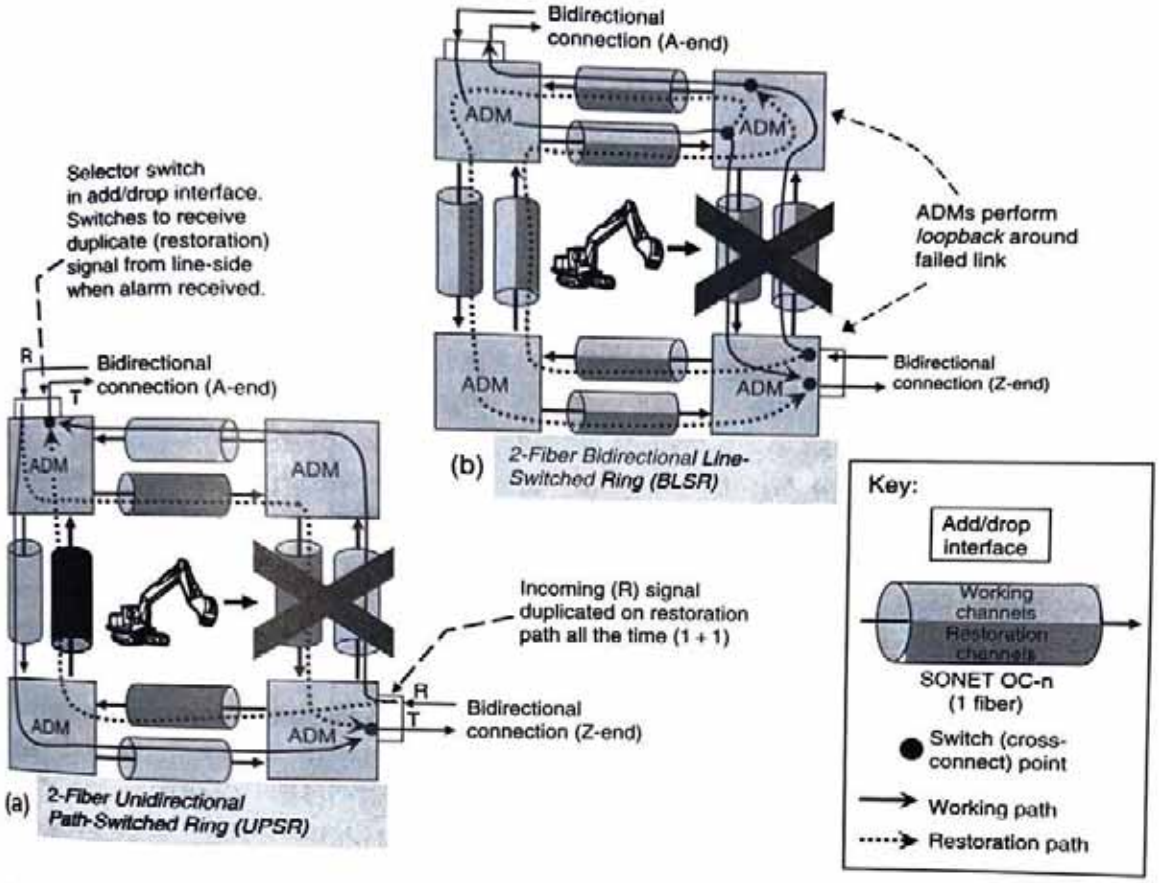


Figure 13.11 SONET self-healing rings. (a) UPSR, (b) BLSR (this figure may be seen in color on the included CD-ROM).

either by being detected to go down due to a network failure or by being intentionally taken out of service for a maintenance event (or, more commonly, its link weight is changed to a very large value), then that constitutes a topology change. Therefore, the IGP reconvergence mechanism becomes, effectively, a restoration method whereby each router recomputes the shortest paths to each destination router port in the network and then adjusts its routing table accordingly. In practice, this generally takes anywhere from 10 to 60 s; however, experiments and analysis have demonstrated that this can be reduced to less than 5 s, if various required timers are carefully adjusted. However, in practice, these are not easy to control.

IP-Layer MPLS Fast Reroute (FRR) This is a standardized restoration procedure in which primary and backup (restoration) *Label-Switched Paths (LSPs)* are established for next-hop or next-next-hop MPLS FRR. When a failure is detected at the routers surrounding the failure (usually via linecard interfaces), the MPLS forwarding label for the backup LSP is "pushed" on the MPLS shim-header at the first router and "popped" at the second router. These labels are precalculated and stored in the forwarding tables, so restoration is very fast. 100 ms or less

restoration has been demonstrated. While enabling rapid restoration, because paths are segmental "patches" to the primary paths, the alternate route is often and capacity inefficient. Furthermore, flows continue routing over the ba paths until the alarms clear, which may span hours or days. To address inefficiency, there are more sophisticated versions in which the FRR pro rapid restoration and then IGP reconvergence occurs and the IP-layer routi recalculated to use more efficient end-to-end paths.

IOS-Layer Shared Mesh Restoration This method is a distributed res tion technique. Links of the IOS network are assigned routing weights. Wh connection is provisioned, its source node computes and stores both the nor route (usually along a minimum-weight path) and a diverse restoration r through IOSs. The nodes communicate the state of the IOS network connecti via topology update messages transmitted over the SONET overhead on the li between the IOSs. When a failure occurs, the switches flood failure messages to nodes indicating the topology change. The source node for each affected conn tion then instigates the restoration process for its failed connections by send connection request messages along the links of the (precalculated) restoration pa Restoration is differentiated among priority and nonpriority connections. T failed priority connections use the diverse paths and get first claim at the spa channels. After a time-out that allows all priority connections to be reroute (restored) and the topology update messages to stabilize, the alternate paths the failed nonpriority connections are dynamically computed and rerouted alo those paths. It is called *shared* restoration because a given spare channel can used by different connections for nonsimultaneous failures. Shared mesh resto tion is generally more capacity-efficient than SONET rings in mesh networks (i.e networks with average connectivity greater than two).

1:1 or 1 + 1 Tail-End Switch (Electronic) The fastest forms of optical-laye restoration are one-by-one (1:1) and one-plus-one (1 + 1) restoration, which switch at the ends of the connections. With 1 + 1, the signal is sent in duplicate across bo the service path and the restoration path; the receiver then chooses the surviving signal upon detection of failure. In 1:1, the transmitted signal is switched to the restoration path upon detection of failure of the service path. Technically speaking this is actually not "optical" restoration, but rather restoration by the electro circuits at the ends of a lightpath, usually SONET-based and similar in behavi to a UPSR with only one channel. These techniques can trigger in as little as 20 (mostly due to hold-down timers to accommodate fault aging). However, the method requires a dedicated backup connection, which results in more than 100 restoration-overbuild of transport resources because of the longer diverse paths.

Circuit-Switched-Layer Dynamic Routing Restoration in this layer consi of routing new calls on alternate paths that avoid failed nodes or links. Existi

calls at the time of failure are lost. Dynamic routing finds the least utilized paths among many intermediate switches. Hierarchical routing can also provide restoration, but generally in highly connected networks dynamic routing provides significantly more paths than methods based on hierarchical routing and provides the same degree of restoration with more efficient use of capacity.

Ethernet-Layer Spanning Tree Protocol The spanning tree protocol generates paths in Ethernet networks via distributed signaling among switches. The principal development for spanning tree protocols was to avoid loops in Ethernet networks. However, it also generates a minimum-hop path and serves as a restoration method after topology updates settle. Rapid spanning tree is a version that accelerates the tree update process and provides a form of fast restoration.

Note that restoration architectures are not widely deployed in Telco networks in the residential loop and feeder networks (up to the first CO). This is mainly due to their tree-like topology (as illustrated in Figure 13.3), which would make alternate routing prohibitively expensive. However, the situation is more complex in business locations. As mentioned earlier, in the United States, the vast majority of business locations are still connected to the network by copper. For those locations it is somewhat similar to residential. However, for the fiber connected locations, as illustrated with the multitenant building example of Figure 13.4, one of the key factors in determining the restoration capabilities for the access network is the layout of the customer building. Generally, the portion of network up to the common space is not diversely routed since riser diversity is not easy to organize in this way. However, larger customers (e.g., other carriers) often will demand such diversity and be able to obtain it because of their large business presence in the building. Note that almost all SONET ADMs are installed as rings in common space. We note that today most access rings are in fact two-node SONET rings.

The most striking observation about Tables 13.2–13.4 is the variation of restoration capability of the layers directly above fiber. In the metro segment, the layer right above the fiber layer is (and has been) the SONET ring layer, which provides restoration. In contrast, in the core network, the next upper layer above fiber is WDM layer, which *does not* provide restoration. Although it is still early, so far, as ROADM technologies penetrate the metro segment, we are seeing this same trend. What are the reasons for this phenomenon? Are there future services or architectures that are driving this trend? The next sections shed more light on the reasons for this architectural trend.

Failures Across Multiple Network Layers

To describe some various restoration approaches and layer inter-relationships, restoration in the core network segment will be discussed first. Please re-examine the core-segment network layers shown in Figure 13.9. Automatic restoration is provided by the IP layer, SONET ring layer, IOS layer, and IP video backbone

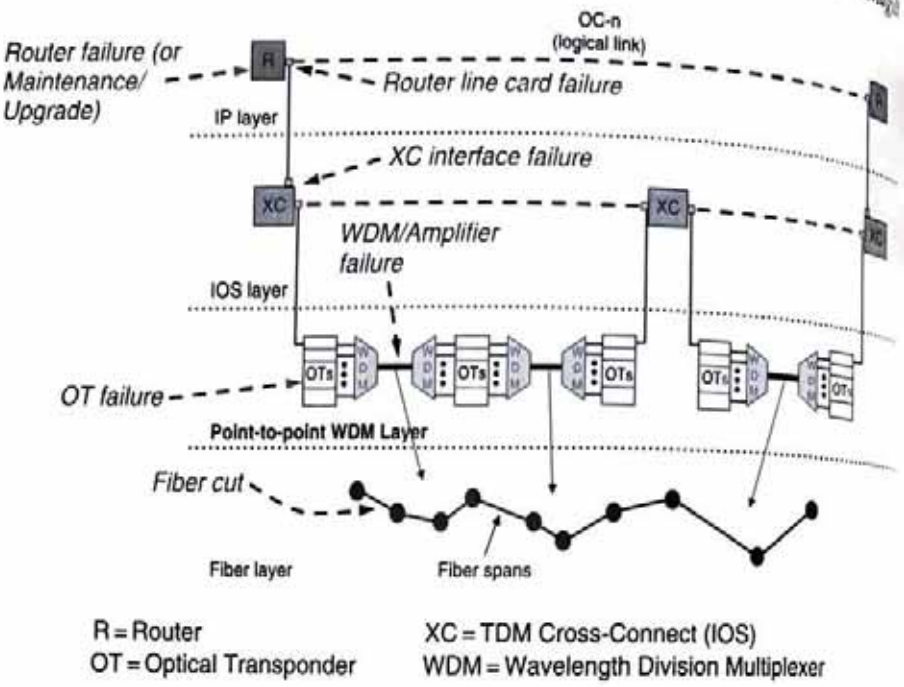


Figure 13.12 Example of Potential network failures at multiple layers (this figure may be seen in color on the included CD-ROM).

layer. As Table 13.4 suggests, at present, no automatic restoration is provided in the WDM layers (other than a very limited number of high-speed SONET private lines with 1:1 protection). Examining the typical architecture for the increasingly important IP layer, one can see that IGP reconvergence is used for restoration. Almost all lower-layer failures in the core IP-layer network of Figure 13.9 result in failure of multiple IP-layer links. In fact, some SRLG failures can cause over 10 links to go down in large ISPs. FRR over MPLS (or other tunneling protocols) also is used by some carriers and by the IP video backbone.

The concept of how failures occur at different layers and how the layers are affected is illustrated in Figure 13.12 for the core network segment. This figure depicts, by simple example, the four layers of Figure 13.9 from the IP layer down, with emphasis on potential network failure. It shows an IP-layer link that routes over the IOS layer, like a link between an AR and BR located in different cities. The links of the IOS layer in turn route over the Pt-Pt WDM layer. The links (DWDM systems) of the WDM layer route over the fiber layer. Figure 13.12 only shows a few of the many component failures that can originate in the layers. For example, an amplifier failure at the WDM layers usually results in the failure of multiple IOS links and, as a consequence, multiple IP links. In this example, the IOS network would reroute its failed SONET STS-*n* connections on different paths. Assuming the rerouting is successful, the IP STS-*n* link would go down for a short time and then after its alarm clears when the rerouting is complete, would be turned back up by the router controller. Alternatively, if some component of the router becomes unavailable (e.g., card failure, fabric failure, system or maintenance, upgrade, etc.), then the

IP layer reroutes the affected flows on different paths via IGP reconvergence. Note that the ability to restore connections at lower layers will not aid the affected IP flows that route through the failed router components. Also, the other overlay networks (frame relay, ATM, and circuit-switched voice) have not been depicted which could also be affected by the lower-layer failures.

It should be clear that to achieve QoS objectives, some form of restoration must be used. Many people in the optical community tend to assume that network operators will provide optical layer restoration as a straight-forward matter of achieving a QoS objective. But, at which layers should QoS objectives be set and how should they be defined? For example as shown, lower-layer failures can cause most of the network unavailability, but lower layers support multiple upper layers each with different QoS objectives. As the reader may have guessed already, the network layering, different service requirements, and choice of restoration architectures make this an inexact science. The situation is exacerbated when commercial carriers must further constrain these decisions by economics. For example, while it is generally true that (per unit of bandwidth) links at lower layers cost less than links at higher layers, it does not always follow that restoration should be fully provided at lower layer(s).

To explain this observation, in Figure 13.10 it can be seen that while some links between ARs and BRs route over the IOS or SONET ring layers, at present all links between core BRs skip the IOS layer and route directly onto the point-to-point WDM or ROADM layers. There is no restoration provided for the router links at these WDM layers. While efficient WDM restoration methods are not widely provided by vendors, in fact the reason for this architectural direction is mostly economic rather than vendor capability. Some of the leading factors for this interesting situation can be briefly summarized:

(1) Many failures (as illustrated by the router component failures in Figure 13.12) originate within the IP layer. Extra link capacity must be provided at that layer for such failures. This extra capacity can then also be used for failures that originate at lower layers. This obviates some of the advantages of lower-layer restoration.
(2) The IP layer can differentiate services by *Class of Service (CoS)* and assign different QoS to them. For example, one such QoS distinction is to restore premium services at a better level of restoration than best-effort services. In contrast, the WDM layer cannot make such fine-grain distinctions; it either restores or does not restore the entire connection supporting an IP-layer link, which contains a mixture of different CoS flows (including non-IP).
(3) Under normal conditions, there is spare capacity available in the IP layer to handle bursty demand. This is because, restoration requirements aside, to handle normal service demand, most IP links are engineered to run below 80% utilization during peak intervals and well below that during off-peak intervals.

Point 1 above is the most crucial driver. To further explain its effect, if we inspect Figure 13.10 in more detail, we see that each AR is *dual-homed* (either via intra-office fiber or via transport over the OL) to two BRs. There are a variety of architectural options to accomplish this. (1) There is just one BR per backbone CO. In this case, for each AR co-located with a BR, it uses intra-office fiber to link to the BR. The link to the mated BR must be transported over the other network transport layers (IOS, SONET ring, WDM, etc.). A remote AR (an AR not co-located with any BR) would use network transport to connect to each of two BRs. (2) Alternatively, there are dual-BRs in each backbone CO. Here, a co-located AR would use intra-office fiber to connect to each BR. A remote AR may either be linked to a mated pair of BRs in one office or link to BRs in different cities.

The reasons for having both ARs and BRs are manifold but, in particular, ARs aggregate lower rate interfaces from various customers. This function requires significant equipment footprint and processor resources (e.g., for customer-facing BGP and VPN processing). Major COs consist of many ARs to accommodate the low-rate customer interfaces. Without the aggregation function of the backbone router, each such office would be a myriad of inter-access-router tie links and inter-office links, which have historically proven to be unmanageable and expensive. To the contrary, backbone routers are primarily designed to be IP transport switches with only the highest speed interfaces. This segregation allows the BRs to be designed for multiterabit per second capacity. When a BR is down (e.g., due to failure of components, upgrade or maintenance), the AR shifts its demand to the alternate BR. This is an essential capability to achieve QoS. For example, key locations serve as peering points to other ISPs, the loss of which causes a significant outage. Thus, sufficient IP-link capacity is installed to reroute the traffic from an AR to its alternate-homed BR if any single-BR fails. However, because of point 2 above, the amount of extra capacity for restoration can be mitigated by allowing utilization levels during a failure event to rise above nonfailure levels. Also note that we have found that the biggest negative impact of the failure of a major BR is the loss of traffic originating at that router, rather than from "through" traffic. When all these three factors were examined the detailed network design optimization studies [11, 12] recommended providing restoration at the IP layer only.

13.3 NETWORK AND SERVICE EVOLUTION

13.3.1 Which Services will Grow and Which will Shrink? Demand and Capacity

Estimating the relative bandwidth impact of services on the network is not as easy as it would seem and, besides the complexity of layering, it also requires some knowledge of network design. First, we must divide service/network sizing into two pieces: *demand* (or *traffic*) and *capacity*.

For example, consider the frame relay service, which is transported over the core ATM overlay network. Define the traffic matrix $d(i,j,t)$ to be the average amount of demand (traffic) from source node i to destination j over time interval t, assuming for simplicity that all flows are point-to-point (in contrast to point-to-multipoint). A typical time interval for measurement purposes is 5 min. Note that packet traffic is typically bursty, so the maximum flow during that interval is even higher. Thus the demand over an interval is the time-averaged demand over that interval. Then the "instantaneous" total network demand at time t is $D(t) = \sum_{i,j} d(i,j,t)$. Of course, for most packet or voice services, $D(t)$ tends to be periodic by time-of-day, week, month, etc. However, in reality it never repeats exactly. We then define the total demand for that layer as $D = \max_t D(t)$ over the period of interest (say, several months). The capacity to carry this traffic is computed by a *network planning process*. In most commercial carriers, this process is akin to a network optimization problem where the objective is to minimize network cost subject to service requirement constraints, such as having enough capacity to provision new service requests and achieving a certain network QoS objective. Suppose for illustration that $D = 30$ Gb/s, the traffic matrix has an average hop count per connection of three hops, and that no link can exceed 80% utilization. Then, the network capacity must be at least 112.5 Gb/s. Of course, in reality this number would likely be higher because links can only be installed in discrete sizes, such as 2.5 or 10 Gb/s. In core networks, because network cost is more influenced by distance, network planners often prefer to express network capacity in units of capacity-distance. This is obtained by converting the capacity of each link in the overlay network to units of an "equivalent bandwidth" and then summing. "OC-48-miles," "DS-3-miles," and "10 Gb/s-km" are examples of such units. But, the major reason it is done this way is the network layering. That is, planners can easily cost out the equipment at the overlay network required to install a link, but it is hard to compute the cost of transporting that link as a connection over lower layers. Continuing the example, suppose these ATM links route over the IOS layer and that the IOS layer provides restoration. Thus, the "demand" for the IOS layer from the ATM layer is 112.5 Gb/s. However, private-line services and other overlays also use the IOS layer. Suppose the IOS layer is composed of 10-Gb/s links. Then the links from the ATM layer plus various private-line services route over the IOS network and share this link capacity. Furthermore, the (normally idle) restoration capacity is shared among these overlays and services. Once the IOS layer is sized, in an analogous, but quite different network optimization process, then one has a capacity estimate for this layer. This capacity then becomes demand for the WDM layer. Then repeat the same process to get demand for the fiber layer.

Now with an idea of how demand and capacity relate to layers, return to the question of how to express the relative size and growth impact of services. Generally, it is best to express it in multiple ways since it is often like comparing apples and oranges. For example, we can express an estimate of the size of the

demand at the highest layer where the service is provided. This is the 30-Gb/s fra relay example above. Then, we can express its impact to the lower layers o the network. In the case of the core network, this is either the WDM layer o fiber layers (or both). Thus, one metric we sometimes use in carrier networks is th amount of equivalent 10 Gb/s of capacity required by the WDM layer for eac service. However, it is clear from the above capacity sharing and restoration aspects, that this is not an easy metric to generate. One approach is to generate network designs down the layers all the way to the WDM layer for a given service with only that service generating all the demand. Interestingly, note that it would be erroneous to add those WDM capacity units over all services and then presume the result is the capacity of the combined network. This is because of the sharing of capacity, discrete units of capacity, and restoration at each layer. But, it does give a useful metric to determine relative size. Of course, such numbers are not easy to generate. Most network planners do not even know them because they usually tend to plan layer by layer.

For optical engineers and researchers, the most significant impact of services on the WDM or optical layers is in the metro segment. This is because the core WDM layer is largely built-out. More efficient long-haul WDM technology will be pursued, but there are not major capacity issues to be tackled. In contrast, many metro networks have virtually no WDM and it has been hard to identify economic drivers for its widespread deployment until recently. In metro networks where ROADM layers are deployed, they are deployed in backbone COs, roughly corresponding with the nodes of the metro backbone SONET rings. Therefore, in terms of impact to the telecommunications optical systems industry, the greatest potential impact will be in the metro segment.

To better understand this, in Table 13.5 provides some very rough capacity impacts on a large Telco metro network based on our study of these networks. Note that these are not official size estimates for any given carrier and are only provided to give a very rough perspective. The second column gives the relative

Table 13.5
Rough estimate of impact of services on WDM/fiber layers in a large metro network

Service	Present estimate of size in terms of WDM layer impact (%)	Relative growth rate
W-DCS (Voice + DS-1 Private Line)	5–10	Small
Business ISP, VPN, FR	0–5	Small
Private Line (DS-3 – OC-12)	5–10	Medium
Switched Ethernet	0–5	Medium
Residential ISP	0–5	Medium
Residential Video	50–75	Large
Ethernet PL (<GigE)	0–5	Small
Wavelength Services & GigE PL	20–30	Large
Dynamic Bandwidth Services	Negligible	Unknown

impact to the WDM and fiber layer caused by that service in today's network, as a fraction of the total capacity of that metro. This is complex and hard to estimate because with only a partial WDM footprint, many links of higher layer networks route over a combination of direct fiber and links of WDM systems. But, it can be clearly seen that in metros where IP networks are installed to carry residential video, this service dominates. High-rate private line and GigE private line are also large contributors. The demand for these services comes mostly from external overlay networks. Many of these customers are themselves carriers.

The last row lists *dynamic bandwidth services*. These dynamic *constant bit rate (CBR)* services are drivers for fast provisioning. They fall into two categories: subwavelength rate and wavelength rate (the dividing line is now 10 Gb/s and likely to move mostly to 40 Gb/s by 2012). We envisage that the main demand for the wavelength-rate dynamic CBR services will come from government, limited e-Science/grid computing, and large, multisite private networks. For example, AT&T offers a "bandwidth-on-demand" type service called Optical Mesh Service (OMS). This service currently provides rapid (subminute) setup of TDM connections between customer access locations via the IOS layer. Its principal customers are large enterprise customers or other carriers with large multisite networks.

An important point is that we assume traditional subwavelength TDM private-line services will continue to be used. This is an important point because many articles in the late 1990s seemed to dismiss these services. However, it is now many years later and today the capacity of the core network (core segment, IOS layer) needed to provide these sub-OC-48 services today is still large and growing. We assume sub-wavelength private line services will continue to have significant demand over the near future. Our rationale is as follows. Following historical growth patterns, the highest optical transport rates of today (i.e., wavelength services) evolve to subwavelength rates of tomorrow. Today most private lines are in fact links of customer IP networks. Many of these "customers" are indeed other carriers or government entities. The business need for such customer overlay networks is not expected to change, just the highest rates. Today the highest (wavelength) rate is 2.5 or 10 Gb/s. This will increase to 40 Gb/s or even 100 Gb/s in the next 5 years. Clearly, only a small fraction of customers will require private-line links at 40–100Gb/s. To provide these subwavelength private-line services, the most likely migration path is that they will be transported over one of the rapidly growing packet-based networks via pseudo-wire circuit emulation [13] with guaranteed minimum latency and QoS. Also, as the packet networks grow to huge size, PVC-type services will be more readily accepted and so these will grow to replace much of the TDM private line.

13.3.2 Network Evolution

Figure 13.13 depicts a potential evolution of the core segment. The first observation is that it is significantly simpler than today's network. Of course, this is

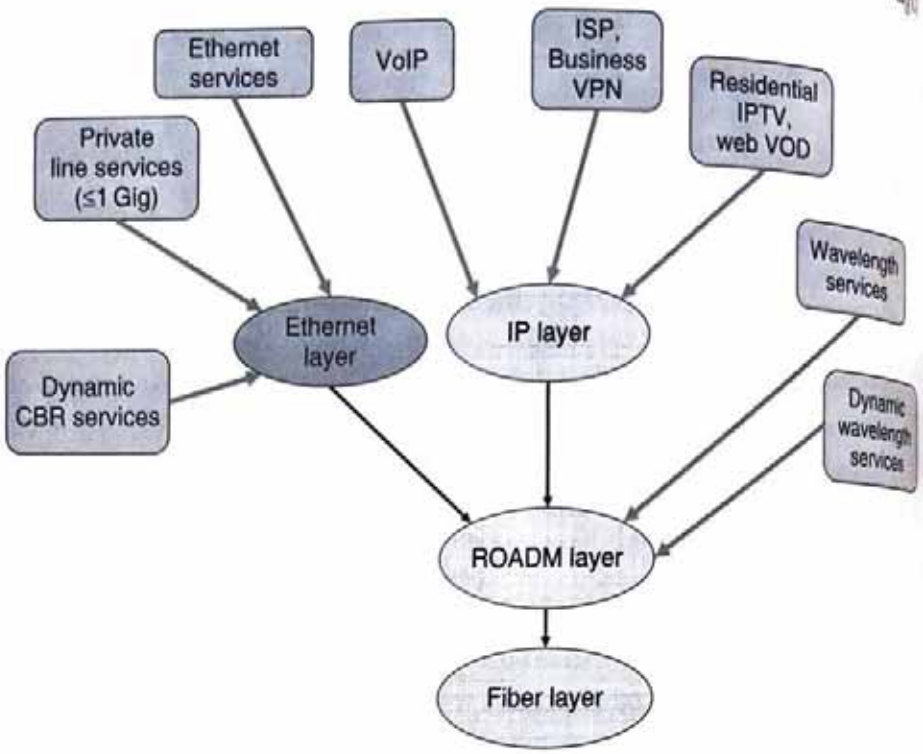


Figure 13.13 Potential future core segment network layers (this figure may be seen in color on the included CD-ROM).

idealistic in the sense that commercial networks have shown a difficulty to retire older technologies and architectures. This picture still retains TDM private-line services. These are likely to exist in some form for a long time because they accommodate the links of other overlay networks. We note that perhaps many customers of TDM private-line services migrate to more circuit-oriented virtual flows, such as Ethernet virtual private line or PVCs in some form (not necessarily frame relay or ATM). It is assumed that an evolved Ethernet layer can accommodate both types of flows: PVCs of some type and TDM circuits via circuit emulation. Both the IP layer and the Ethernet layer mostly transport IP packets and their higher layers, but the Ethernet layer is more focused on connecting LANs and transporting PVC-type connections.

To achieve this evolution, some enhancements over today's core packet networks will be needed. The challenges will be to demonstrate when pseudo-wire setup can be optimized for speed and that the latency of circuit emulation (due to its playout buffers) will meet grid-computing requirements. However, note that carrier-grade circuit emulation technologies exist, are generally available (GA) today, and have been shown to be able to meet the private-line service requirements [13, 14]. Uncertainty in the timing of this migration lies more in the metro segment (which evolves more slowly) and the economics of transitioning from legacy network architectures. Furthermore, ITU G.8261 has been established to enable timing (clock) distribution adequate for carrier-grade circuit emulation in Ethernet networks.

Dynamic wavelength-rate CBR services will require an optical control plane that supports fast provisioning and WDM transmission systems and cross-connects that can appropriately meet transient, power, and other optical requirements or, if not possible, cleverly preprovisioned O-E-O methods.

Voice is expected to migrate to VoIP. We note that this migration has been forecast more aggressively in the past, yet has defied predictions. However, despite the economic analysis or preference of carriers and predilection of many customers against change, equipment to feed the circuit-switched layer is being capped by many equipment suppliers and so its lifetime is very limited.

13.3.3 "Killer Apps"

What sorts of future applications will require massive link bandwidth and/or advanced network provisioning and management capabilities? There are many ways to interpret how future applications will affect the various networks discussed, but given the focus of this book, this discussion will be specific to the impact on the optical layers. Just as with the first sections of this chapter, we need to examine this segment by segment and layer by layer.

For the metro and access segments, referring back to Table 13.5 (based on present growth rates), the services that are projected to have the biggest impact on the metro fiber layer are residential IP and video services. The impact on the fiber layer of these two services can probably be lumped together. This is because IP-based transport of residential video services is estimated to become predominant ($\geq$50% of capacity) by 2015 and residential ISP services are carried over the same IP layer. Note that we can lump residential wireline voice services (VoIP) in as well, but its bandwidth usage is so insignificant that it essentially has no impact on the fiber layer. The impact on the DWDM and fiber layers of the metro segment for these services will be gated by the access segment, and therefore it is important to discuss the main technologies available for video transport over the access segment and how they might evolve.

Access Segment

Access-segment technologies fall roughly into five transport architectures: (1) IP layer over copper (xDSL), (2) IP layer over fiber, (3) satellite, (4) RF carrier over fiber, and (5) RF carrier over coaxial cable. A key difference is that architectures 1 and 2 are switched, while 3–5 are broadcast. That is, architectures 1 and 2 transport only the services/flows required to each customer at a given time, while 3–5 transport to each customer the union of services/flows over all customers in a given access subnetwork (or serving area) at a given time. This is a critical point because, under architectures 3, 4, and 5, an increase in the amount of available video entertainment channels plus growth of ISP service requires a commensurate increase in the access transport bandwidth. However, the real-time bandwidth actually required by each household has grown relatively slowly, most recently rising due to the advent of high-definition TV (HDTV) and increased demand for

ISP bandwidth (in the 3–20 Mb/s range). In contrast, the access bandwidth requir for architectures 1 and 2 is governed by the maximum cumulative bandwidth required by each household at any instance of time. Thus, for architectures 1 and 2 the breadth of entertainment channels is not limited by access-segment bandwidth and therefore can grow without bound.

Providing video and ISP services under architectures 1 and 2 (which are relatively new compared with architectures 3–5) has required substantial investment in the access and metro segments because the required lower layers of the infrastructure were not pre-existing. The accompanying investment in the core segment has been relatively very small; but we will return to discuss the core segment later. However, while the industry is experiencing an enormous economic "bump" from the access- and metro-segment deployments to realize architectures 1 and 2, based on the argument above and our assumption of predominant IP transport in 5 years, this growth will likely flatten out once the bulk of the network and its layers is deployed in most metropolitan areas. Then, network growth will be gated by growth in the maximum bandwidth per household. At present, based on an ISP service (1–20 Mb/s), four simultaneous video receivers, and anticipated standard definition (SD) (1–3 Mb/s peak) and HDTV (6–8 Mb/s peak) compressed data rates, the anticipated per-household peak demand over the next 5 years has been estimated anywhere from a low of 20 Mb/s to a high of 50 Mb/s. Given these estimates and assuming architecture 1 or 2, the variety of entertainment channels, video-on-demand (VoD), and other applications, which can be more easily accommodated by a switched IP infrastructure, can grow at a frenetic pace [e.g., person-to-person video, home-monitoring, YouTube-type video, intense gaming (high-speed, high-resolution, multiplayer), etc.], yet continue to be accommodated by the anticipated maximum access bandwidth. Therefore, for the residential access segment we would need to see new killer apps which require ~10 Mb/s or more to cause the next "bump" in commercialization or spending.

Alternatively, for business access, there is a completely different situation. Business customers who require large bandwidth usually aggregate many employees or serve many customers, either directly or indirectly through other businesses. While a maximum of 20–50 Mb/s may be adequate for most residential locations, many businesses today require multiple gigabits per second. The distinction, however, is that there are relatively few of these large enterprises when compared with the much larger number of residential customers (by more than three orders of magnitude). Also, almost all business locations with significant bandwidth requirements already have fiber to their locations. We do not anticipate any business killer apps driving access and metro bandwidth requirements much differently than the estimates in Table 13.5, although we will discuss a special class of academic/government customer when we discuss the impact on the core segment later.

Metro Segment

To examine the impact of residential IP (video + ISP service) on the metro segment in more detail, look at how entertainment video might evolve. Today, most entertainment

TV shows are still broadcast on a schedule, typically 300–500 simultaneous SD channel feeds plus 10–100 HD channels, assembled from a variety of entertainment content providers. Some satellite carriers are trying to get a jump on terrestrial carriers by offering hundreds of all-HD channels. To deliver this type of service to residential customers over the metro network, IP *multicast* [using *protocol independent multicast* (PIM)] can be used. Under this methodology, a *video hub office* (VHO) or *headend* for a metro area transmits the IP streams for each channel to downstream routers or switches. Each router or switch in the IP layer takes the channel stream received and duplicates it to each of its downstream links, according to a multicast routing tree that is set up via the PIM signaling protocol for each channel (note that to lower costs, many implementations of IPTV install Ethernet switches closer to the customer. These switches make use of a protocol called Internet Group Management Protocol (IGMP) for multicast, rather than PIM, which is mostly confined to routers). The last router before the customer (usually located in an RT or mini-headend) only transmits the channels that the customer requires in real-time. This tends to equalize the cost of infrastructure when compared to alternate architectures (like 3–5 above) which broadcast all channels to each customer. In contrast, if IP *unicast* is used instead (in which case virtual point-to-point connections are made between the VHO and each customer), then the metro transport cost is much higher. Today, both multicast and unicast are required. Unicast is mostly used for ISP service and VoD, where customers choose their entertainment from a stored catalog of programs.

Continuing with the evolution of residential video entertainment, in 2007 it is estimated that of the 200–500 TV channels, only 10–20 actually require real-time streaming (e.g., sports events, news, and live events). The principal question, then, is how prevalent or even predominant will the VoD model become? Taking this to the next step, how will VoD, as offered by video entertainment providers today, differ from that of video delivered via the Internet? These two forms of entertainment are converging. The main differences today are quality and licensing: TV offers consistently high-quality video and more heavily licensed movies and TV shows, whereas the Internet is a wild repository of anything from content similar to TV itself, to homemade, low-quality webcam movie shorts with no licensing or copyright impact. But, from the standpoint of network layers and technologies, there is little difference! These two services (ISP and IPTV) are only separated by virtual streams inside the IP transport pipe. If the latter model is realized in full (i.e., almost all forms of video entertainment are VoD), this will increase the bandwidth requirements substantially on the metro transport layers because, without many simultaneous channels, multicasting provides little advantage. Thus, the metro transport networks will continue to grow beyond their initial deployments, even though the maximum residential bandwidth remains constant. But, as with the access segment itself, the capacity (e.g., cumulative capacity in units of 10 Gb/s-km) of the metro-segment residential-IP layer will reach a maximum on the order of nd^2 [also denoted $O(nd^2)$], where n is the number of residential customers and d is a parameter that estimates the geographical extent of the access network served (e.g., "network diameter"). Alternatively, if this IP

layer of the metro segment is dominated by multicast transmission, then the capacity scales only as $O(d^2)$. We anticipate that the VoD model will predominate after 5 years. We do not base this prediction on the desires of the carriers, who need to receive return on their rather substantial investment, but rather by the enormous competitive pressure asserted by the Internet, coupled with the lifestyle shift currently underway. These changes include the use of "time-shifting" devices (such as Tivo®) and increasing numbers of teens and younger adults taking entertainment from the Internet. This shift also significantly changes what type and where advertising can be placed, with substantial impact on the overall economics. However, we note that in the past many traditional entertainment distribution models were turned upside down by Internet-based sharing models (such as sharing of music via free music distribution sites), only to be reigned in by the US tort system and copyright and licensing laws. Therefore, we predict this evolution will not be a smooth transition but will move ahead in spurts.

As for the ISP service itself and the previously mentioned estimate of 20–50 Mb/s per household maximum, what sorts of applications will even get close to this limit or, ultimately, push bandwidth beyond this? Since high-quality video channels will be available via VoD for the traditional SD/HD streaming (or VoD) video entertainment portion of the IP pipe, applications that are more distributed and not mass produced will be needed to drive this portion. A very likely candidate is monitoring or surveillance. With heightened concern for safety, two-income households, inexpensive and robotic web-cams, and just plain curiosity, there will be a large increase in local video monitoring. For example, parents might want to monitor their children at home, daycare, or school, or to monitor their pets or check on the security of their home or other locations. More people are also using video conferencing while working at home. Is this enough to stress the 20 Mb/s maximum access bandwidth? The answer is "no," unless the video quality becomes more like SD and there are multiple, simultaneous feeds required. However, another interesting driver for increased access bandwidth and a less stringent metro network has been identified [15]. Because of the current video streaming model described above, the IP layer of the metro network has to deliver incredibly reliable performance. Jitter and small failures or packet loss are not well tolerated. Furthermore, video features, such as "pause," "fast-forward," or "rewind," are somewhat limited under VoD. For example, to maintain the proper encoding and sequencing of differential video frames, a show/movie is usually prerecorded in a separate compressed date file for each "fast-forward" or "rewind" speed offered. When the customer selects a particular direction and speed, the specific prerecorded file is accessed and provided by a separate stream. Some of this can be provided by a digital video recorder (DVR), but it still requires some degree of streaming limitation. However, if we hypothesize an access link with huge bandwidth, then for non-real-time content there is no need for streaming. The customer (or the system) can simply download (or pre-fetch) all or most of the video file and then have a full set of features, similar to a DVD. But, most importantly to the transport network, it alleviates most of the high-availability and low-latency

characteristics of the metro IP layer and furthermore alleviates much of the need for IETF-motivated *diffserve* performance models for high-priority streaming data. Is this enough of a driver to force up the required bandwidth of the access link? We are not prepared to predict this, but it is an interesting viewpoint.

Core Segment

Although some requirements on DWDM and optical technology might be unique to the core segment, its size will be bounded by a function of the access limitations identified above. A key historical factor in the core segment is the $O(n)$ vs $O(n^2)$ effect described in Section 13.1.3. This continues to be a key factor which will be illustrated with an example. Today, residential ISP customers download a VoD service from an IPTV carrier or something similar from the ISP from a few, centralized sources. This, to date, has also applied to Internet gaming, wherein there are large numbers of end users, but who interact mostly via centralized servers. This implies that the impact on the core segment is $O(n)$ because of the small, fixed number of sources compared to the total number of end customers, n. However, there are point-to-point applications (or in Internet lingo *peer-to-peer*) such as residential monitoring and peer-to-peer video, multimedia, and music applications. If the video bandwidth requirement of such a point-to-point session increases to become closer to that of a VoD movie or TV show, then the peer-to-peer applications will dominate the centralized services. Most importantly to the core network segment, its traffic matrix (the matrix $\{d_{ij}\}$ whose entries are demand from AR i and AR j) will have density (number of entries with significant demand) more like $O(n^2)$, rather than $O(n)$. This will increase the required network capacity. However, note that the impact on the household maximum bandwidth in the access segment is unchanged. Rather, it is the networking effect that mostly impacts the core segment layers.

Our observation is supported by the current and historical distribution of types of Internet services. Peer-to-peer applications have constituted a large portion of long-distance ISP traffic and we now estimate they comprise over 25% of flow bandwidth in long-distance ISP networks. However, we note that quoting a reliable distribution of services will be increasingly difficult because many peer-to-peer applications on the Internet are disguised within protocols or encrypted and, as such, are hard to measure.

Distributed sensor networks are another interesting application. These could include extremely low-energy radio signaling between inventory and inventory control systems via RFID. While this application has the potential to be truly massive in n, it is not yet clear if the individual bandwidth requirements are substantial (e.g., on the order of video). It is also unclear whether these connections will aggregate to mostly business locations or mostly residential and whether this will have a $O(n)$ vs $O(n^2)$ effect on the core network.

One particular type of business service that has been highly touted to have a large impact on the core network is grid computing and e-science applications. These include the distributed processing of massive data files from particle accelerators in different countries and advanced distributed genome processing. They

have also spurred some interest in educational and government circles for bandwidth-on-demand services or interactions among control planes to dynamically adjust topologies across network layers. In these applications, the bandwidth connections are truly massive and, for various distributed computer-to-computer interface requirements, need to bypass typical data layer protocols as found in commercial IP layer networks. These applications are of great scientific interest and have spurred a whole academic field of integrated IP/DWDM network management and switching technologies (such as optical burst switching). And while they will generate focused, dynamic, and high-bandwidth connections, these applications have not been shown to generate much impact on total core network capacity due to the paucity of locations. Current large business customers also have interest in bandwidth-on-demand. These large business customers generally have large private networks and bandwidth-on-demand provides them flexibility for their own IP-layer topologies because of the rapidity of ordering point-to-point long-distance connections for their IP-layer links. We feel that the bandwidth-on-demand model would fit both of these applications well [16].

13.4 SUMMARY

The US telecom network has been divided into access, metro, and core segments and their various stages of evolution were illustrated. While the core segment has an almost ubiquitous penetration of fiber and WDM technologies, the opposite is true for access and metro segments in most areas. Thus, while much of the optics literature and industry focuses on advanced optical technologies for the long-distance network, most of the investment and opportunity for growth resides in the metro/access portion. As study shifts through the core, metro, and access segments, it will be seen that networks evolve more slowly and more nonuniformly. Beyond network segmentation, since the optical layer is essentially the workhorse for higher network layers, an understanding of network layering is crucial to understanding the requirements and evolution of the optical layer. A prime example of this correlation is network restoration. One cannot study or propose practical restoration methodologies for the optical layer unless the functionality and formulation of the restoration problem is understood at all network layers. This chapter has attempted to provide the reader a basic understanding of commercial optical network architectures and brought out some of their more practical commercial aspects.

LIST OF ACRONYMS

1 + 1	One-plus-one (signal duplicated across both service path and restoration path; receiver chooses surviving signal upon detection of failure)

1:1	One-by-one (signal switched to restoration path upon detection of failure)
AAL	ATM adaptation layer
ADM	Add/drop multiplexer
AR	Access Router
ATM	Asynchronous transfer mode
B-DCS	Broadband digital cross-connect system (cross-connects at DS-3 or higher rate)
BER	Bit error rate
BGP	Border Gateway Protocol (IP Protocol)
BLSR	Bidirectional line-switched ring
BPON	Broadband Passive Optical Network (each PON downstream carries 622 Mb/s on one wavelength and (optionally) analog video on another; upstream 155 Mb/s; up to 32 endpoints)
BR	Backbone Router
CLEC	Competitive local exchange carrier
CO	Central office
CoS	Class of Service
CPE	Customer premises equipment
CWDM	Coarse wavelength-division multiplexing
DCS	Digital cross-connect system
DLCI	Data link connection identifier
DS-0	Digital signal—level 0 [a plesiosynchronous (pre-SONET) signal carrying one voice-frequency channel at 64 kb/s]
DS-1	Digital signal—level 1 (a signal carrying 24 DS-0 signals at 1.544 Mb/s)
DS-3	Digital signal—level 3 (a signal carrying 28 DS-1 signals at 44.736 Mb/s)
DSL	Digital subscriber line
DSLAM	Digital subscriber line access multiplexer
DSX	Digital cross-connect
DVR	Digital Video Recorder
DWDM	Dense wavelength-division multiplexing
E1	European plesiosynchronous (pre-SDH) rate of 2.0 Mb/s
EPON	Ethernet Passive Optical Network (each PON downstream carries 1 Gb/s; upstream 1 Gb/s; up to 32 endpoints)
FCC	Federal Communications Commission (an agency of the US Government)
FE	Fast Ethernet (100 Mb/s)
FRR	Fast Reroute
GigE	Gigabit Ethernet (nominally 1000 Mb/s)
GPON	Gigabit-per-second-capable Passive Optical Network (each PON downstream carries 2.4 Gb/s on one wavelength and

	(optionally) analog video on another; upstream 1.2 Gb/s; up to
	endpoints)
HD	High definition (short for HDTV)
HDTV	High definition (television with resolution exceeding 720 x 12
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
IGP	Interior Gateway Protocol
IOS	Intelligent optical switch
IP	Internet Protocol
IPTV	Internet Protocol television (i.e., entertainment-quality vid
	delivered over IP)
IS-IS	Intermediate-system-to-intermediate-system
ISO	International Organization for Standardization (not an acrony
ISP	Internet Service Provider
ITU	International Telecommunication Union
LATA	Local access and transport area
LGX	Lightguide cross-connect (fiber patch panel)
LOS	Loss of service
LSA	Link state advertisement
LSP	Label-switched path
MAN	Metropolitan Area Network
MPLS	Multi-Protocol Label Switching
MSP	Multi-Service Platform
MTBF	Mean time between failure
MTTR	Mean time to repair
N-DCS	Narrowband digital cross-connect system (cross-connects at DS
	rate)
NPE	Network premise equipment
OC-*n*	Optical carrier—level *n* (designation of optical transport of
	SONET STS-*n*)
O-E-O	Optical-to-electrical-to-optical (3R regeneration)
OL	Optical layer
OSPF	Open shortest path first
PBX	Private branch exchange
PIM	Protocol independent multicast
PL	Private Line
P-NNI	Private network-to-network interface
POP	Point of presence
PON	Passive Optical Network (access network with only pass
	components (glass) in the distribution plant; no electric pow
	needed except at CO and endpoints)
POS	Packet over SONET/SDH
PPP	Point-to-Point Protocol
PPPoE	Point-to-Point Protocol over Ethernet

PVC	Permanent virtual circuit
PWE3	Pseudo-wire emulation edge-to-edge
PXC	Photonic cross-connect
QoS	Quality of service
RF	Radio frequency
RFID	Radio frequency identification (a technology for very low power, low bandwidth communication using small tags)
ROADM	Reconfigurable optical add/drop multiplexer
RST	Rapid spanning tree
RT	Remote terminal
RTP	Real Time Protocol
SD	Standard definition (television with resolution of about 640×480)
SDH	Synchronous digital hierarchy (a synchronous optical networking standard used outside North America, documented by the ITU in G.707 and G.708)
SLA	Service level agreement
SLRG	Shared-link-risk group
SONET	Synchronous Optical Network (a synchronous optical networking standard used in North America, documented in GR-253-CORE from Telcordia)
STS-n	Synchronous transport signal—level n (an electrical signal level of the SONET hierarchy with a data rate of $n \times 51.84$ Mb/s)
SVC	Switched virtual circuit
TCA	Threshold crossing alert
TDM	Time division multiplexing
ULH	Ultra-long haul
UPSR	Unidirectional path-switched ring
VCAT	Virtual concatenation
VCI	Virtual circuit identifier
VoD	Video on demand
VLAN	Virtual Local Area Network
VoIP	Voice over Internet Protocol
VPLS	Virtual Private LAN Service (i.e., Transparent LAN Service)
VPN	Virtual Private Network (IP Protocol)
VT1.5	Virtual Tributary group (encapsulation of asynchronous DS-1 inside SONET STS-1) payload; other VT groups (VT2, VT3, VT6) are similarly defined
WAN	Wide Area Network
W-DCS	Wideband digital cross-connect system (cross-connects at DS-1 or SONET VTn rate)
WDM	Wavelength-division multiplexing
xDSL	various digital subscriber line technologies, where x $\in$ {null, A, H, or V}
xPON	various Passive Optical Network technologies, where x $\in$ {B, E, or G}

REFERENCES

[1] H. Zimmermann, "OSI reference model—The ISO model of architecture for open system interconnection," *IEEE Transactions on Communications*, 28(4), 425–432, April 1980.

[2] IETF, Pseudo Wire Emulation Edge to Edge (PWE3) Working Group, http://www.ietf.org/html.charters/pwe3-charter.html.

[3] J. Rosenberg and H. Schulzrinne, "Signaling for Internet Telephony," in Proc. *ICNP 1998*, Sixth International Conference on Network Protocols, IEEE, pp. 298–307.

[4] http://www.telecomspace.com/latesttrends-ims.html

[5] S. Chaudhuri, G. Hjálmtýsson, and J. Yates, "Control of lightpaths in an optical network," *Optical Internetworking Forum Submission*, OIF2000.04, January 2000 and *IETF Internet Draft*, February 2000.

[6] R. D. Doverspike, J. Morgan, and W. Leland, "Network design sensitivity studies for use of digital cross-connect systems in survivable network architectures," *IEEE J. Sel. Areas in Commun. (JSAC)*, Issue on Integrity of Public Telecommunication Networks, 12(1), 69–78, 1994.

[7] C.-W. Chao, H. Eslambolchi, P. Dollard et al., "FASTAR – A Robust System for Fast DS3 Restoration," in Proc. *GLOBECOM'91*, Phoenix, Arizona, December, 1991, pp. 1396–1400.

[8] T. Sheldon, "Encyclopedia of Networking & Telecommunications," McGraw-Hill, 2001.

[9] R. Doverspike, S. Phillips, and J. Westbrook, "Future transport network architectures," *IEEE Commun. Mag.*, 37(8), 96–101, August 1999.

[10] K. Oikonomou, R. Sinha, and R. Doverspike, "Multi-layer network performance and reliability analysis," Operations Research, to appear.

[11] G. Li, D. Wang, R. Doverspike et al., "Economic Analysis of IP/Optical Network Architectures," in Proc. *OFC 2004*, Los Angeles, CA, March 2004.

[12] G. Li, D. Wang, J. Yates et al., "IP over optical cross-connect architectures," *IEEE Commun. Mag.*, 45(2), 34–39, February 2007.

[13] T. Afferton, R. Doverspike, C. Kalmanek, and K. K. Ramakrishnan, "Packet-aware transport for metro networks," *IEEE Commun. Mag.*, 120–127, March 2004.

[14] J. Wei, K.K. Ramakrishnan, R. Doverspike et al, "Convergence through Packet-Aware Transport," *The Journal of Optical Networking*, Special Issue on Convergence, 5(4), 221–245, April 2006.

[15] A. Odlyzko, "Resource/traffic Management Architectures for NGI." http://www.dtc.umn.edu/~odlyzko/talks/itc2007.pdf

[16] S. Beckett and M. Lazer, "Optical Mesh Service, Service Strategy Capitalizing on Industry Trends," http://www.oiforum.com/public/documents/061016-AT&T.pdf

15

Future optical networks

Michael O'Mahony

Department of Electronic System Engineering, University of Essex, Colchester, UK

(Portions reprinted, with permission, from *Journal of Lightwave Technology*, Vol. 24, No. 12, December 2006, © 2006 IEEE)

15.1 INTRODUCTION

All developed regions of the world are experiencing huge growth in the volume of digital data being generated. Figure 15.1 shows the historical growth, where for example between 2000 and 2006 the volume of data grew by a factor >50, from 3 to 160 BGB, an unimaginable growth a decade or so ago. In this context, it is also predicted that by 2010, 70% of this data will be generated by consumers. Although all of this data will not flow across networks, increasing amounts will, and hence current networks must evolve to support this staggering growth and provide appropriate services to the end users for manipulating and processing their data.

This growing data wave and its consequent demands on communication networks for capacity and services arise from many different user communities spread across the globe, a phenomenon which has accelerated in the twenty-first century. Figure 15.2 illustrates the global networking situation, wherein different user groups have their own networking domains, but increasingly wish to interconnect and share networking resources. These networks may, e.g., be domestic/national telecommunication networks, national research and educational networks (NRENs) (e.g., interconnecting national research and educational institutes), major research test beds (e.g., funded by governments), or enterprise/business networks. Currently, major processing and storage resources (such as supercomputers) distributed on a global basis are shared by high-end (scientific) users (grid computing) and will likely, in time, become accessible to domestic users and other groups; however, moving from specialized networks designed for thousands of scientific users to one designed for sharing globally distributed resources to millions of users requires significant research and development. Requirements from some of these users groups are as follows:

Optical Fiber Telecommunications V B: Systems and Networks
Copyright © 2008, Elsevier Inc. All rights reserved.
ISBN: 978-0-12-374172-1

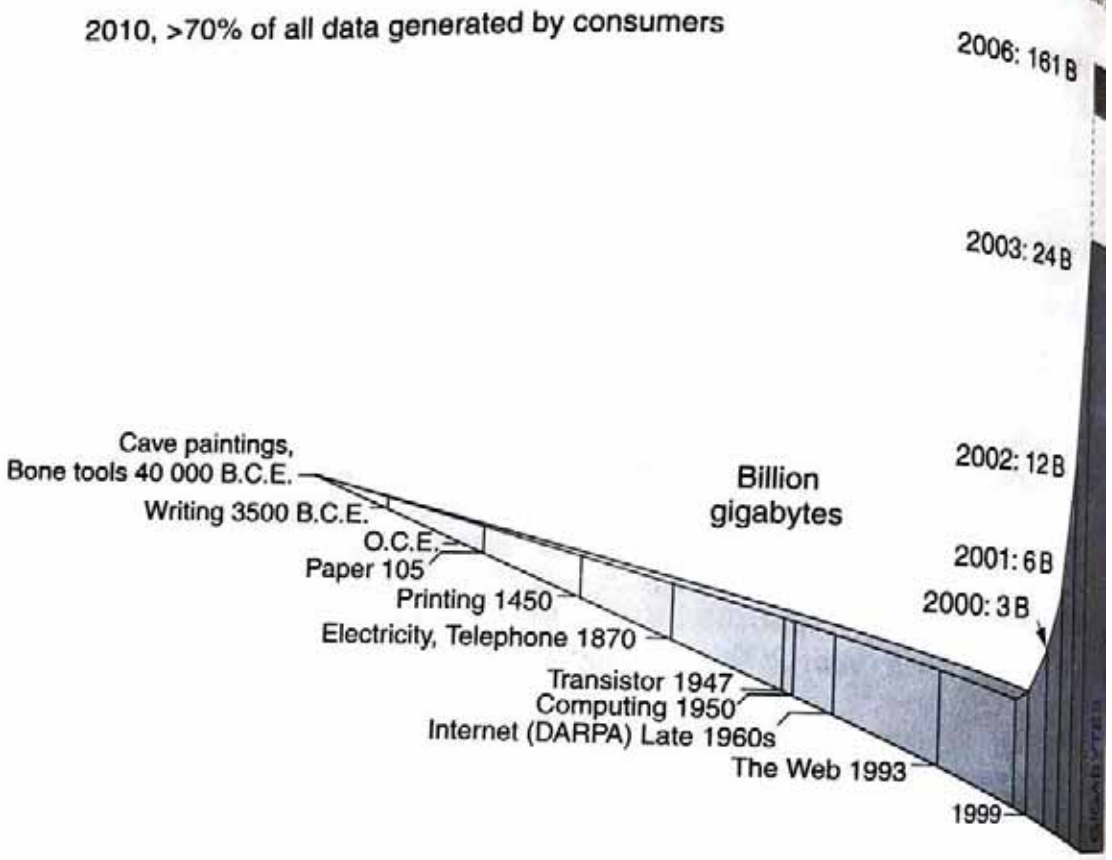
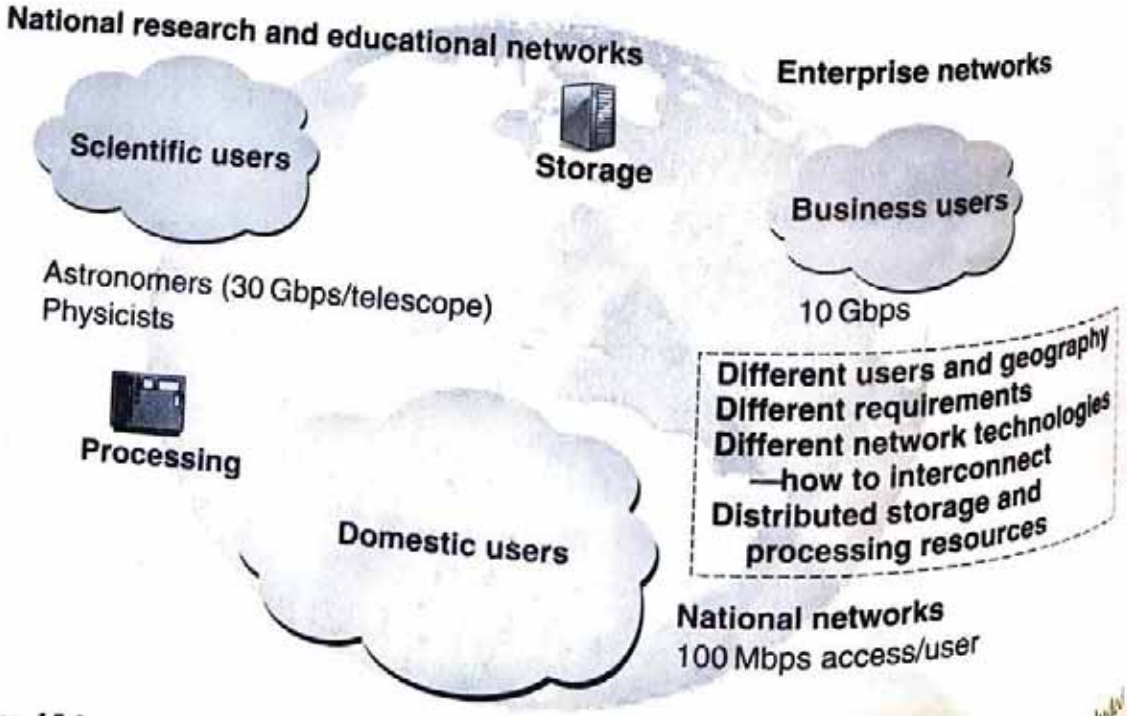


Figure 15.1 Growth of data (derived from idea by da Silva, Director Network & Communicati Technologies, CEC.) (This figure may be seen in color on the included CD-ROM).



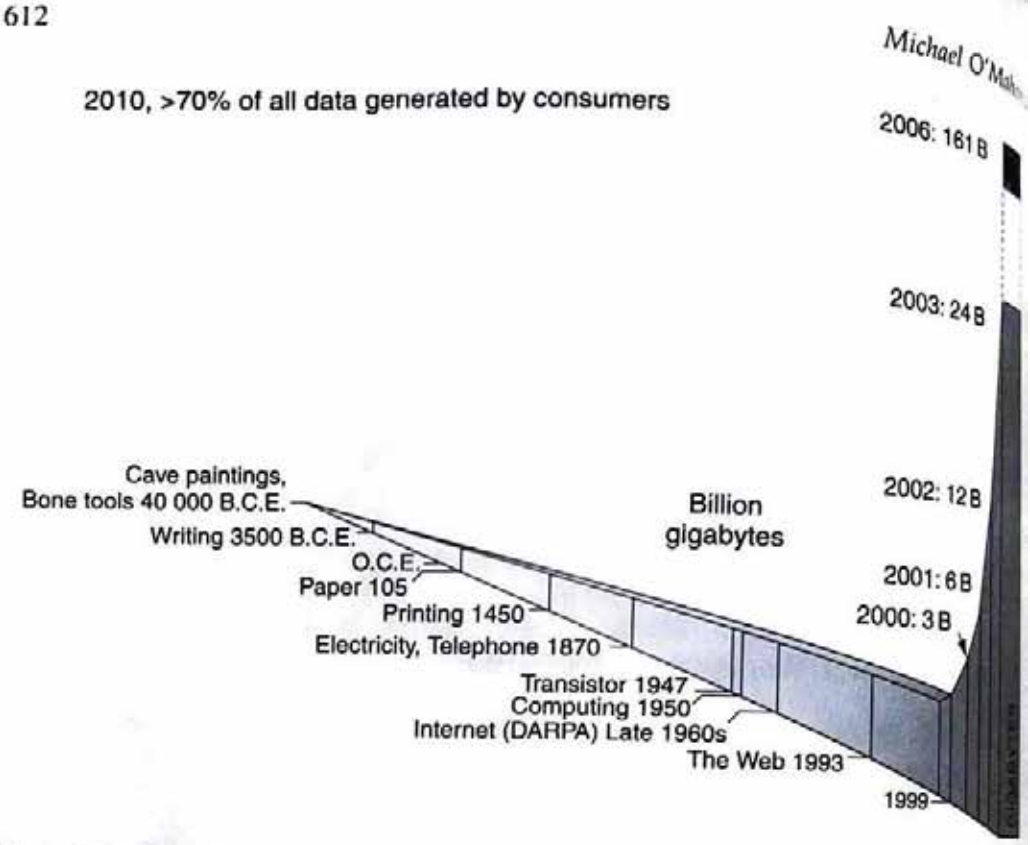


Figure 15.1 Growth of data (derived from idea by da Silva, Director Network & Communication Technologies, CEC.) (This figure may be seen in color on the included CD-ROM).

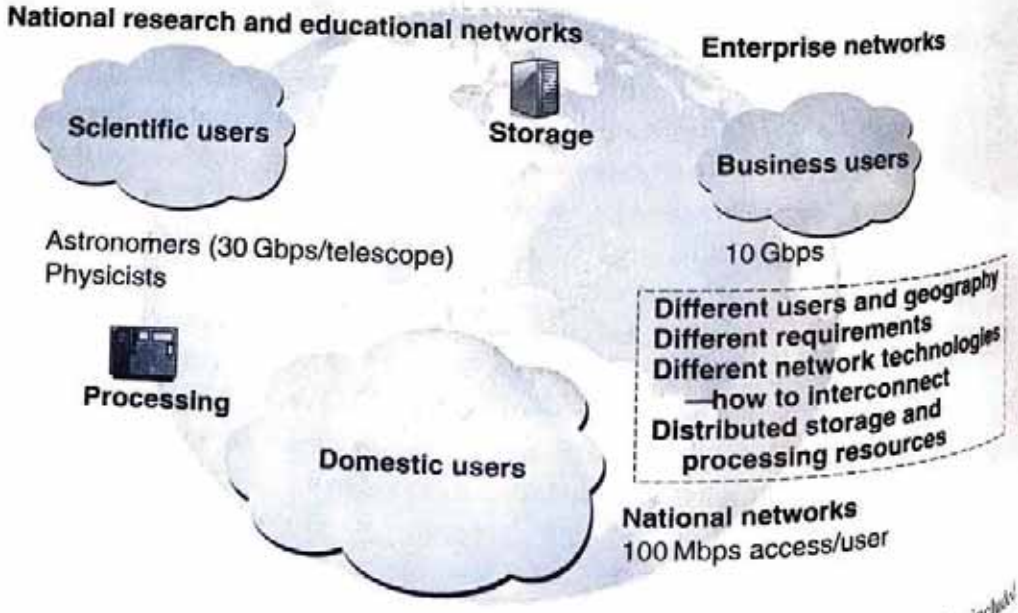


Figure 15.2 User communities and networks (this figure may be seen in color on the included CD-ROM).

15.1.1 Domestic Users

Based on national telecommunication networks, domestic users will increasingly look to store and manipulate video images, play interactive games, and download increasingly large files. Much of this capability is enhanced through the greater penetration of broadband access, which in its ultimate form will be provided via optical fiber interconnections offering bandwidth of 100 Mbps, as currently in countries such as Japan [1]. As summarized in Table 15.1, access at 100 Mbps is seen as a target that would support most services being considered, including "Triple Play" services, which include voice data and video; mobile users will also require high-bandwidth access, estimated as 30 Mbps in this model. These figures would be greatly increased, if super high definition TV would also become a broadcast service, requiring Gbps access speeds.

It is likely that, in the future, such users will make use of network resources such as storage (already BT Vault, in the UK, is a service for storage) and processing power. As mentioned above, networking technologies such as the grid computing [2], where global computing resources can be accessed for any location, will become available to domestic users (consumer grids), enabling new services, e.g., video processing, to be done online. In the era of mobility, it is likely also that personal storage will follow the user across the globe, so that wherever he/she is, all personal data can be instantly accessed. Major challenges here include security, as well as adaptable network architectures.

15.1.2 Large Business/Enterprise Users

They require access to high symmetric bandwidths (e.g., up to 10 Gbps) for virtual private network (VPNs), disaster recovery, storage, etc. Such services will be supplied through Fiber to the Premises (FTTP) technologies.

Table 15.1

Residential service requirements.

Application	Downstream requirement	Upstream requirement
HDTV (3 per home at 20 Mbit/s each) Standard TV = 4.5 Mbit/s	60 Mbit/s	<1 Mbit/s
Online gaming	2–20 Mbit/s	2–20 Mbit/s
VoIP Telephone (3 per home at 100 kbit/s)	0.3 Mbit/s	0.3 Mbit/s
Data/Email, etc.	10 Mbit/s	10 Mbit/s
DVD download for rental Assume download must take <10 mins, i.e., ~ the time to get one from a rental store	14 Mbit/s	<1 Mbit/s
Total	**~100 Mbit/s**	**~30 Mbit/s**

Scientific Users (High-End Users)

These users are currently served by (NRENs) or large dedicated test beds. Application examples include:

High-Energy Particle Physics: The next generation of experiments at the Large Hadron Collider (LHC) in European Laboratory for Particle Physics (CERN) will produce data sets measured in tens of petabytes per year that can only be processed and analyzed by globally distributed computing resources. Experiments require deterministic transport of 10–100 terabyte data sets, and a 100 terabyte data set requires a throughput of 10 Gbps for delivery within 24 hours. Thus optical network services will be crucial to this discipline where dedicated and guaranteed bandwidth is required for periods of days.

Very-Long-Baseline Interferometry (VLBI): VLBI is used by radio astronomers to obtain detailed images of cosmic radio sources, where the combination of signals from two or more widely separated radio telescopes can effectively create an instrument with a resolving power proportional to their spatial separation. e-VLBI [3] will use high-speed networks to transfer telescope data to a correlator and the availability of optical network services at multi-Gbps (10–40 Gbps) throughput will greatly increase capability.

e-Health: Remote mammography poses challenges for the deployment of supporting IT systems due to both the size and the quantity of images, with networks required to transport 1.2 GB of data every 30 s. The availability of optical network services offering real-time guarantees is important in this field.

15.2 REGIONAL ACTIVITIES

With these new demands for data services, the traditional voice-centric telecommunications network is in the process of transforming to a data-centric network. The vision is to have a communications environment comprising wired and wireless networks where any individual can access information and network resources in an effortless manner. To achieve this, the ultimate network will have as its building blocks a fixed optical network platform, accessed though wireless (Wi-Fi, WiMax, UMTS) and wired infrastructures, such as FTTP. Such an all-pervasive networked society, however, puts increased demands on network reliability and security.

The importance of future national network infrastructures is mirrored in the ongoing discussions on how a new Internet might be realized. It has been commented [4] that the Internet is expanding from an "Information Service" to a "Critical Infrastructure" for all aspects of society and so new network architectures must evolve to overcome many of the problems inherent in the current Internet. NRENs as well as experimental network test beds are currently being used (in part) to understand how a new Internet might be constructed. The inference is that at all levels communication networks are no longer just desirable but critical for a nation's development and security.

Regional views on expected growth in capacity demands, traffic profiles of new services, and the need to move to data-centric networking are reflected in the scope of the major research programs and their associated projects funded by regional governments.

In *Europe*, currently there are no overarching roadmaps for future networks, but a number of very large "Integrated Projects" funded through the European Commission (EC) identify their own vision of the future. The NOBEL project [5], e.g., studies the evolution of core and metropolitan optical transport networks, supporting end-to-end quality of service (QoS), with intelligent data-centric solutions based on automatic switched optical network (ASON) [6] and generalized multiprotocol label switching (GMPLS) [6], together with optical burst and packet switching; these topics are discussed later in this chapter.

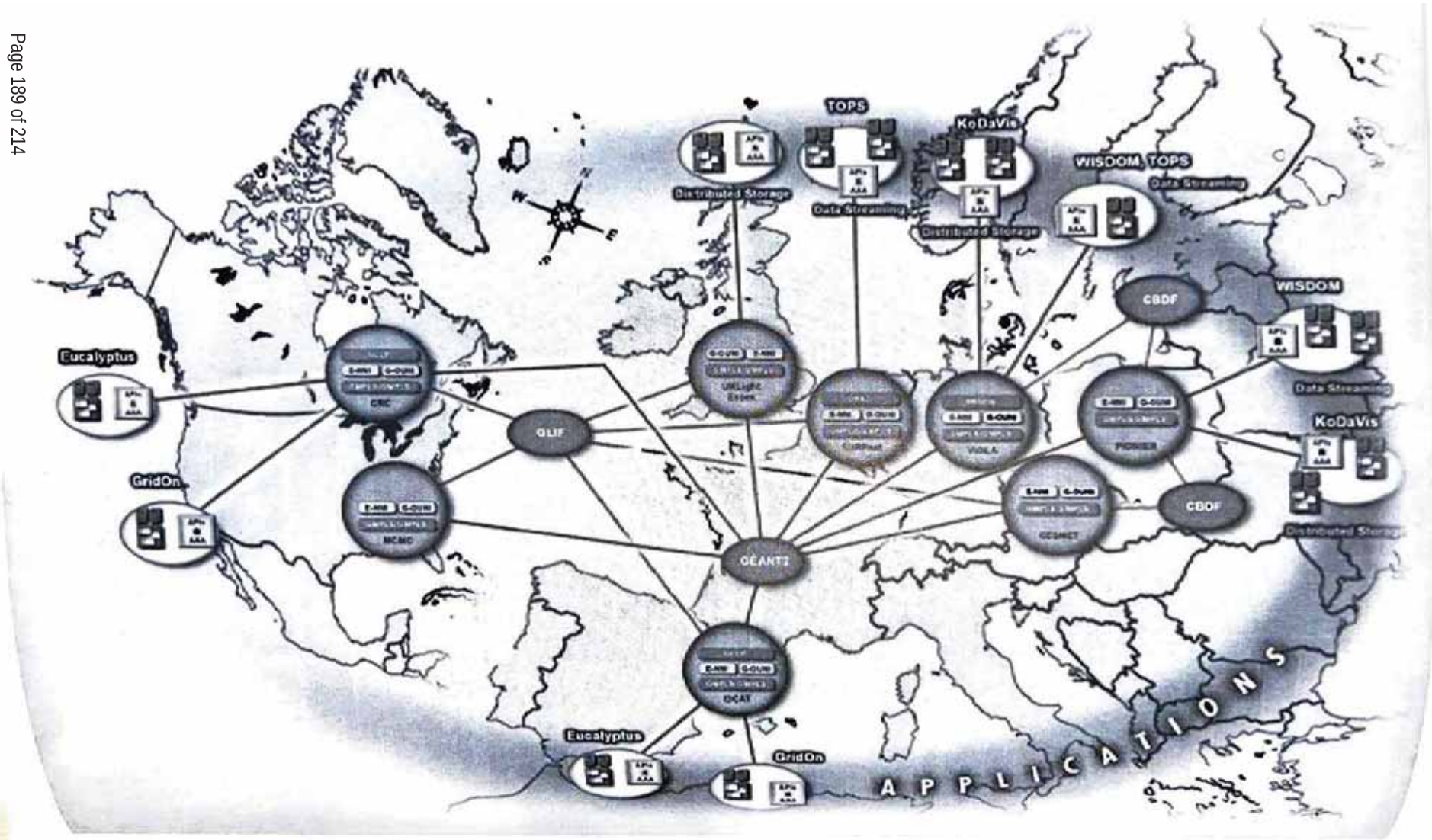
Optical packet switching (OPS) has had long-term support within the EU and national programs, with projects such as DAVID [7] and OPSNET/OPORON [8, 9] developing the technology and its application over a period of about 15 years; however, it is still seen as a very future technology. By contrast, optical burst switching (OBS) and GMPLS approaches are viewed as realistic possibilities for more near-term deployment, and research in these areas is also echoed in national funding in many countries.

Recently, BT (UK) has committed itself to moving to a converged national network solution (BT 21CN Network), with an Internet protocol/multiprotocol label switching (IP/MPLS) core [10]. This change is also under way in other countries (Netherlands, Australia) and is significant in that the architecture opens the path to a full optical transport network at some point in the future, with the possible replacement of optical-electronic-optical (OEO) switches and regenerators with optical-optical-optical (OOO) technologies.

Europe is characterized by a large number of NRENs which exist in most member countries. These networks are nationally funded and commonly used for research into scientific applications (of the type discussed above). The EC funds an overlay network GEANT [11], which provides interconnections between these national networks and provides international connections, e.g., to the USA. These networks look to lambda networking based on scheduled or dynamic provision of lightpaths.

The Pacific Rim countries of Japan and Korea are well known for their ambitious plans in relation to broadband deployment and enhanced network infrastructures.

Japan has well-defined R&D programs; some are initiated by the Japanese Government with a view to evolving a legacy telecommunications network to an IP over wavelength division multiplexing (WDM) network. Such programs move in tandem with operator plans, e.g., NTT plan to migrate 30 million customers to FTTP and IP telephony by 2010. The most recent program focused on developing an all-optical transport network with terabit capability [12]; this research comprised the study and development of photonic nodes such as fast optical cross-connect (OXC) based on MEMS together with control plane (such as GMPLS), OBS, and high-bit-rate, ultralong transmission based on dense wavelength division multiplexing (DWDM) (up to 1000 channels) and optical time division


TOPS
KoDaVis
WISDOM, TOPS
Distributed Storage
Data Streaming
Distributed Storage
CBDF
WISDOM
Eucalyptus
GLIF
KoDaVis
GridOn
CBDF
GEANT2
Distributed Storage
Eucalyptus
GridOn
APPLICATIONS

APPLICATIONS
WISDOM
KoDaVis
CSDF
WISDOM TOPS
TOPS
GLIF
GÉANT2
OrdiOn
Eucalyptus

multiplexing (OTDM). For photonic routing, targets are related to the feasibility of 10-Tbps routers and network architectures appropriate to a Tb-class wavelength-routed optical network. The current program seeks to understand how this optical platform can support new bandwidth, demanding applications such as grid computing, and real-time applications like video, digital cinema, and network storage. Key targets include 160 Gbps multilevel transmission systems, using DQPSK/QAM, etc. aiming at bandwidth utilization of more than 2 bits/Hz by 2010.

In the *USA* funding for research (nonindustrial) is through the National Science Foundation and for large projects often through DARPA. Current major photonic activities include studies on optical code division multiple access (OCDMA) and terabit router technology, the latter represented by projects IRIS [13] and LASOR [14], whose goal is to realize 100-Tbps routers. Both projects take the route of OPS, which offers attractions in terms of footprint and power requirements.

The strong interest in optical networking in the USA is reflected in the existence of a number of national test beds, which enable the interconnection of scientific users and support research into future networks. For example, National LambdaRail (NLR) [15] is a high-speed national computer network which is also used as a network test bed for experimentation with next-generation large-scale networks. Links in the network use DWDM at 10 Gbps/channel. NLR's services are already in use by many network research projects, e.g., the NSF OptIPuter project and Internet 2's Hybrid Optical Packet Infrastructure (HOPI) project, which looks at a future infrastructure comprising an IP core network together with an optically switched wavelength set, for dynamic provisioning of high-capacity paths. Currently being proposed is a new national facility GENI [16], which includes a global experimental facility designed to explore new network architectures with the broad scope of understanding new paradigms for Internet-type networks.

The global nature of communications means that there is much interaction and common research between world regions. Organizations like Global Lambda Integrated Facility (GLIF) [17] provide a structure to enable global test beds to be interconnected to undertake research activities of common interest. Thus there exists a global test bed interconnection, as illustrated in Figure 15.3.

Figure 15.3 Global test bed (Source: IST-PHOSPHOROUS Project-D. Simeonidou.) (This figure may be seen in color on the included CD-ROM).

15.3 A BRIEF HISTORY OF OPTICAL NETWORKING

The invention of the laser by Schawlow and Townes in 1958 followed by the work of Kao and Hockham on optical fibers (in 1965) and the subsequent demonstration of optical fiber as a practical communication medium by Maurer, Keck, Schultz, and Zimar in 1970 brought into being a technology platform capable of supporting national and global communication requirements for the twenty-first century and beyond. In the late 1970s, fiber began to replace coaxial cable as the transmission medium in the trunk systems of telecommunication networks bringing many advantages both technical and economic. The creation of the Internet (with TCP/IP) in 1983 and subsequently the World Wide Web in 1993 sparked the growth of data

Michael O'L

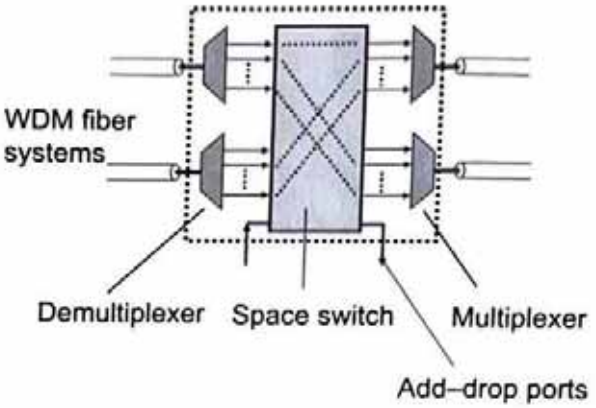




Figure 15.4 Optical cross-connect (this figure may be seen in color on the included CD-ROM)

traffic on the network, and in 2002, or thereabouts, the amount of network data traf exceeded that of voice traffic. In the decade from 1985 to 1995, four significa events heralded the possibility of optical networking where both transmission a switching might be based on optics. These were (1) the realization of optic amplifiers allowing (2) the economic deployment of WDM, (3) the demonstratio of an OXC enabling the rapid reconfiguration of lightpaths based on wavelengt channels, and (4) the convergence of service and transport transmission rate.

The early view of optical networking considered an "Optical Layer" to form a extension to the existing SDH/SONET network layers. Figure 15.4 shows an OXC (and its realization), which comprises input demultiplexers, a space switch, and output multiplexers. Each incoming fiber supports a number of wavelengths; these are demultiplexed and then either switched (by the space switch) to an outpu multiplexer and hence outgoing fiber or dropped off locally. Hence the OXC is defined as a general wavelength switch which can be realized in (a) an all-optical (transparent) manner (OOO—optical input, optical switch fabric, optical output) or in an opaque manner (OEO—optical input, electrical switch fabric, optical output through choice of technologies.

Figure 15.5 shows the original concept of an optical layer, which was envisaged as an extension to the existing SDH/SONET network layers. In the UK, e.g., it was typical that about 60% of the traffic entering a main node was transit traffic; thus an OXC might be deployed within the optical layer to enable long-haul transit traffic to bypass the main switch nodes and hence reduce the size and cost of the digital cross-connects. Demonstrations of such reconfigurable networks were carried out in Europe and the USA in 1994 [18, 19].

The convergence of service and transport wavelength bit rates around the year 2000 (Figure 15.6), at a bit rate of 10 Gbps, opened the possibility of direct interfacing between, e.g., an IP network and an optical transport network



axer

orts

y be seen in color on the included CD-ROM).

bouts, the amount of network data traffic de from 1985 to 1995, four significant etworking where both transmission and se were (1) the realization of optical yment of WDM, (3) the demonstration ion of lightpaths based on wavelength and transport transmission rate.

nsidered an "Optical Layer" to form an ork layers. Figure 15.4 shows an OXC t demultiplexers, a space switch, and pports a number of wavelengths; these d (by the space switch) to an output opped off locally. Hence the OXC is ch can be realized in (a) an all-optical optical switch fabric, optical output) or lectrical switch fabric, optical output)

an optical layer, which was envisaged network layers. In the UK, e.g., it was a main node was transit traffic; thus, cal layer to enable long-haul transit ience reduce the size and cost of the such reconfigurable networks were 18, 19].

wavelength bit rates around the year ps, opened the possibility of direct and an optical transport network

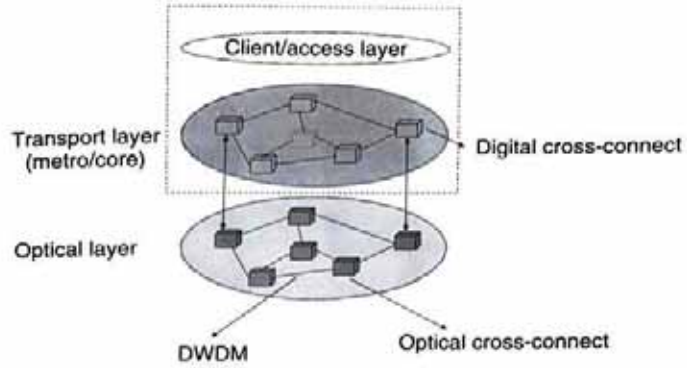


Figure 15.5 The optical layer (this figure may be seen in color on the included CD-ROM).

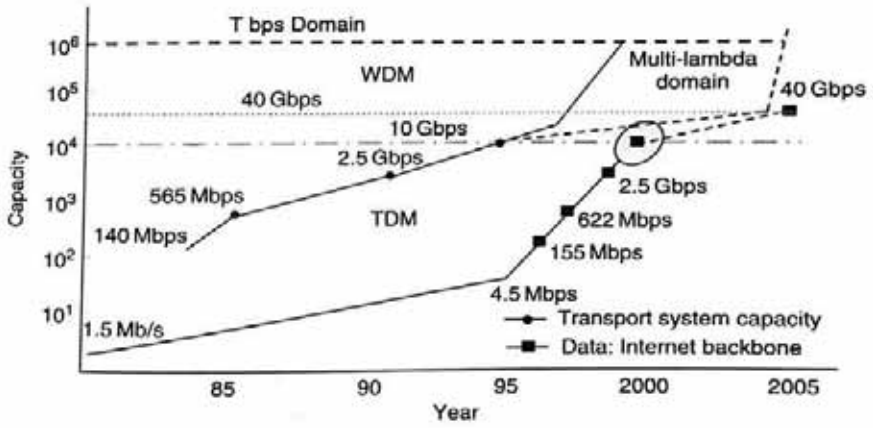


Figure 15.6 Service and line rates.

employing WDM and OXCs, where the granularity of the network directly matched the router interface rate. This was an important step in the evolution of optical networks as a router output stream could flow directly on a wavelength channel. The diagram also shows that convergence at 40 Gbps occurred in 2005, with the availability of 40-Gbps routers and engineered 40-Gbps DWDM transmission systems [20]. Each wavelength, of course, may support many traffic streams.

Figure 15.7 is the schematic of a possible future telecommunications network showing core, metro, and access layers. The core cloud represents an optical network comprising a number of nodes interconnected by amplified fiber links employing DWDM. The nodes comprise an OXC in conjunction with a network service element (SE). The OXC supports the bypass function and allows specific wavelength channels to be dropped to the SE, which e.g., could be an

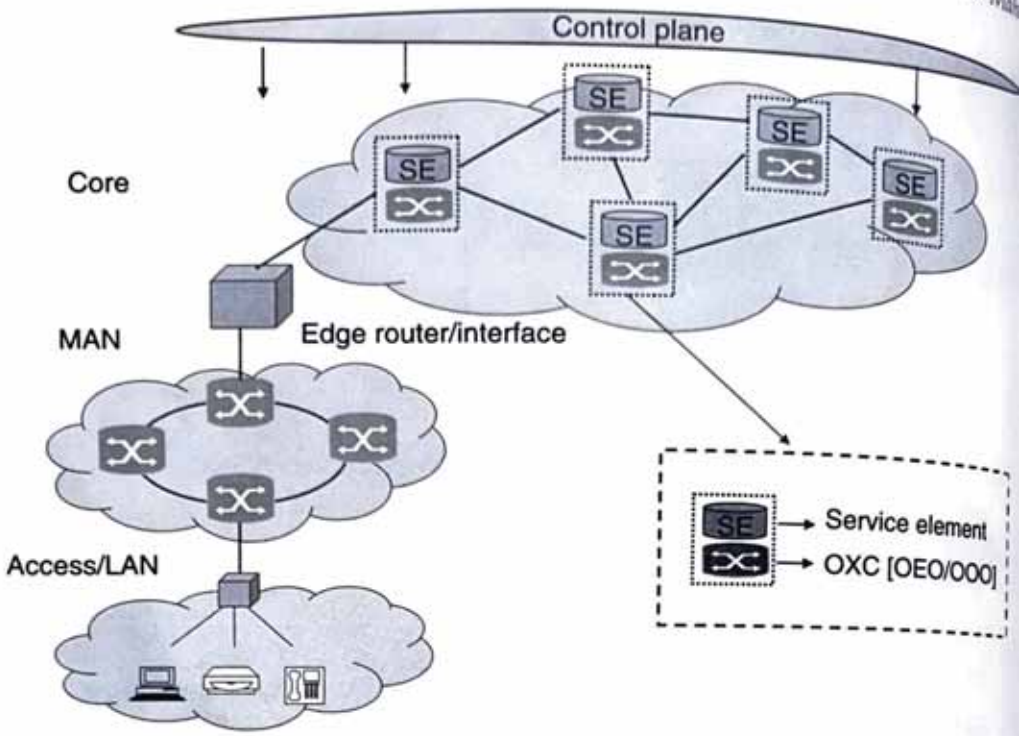


Figure 15.7 A national telecoms network (this figure may be seen in color on the included CD-RO

Dr. Dan

IP/MPLS router, an SDH/SONET digital cross-connect, or an optical burst, packe or Ethernet switch, as discussed later. At the network edge, traffic is mapped onto t network services via an edge interface/router, which can perform either Us Network Interface (UNI) or Network Network Interface (NNI) functionality defined by the Optical Internetworking Forum (OIF) [6] depending on what connected to the core. A control plane is required to establish paths across the da plane as requested at the network edge, mapping, e.g., an IP/MPLS stream from t MAN onto a specific wavelength; this path establishment can be done in a centra lized or distributed manner. The deployment of optical technology brings man potential benefits, as outlined in Figure 15.8.

Some of these are:

Transmission: Optical systems enable low-cost high-capacity systems based o high-speed channels combined with DWDM technology.

Switching: A major benefit from optical switching (discussed below) is th reduced power and size in comparison with electronics for the same throughput. This is a very important factor as exchange buildings are often limited in size, especially within urban environments.

Interoperability: As discussed above, there are many different user communities distributed globally who wish to interconnect (e.g., for grid computing) for purposes of experiments or sharing resources. Interoperabilty is a key issue as each network domain may use different technologies at data and control plane level. It is likely that networks based on optical technologies will offer much

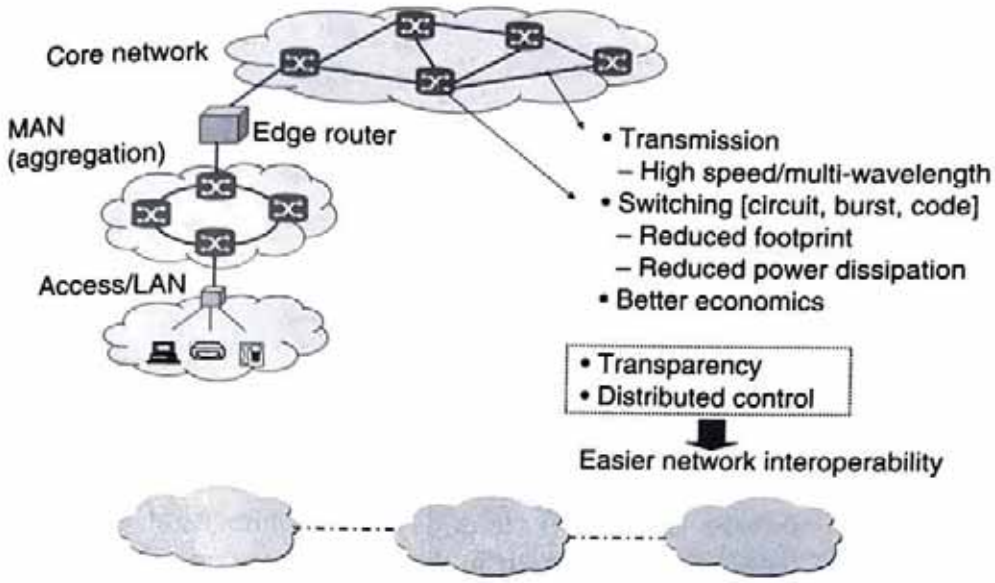


Figure 15.8 Optics and the network (this figure may be seen in color on the included CD-ROM).

greater ease of interoperability, through e.g., the exploitation of transparency (signals remaining in the optical domain across the network), allowing wavelengths at the network edge to be dynamically aligned to another domain, or support distributed control though the use of fast optical switches.

This chapter addresses some of the key optical functionalities of future optical networks, rather than the detail of all the technology issues which are inevitable part of any evolution. It is recognized that a number of functions, e.g., dispersion compensation, are currently migrating to the electronic domain, but these topics are not considered here. As discussed earlier, optical networking includes the concept of OXCs (wavelength switches) using either an OEO switching or an OOO approach. The OEO approach requires that all the wavelength paths terminating at the switch go through OE conversion prior to the switch and EO conversion following the switch; this is costly but enables a well-established electronic switch technology to be used; it also supports finer granularity switching. The OOO approach has the great attraction that no OE/EO conversions are needed, but the transparent systems it offers (with signals staying in the optical domain across the network) mean that careful design of the use of wavelengths is needed and make it likely that wavelength conversion (ideally all-optical) would be needed.

15.4 A DIVERSITY OF ARCHITECTURES AND TECHNOLOGIES

As Figure 15.7 shows, national networks comprise a number of interconnected subnetworks, access (user to local exchange), metro (interconnection of local exchanges to the core), and core network which transports data over long

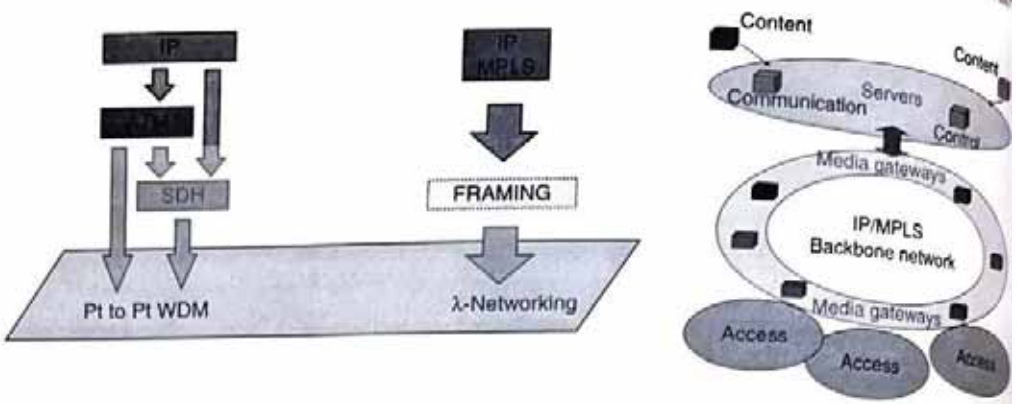


Figure 15.9 The changing network (this figure may be seen in color on the included CD-ROM).

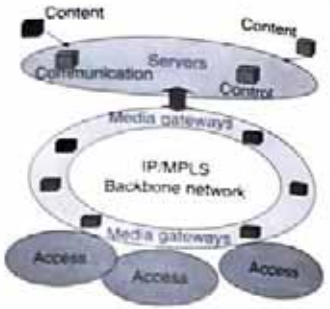
distances. Each of these subnetworks has its own architectural designs which respond to issues specific to its position in the overall network; for example, the access connects to every individual and hence cost is a major issue; the core networked carries data from many users and so its costs are shared. In this discussion, we focus on the often-called transport network, which spans metro and core carrying aggregated traffic across the network. As the total traffic across the network increases and the demand for new fast services increases, design of the transport network and its control and management are increasingly important.

Much research has focused on moving from the traditional circuit-based switching to a more dynamic and data-centric network enabling rapid lightpath reconfiguration and providing subwavelength granularity, as needed by the new applications discussed earlier. In its present form, the network has a complex layering to allow the simultaneous support of data and voice services as shown in Figure 15.9.

The left-hand side of Figure 15.9 shows how data (IP) may be encapsulated into ATM cells (or SDH/SONET frames) for transmission across the point-to-point connections in the WDM network; currently, the network is changing to a more data-oriented version of SDH/SONET [next generation (NG) SDH/SONET]. Much effort has been made to minimize this complex layering, which is expensive to maintain. The right side of the diagram shows an example (discussed again below) where the core of the network is based on IP data streams with MPLS at the network nodes—a purely data network.

Figure 15.10 outlines a possible evolution route for the network structures and technologies that may appear in the future optical (transport) network; as illustrated, it is a version of many such diagrams presented over the years. Working from the bottom left-hand corner of the diagram:

Progress toward optical networking has been much slower than envisaged in the late 1990s, and currently the first real steps toward networking are seen in the deployment of reconfigurable optical add–drop multiplexers (ROADMs), in particular those based on a multiport wavelength-selective switch (WSS) [21]; these devices have the functionality of OXCs (Figure 15.3) with an optical switch



be seen in color on the included CD-ROM).

its own architectural designs which
he overall network; for example, the
nce cost is a major issue; the core
nd so its costs are shared. In this
nsport network, which spans metro
e network. As the total traffic across
fast services increases, design of the
ment are increasingly important.
) the traditional circuit-based switch-
ork enabling rapid lightpath reconfi-
anularity, as needed by the new
form, the network has a complex
f data and voice services as shown

v data (IP) may be encapsulated into
nsmission across the point-to-point
the network is changing to a more
t generation (NG) SDH/SONET].
mplex layering, which is expensive
hows an example (discussed again
on IP data streams with MPLS at the

route for the network structures and
ptical (transport) network; as illu-
presented over the years. Working
m:
en much slower than envisaged in
toward networking are seen in the
drop multiplexers (ROADMs), in
ngth-selective switch (WSS) [21];
Figure 15.3) with an optical switch

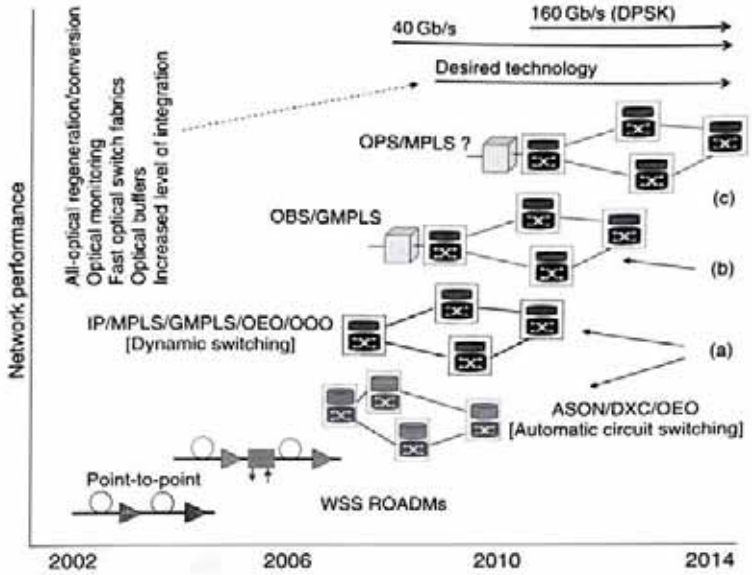


Figure 15.10 Network evolution (this figure may be seen in color on the included CD-ROM).

core; a selected wavelength on an incoming port can be dropped or transmitted under control and so represents the first introduction of optical wavelength switching.

15.4.1 Network Switching

Figure 15.10 illustrates possible further stages in the network switching evolution. In the figure, (a) represents the move to a more data-centric and dynamic switching model using an ASON [6] architecture which would allow automated lighpath provisioning and supports NG-SDH/SONET with digital cross-connect (DXC) or OXC (OEO) switching in the data plane. Figure 15.10(a) also shows that a move to an IP/MPLS (i.e., IP routers) or GMPLS (with OEO or OOO wavelength switches) architectures is foreseen, which provides an enhanced dynamic capability. GMPLS allows all transport modes, circuits, burst, and packets to be supported and can be deployed in either a centralized or a distributed mode. This represents one of the options to build a "converged network," where the backbone is a multitechnology IP/GMPLS/OEO/OOO network supporting all services (voice, data, video), which may overtake the ASON architecture. It is also the case that in recent times Carrier Ethernet [22] (based on native Ethernet or MPLS) looks increasingly attractive across all layers of the network, and indeed within the UK some (small) network providers already operate national converged networks with Ethernet switching

elements. The move toward 100 GbE standards illustrates the importance of technology and hints of future major roles in the next-generation networks. Fig 15.10(b) represents a move to a user-centric design, based on OBS with GMP (OBS/GMPLS); this technology provides subwavelength granularity and is also interest to future optical grid networks [23]. Finally, Figure 15.10(c) represents move to an OPS network where MPLS provides a common (across electric optical domains) control plane. OPS offers the finest granularity and is still seen the ultimate switching technique, but its success will depend on many technolo advances; these options are discussed in more detail in the following sections.

A recent development relating to network architectures is the move towa programmable networks. Figure 15.11 illustrates the concept. In this case, the co and edge nodes of the network are directly associated with electronic processo allowing node functions to be dynamically altered.

The Metro/aggregation network (Figure 15.8) delivers data from the access the edge device for processing, and will likely use DWDM supporting, e.g., Carri Ethernet, but it is the access network that provides the key to the future, particular the move toward FTTP.

Access-PONs: FTTP supports the Passive Optical Network (PON) concep which has been studied for over 20 years, but now increasingly deploye PONs offer multiservice (voice, data, video, and telemetry) and multiprotoc (IP, TDM, ATM) support and thus are a very flexibly infrastructur A number of possibilities have been studied, which exploit the basic PO concept, but with a view to reduce costs. Two examples follow:

Long-reach PONs: Some operators [24] see the possibility of merging metr and access in a long-reach (amplified) PON. Optical amplification included to boost the power budget and increase bandwidth, range, an number of splits. The long reach is used to bypass the metro network an terminate at a core edge node; this enables the removal of the loc

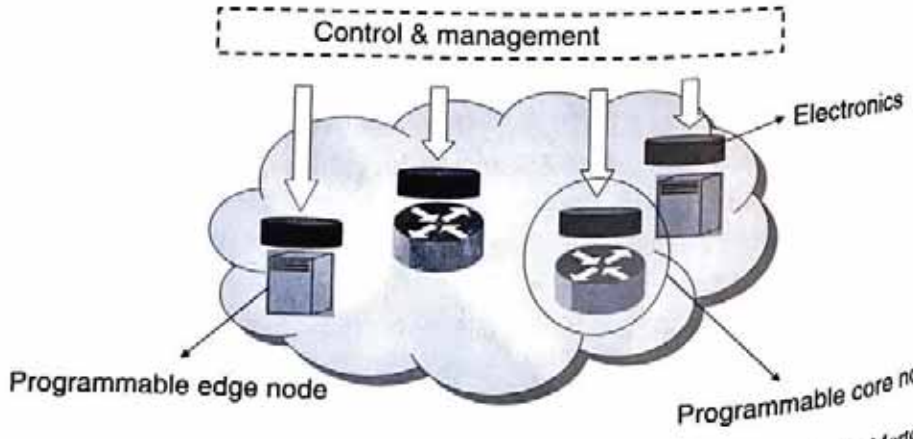
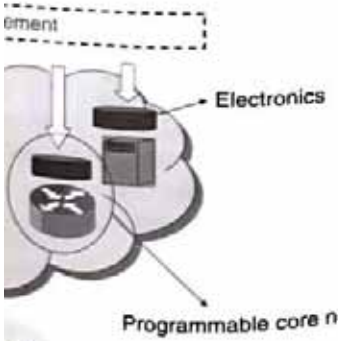


Figure 15.11 Progammable Networks (*Source*: FON Workshop OFC 2006: P Morton, GEN Simeonidou IST-PHOSPHOROUS.) (This figure may be seen in color on the included CD-ROM)

dards illustrates the importance of this
in the next-generation networks. Figure
ic design, based on OBS with GMPLS
ubwavelength granularity and is also of
. Finally, Figure 15.10(c) represents the
provides a common (across electrical/
the finest granularity and is still seen as
ıccess will depend on many technology
ıore detail in the following sections.
ork architectures is the move toward
ıtrates the concept. In this case, the core
associated with electronic processors,
altered.
15.8) delivers data from the access to
ly use DWDM supporting, e.g., Carrier
ıat provides the key to the future, in

sive Optical Network (PON) concept,
years, but now increasingly deployed.
video, and telemetry) and multiprotocol
us are a very flexibly infrastructure.
studied, which exploit the basic PON
osts. Two examples follow:

24] see the possibility of merging metro
ıplified) PON. Optical amplification is
lget and increase bandwidth, range, and
is used to bypass the metro network and
this enables the removal of the local



ON Workshop OFC 2006: P Morton, GENI:
be seen in color on the included CD-ROM).

exchange or remote concentrator site. In the UK this would require 100 of these core edge nodes with long-reach spans of 100 km and bit rates of at least 10 Gbps.

CWDMA/OCDMA PONs: Coarse WDM (CWDM) allows low-cost WDM to be deployed (as device cooling is not required); CWDM uses channel spacing of 20 nm, so only eight or so channels can normally be deployed. OCDMA (see below) has been demonstrated [25, 26] as a robust complementary technology that could also be deployed in conjunction with CWDM to increase the number of users. As with WDM, OCDMA offers the possibility of translation from one channel to another (code translation), opening many interesting possibilities.

Transmission speed

Current networks employ amplified DWDM systems with individual channel bit rates up to 10 Gbps to connect main switching centres, and currently 40-Gbps systems are being deployed. There are many reasons for believing that bit rates will increase beyond 10 Gbps and perhaps even to 160 Gbps (Figure 15.6). For example, (a) in the past it has always been advantageous to move to high bit rates from a cost viewpoint; (b) although optical amplifiers are reasonably bit rate agnostic, most of the more advanced functions currently considered for future networks such as all-optical regeneration, wavelength conversion, dispersion compensators, etc. can only operate on a single wavelength; (c) when future OOO switches are deployed then increasing the bit rate can help in reducing the port count required of the switch fabrics [26], as large OOO switch fabrics are difficult to realize, e.g., in the case of an optical packet switch where it is difficult to scale fabrics; for wavelength switching, waveband approaches [24] can be used to mitigate these technology issues, and (d) studies show that as the overall traffic on a network increases then the optimum line rate also increases—thus as network traffic increases by factors of 10 or 100 over the next decade, high bit rates may be of increasing interest. Increasing bit rate, however, leads to a more demanding requirement for system design to minimize the effects of dispersion (chromatic and polarization) and nonlinear effects; thus there is the increasing interest in modulation techniques such as differential phase-shift keying (DPSK), which is more robust to transmission impairments than intensity modulation; PSK systems are already being deployed and 160-G systems are being investigated in field trials.

15.5 KEY TECHNOLOGIES AND SUBSYSTEMS

15.5.1 Switching Technologies

Figure 15.12 is the schematic of the switching options currently under consideration for future networks, namely optical circuit, burst, and packet switching, together with optical code switching and possible hybrid solutions involving

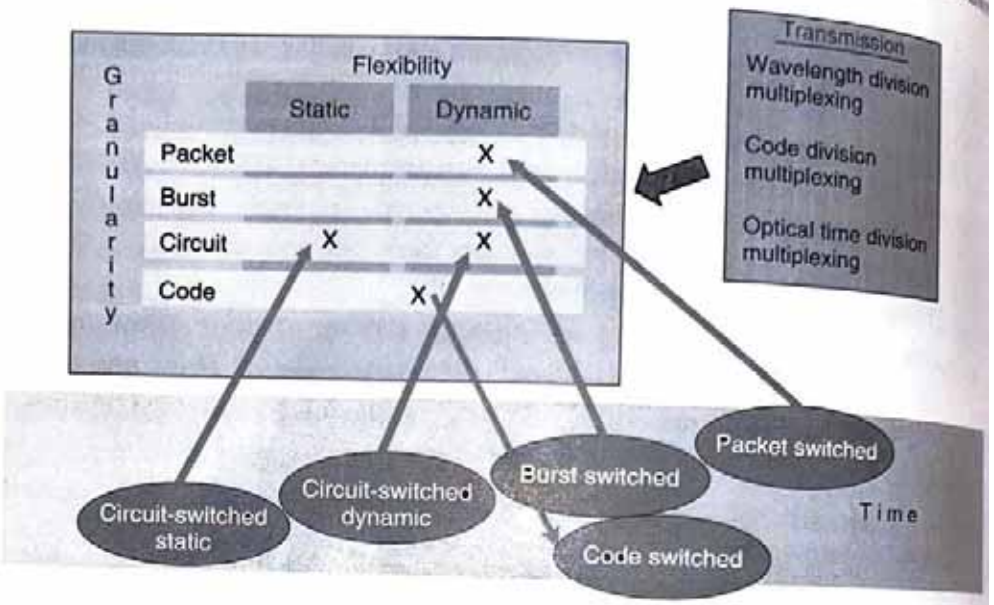


Figure 15.12 Network switching options (*Source*: IST-OPTIMIST Project.) (this figure may be seen in color on the included CD-ROM).

burst and circuit switching. The diagram illustrates the key features of these differing options, particularly in respect to granularity and flexibility. Granularity concerns, in these cases, the ability to have many users operate across one wavelength channel, thus as described below, burst-, packet-, and code-based systems demonstrate this feature; in contrast, circuit switching simple sets up a wavelength path for a specific information stream. Flexibility implies the ability to dynamically respond to network demands, setting up and tearing down new paths etc. All the technologies discussed can be deployed in a dynamic manner by the incorporation of optical switches and a suitable control plane to enable rapid configuration. As we move to more dynamic and granular solutions of course the technology becomes more complex and deployment moves out in time. These switching options are now discussed in greater detail.

Dynamic Optical Circuit Switching

Circuit switching technology, where end-to-end circuits are established upon request, is currently moving toward the deployment of fast and automatically reconfigurable nodes with switching granularity at the wavelength level and is represented by the ASON architecture standard. To enable dynamic networking the provisioning of lightpaths is automated; this is illustrated in Figure 15.13. The diagram shows an advanced control plane (associated with the data plane) which provides the necessary signaling capabilities for configuring paths through the network. The user can request, through the UNI, a new lightpath, and the controllers in the network nodes exchange, through the NNIs, the necessary information for the lightpath setup (allocation of wavelengths, etc.) activating the node

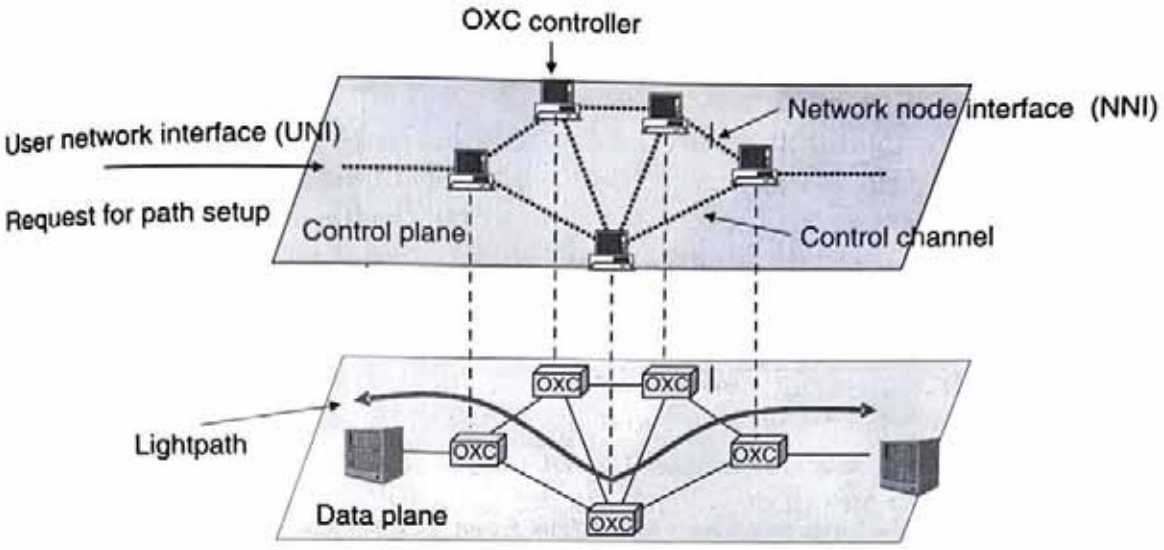


Figure 15.13 Dynamic circuit switching (this figure may be seen in color on the included CD-ROM).

switches. The node switches themselves can be DXCs or OEO–OXCs as illustrated in Figure 15.13.

As currently viewed, the ASON network, which is based on NG-SDH/SONET, has advantages in terms of QoS, management, and security. However, due to the domination of IP-centric traffic, future solutions are focusing on bursty networking models able to handle dynamic segments instead of continuous data, or combinations of burst and circuit switching.

Optical Burst Switching

In recent years, much research has focused on OBS. It is a technology which is suited to bursty traffic (and can be viewed as application-centric) but can be realized in a less complex way than OPS (see below). Indeed some commercial equipment is now available [27].

OBS [28] is based on the separation of data and control plane and is seen as a cut-through technology, i.e., data transfers directly though the network without being stored at intermediate nodes. It has become of great interest as it enables many data streams to share a wavelength channel, i.e., it offers subwavelength granularity. This sharing of a wavelength channel will become increasingly important as channel bit rates increase to 40 Gbps and beyond. Figure 15.14 illustrates the operation of OBS. Prior to data burst transmission, a burst control packet (BCP) is created and sent toward the destination by an OBS ingress node (the edge router). The BCP is typically sent out-of-band over a separate signaling wavelength and processed at intermediate OBS routers. It informs each node of the impending data burst and sets up an optical path for its corresponding data burst. Data bursts remain in the optical plane end-to-end and are typically not buffered as they transit the network core. The bursts' content, protocol, bit rate, modulation

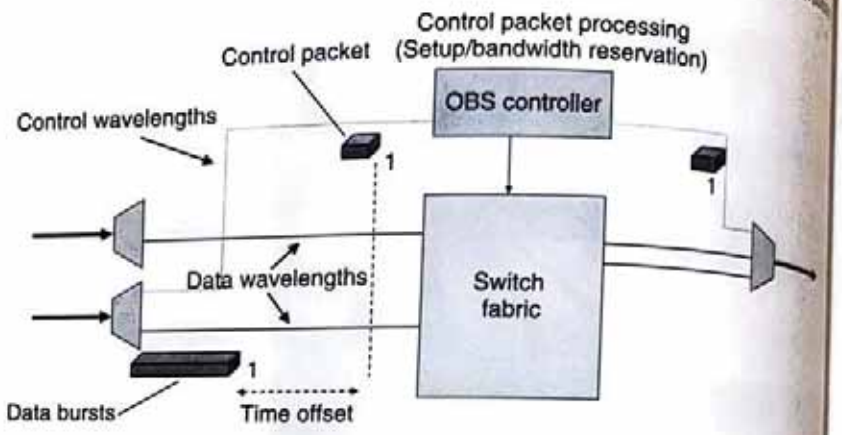


Figure 15.14 Optical burst switching concept (this figure may be seen in color on the include CD-ROM).

format, and encoding are completely transparent to the intermediate routers. The main advantages of the OBS in comparison with the other optical networking schemes are that unlike the optical wavelength switched networks, the optical bandwidth is reserved only for the duration of the burst, and that unlike the OPS network it can be bufferless; but it also needs a switch reconfiguration speed in the order of microseconds; switch speed is important as it has an impact on the size of burst that can be handled. Figure 15.15 shows a schematic of an OBS network scenario. At the edge of the network, Figure 15.15, data are aggregated into classified bursts, giving some, e.g., short burst lengths for QoS. A GMPLS control plane supports the data plane and switch reconfiguration; the node switches ideally have reconfiguration speeds in the microsecond regime.

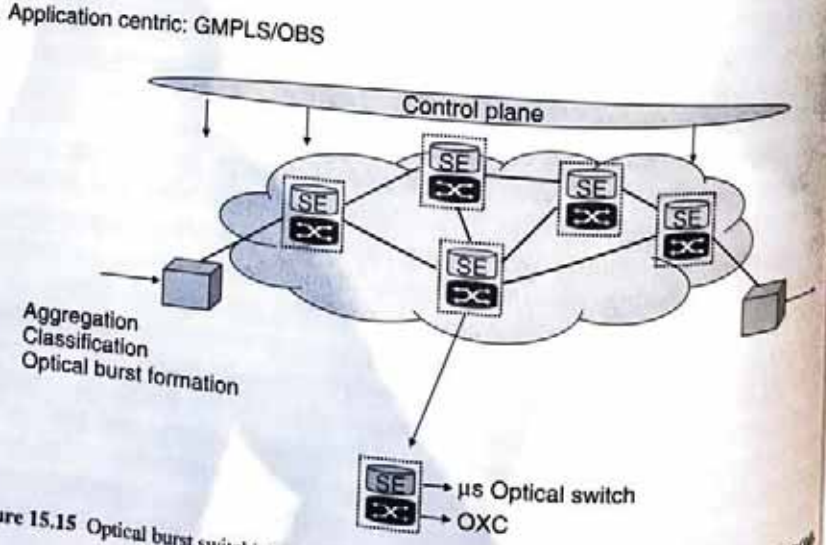


Figure 15.15 Optical burst switching network (this figure may be seen in color on the included CD-ROM)

Due to these implementation issues (easier processing, bufferless/reduced buffer operation, and relatively slow switching requirements), OBS is seen as a practical solution and is considered as the next evolution step in future optical networking. The feasibility of OBS technology can be identified by a number of reported results from field trials and test beds. The most complete demonstrator was set up in Japan under the "OBS network" project and accommodated six nodes with MEMS-based switches, achieving switching times of 1 ms for bursts with a minimum size of 100 ms [29].

OBS is currently being considered for deployment in Metro/WAN networks. Currently, the only way to build such networks with more than 10 GBPS of bandwidth requires the use of DWDM technology to enable point-to-point circuits to be established for every required path across the network; clearly, this can be both expensive, inefficient, and difficult to manage. OBS, with its multiplexing capabilities, offers a new approach by enabling communications with multiple destinations across a network. This means no circuits need be preprovisioned, and high-speed transponders need not be dedicated for every single communication path. This frees up capital, simplifies network design, and enables the creation of packet metro aggregation networks where bandwidth shifts in real-time to where it is needed in the network. On the negative side, the use OBS may have a detrimental effect on network utilization, a topic still under study. The lack of buffering at the OBS nodes and the challenges in efficient scheduling of variable size bursts, which lead to poor utilization, make this a serious aspect of network design.

OBS has been also identified as a compatible solution for the physical layer infrastructure in grid computing applications, with possible realization on NRENs [30].

Optical Packet Switching

Figure 15.12 shows that OPS is seen as the highest granularity of the proposed schemes; it is also the most difficult and complex to realize; hence, for many years it has appeared on the edges of any network evolution roadmap (such as in Figure 15.10). OPS is a store and forward technology so optical packets are buffered at network nodes; hence, buffering is a key issue and challenge in OPS networks.

Figure 15.16 illustrates the operation of an optical packet switch. The top of the diagram shows the form of an optical packet, which comprises a number of IP packets aggregated to form a payload with a serial header, header, and payload separated by a time duration sufficient for the optical switch to reconfigure; this time is typically in the order of nanoseconds. At the switch node, the header is decoded, by means of detection followed by electronic processing, and used to set the switch fabric. The switch operates on a store and forward mode, so if contention exists where more than one input seeks the same output, buffering is necessary. This is problematic for optics as, to date, it is only possible to delay a packet (rather than store) usually by means of a length of fiber; in a discussion below, a means of buffering using advanced technology is outlined. At the switch output, a

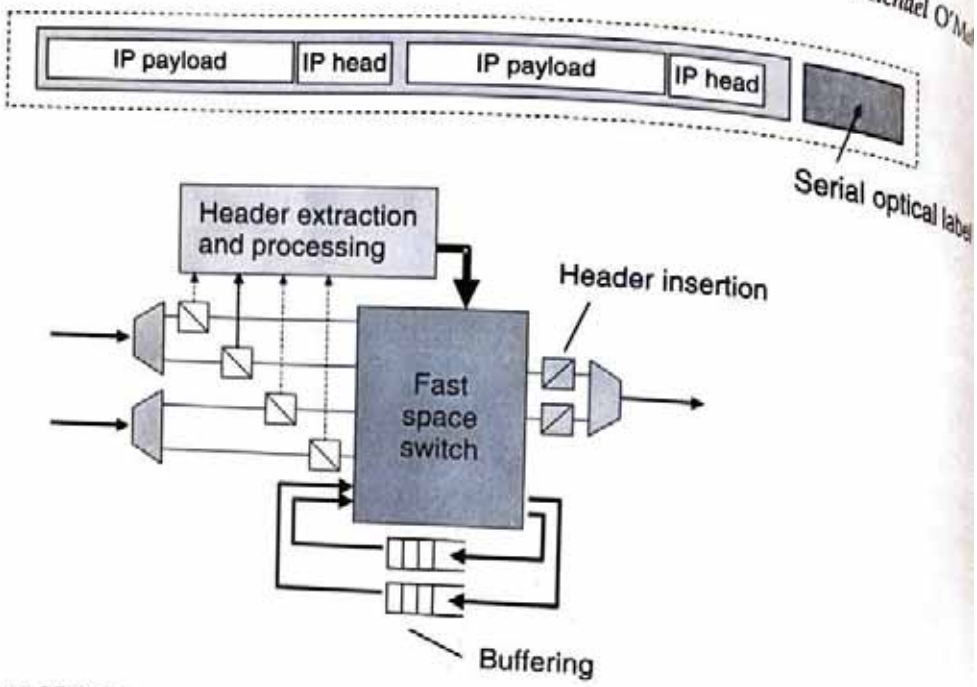


Figure 15.16 Optical packet switching (this figure may be seen in color on the included CD-RO

new optical header is inserted (and the old deleted) before the packet exits to t network. OPS requires the use of more demanding subsystems (than OBS) w intrinsic intelligence to realize adequate packet processing and routing on the The main challenges in OPS are the implementation of the optical header proc sing mechanism, the development of an intelligent switch controller, the reali tion of ultrafast switching in nanoseconds timescale, and the exploitation buffering mechanisms to reduce packet blocking.

A common realization of a switch node is illustrated in Figure 15.17, whic shows a possible switch fabric but does not include other necessary subsyste such as buffering or routing table implementations; buffering is commonly re lized using fiber delay lines or their equivalent [31]. Each incoming fiber (a total *m* fibers) supports a number of wavelengths (n), which are demultiplexed at th switch input. The stream of optical packets (on each wavelength) is first sample by a splitter, to allow the optical header to be decoded, converted to an electrica signal for electronic processing, then used as a control signal to the tunable wavelength converter (TWC). The TWC in conjunction with the arrayed wave guide (AWG) forms the equivalent of a space switch functionality as follows The path through the AWG is determined by the wavelength at its input port. Th header decoder and processing sets the electrical signal to the TWC to change the wavelength of the optical packet payload to that required to ensure it exit the AWG at the port specified by the packet header. The combination of tunable laser and AWG enables switch times in the order of nanoseconds to be achieved and this enables relatively short packets to be formed in an efficient manner.

In contrast to OBS, complete OPS demonstrators are mainly restricted to the development of fully functional but small switching elements that simply show the

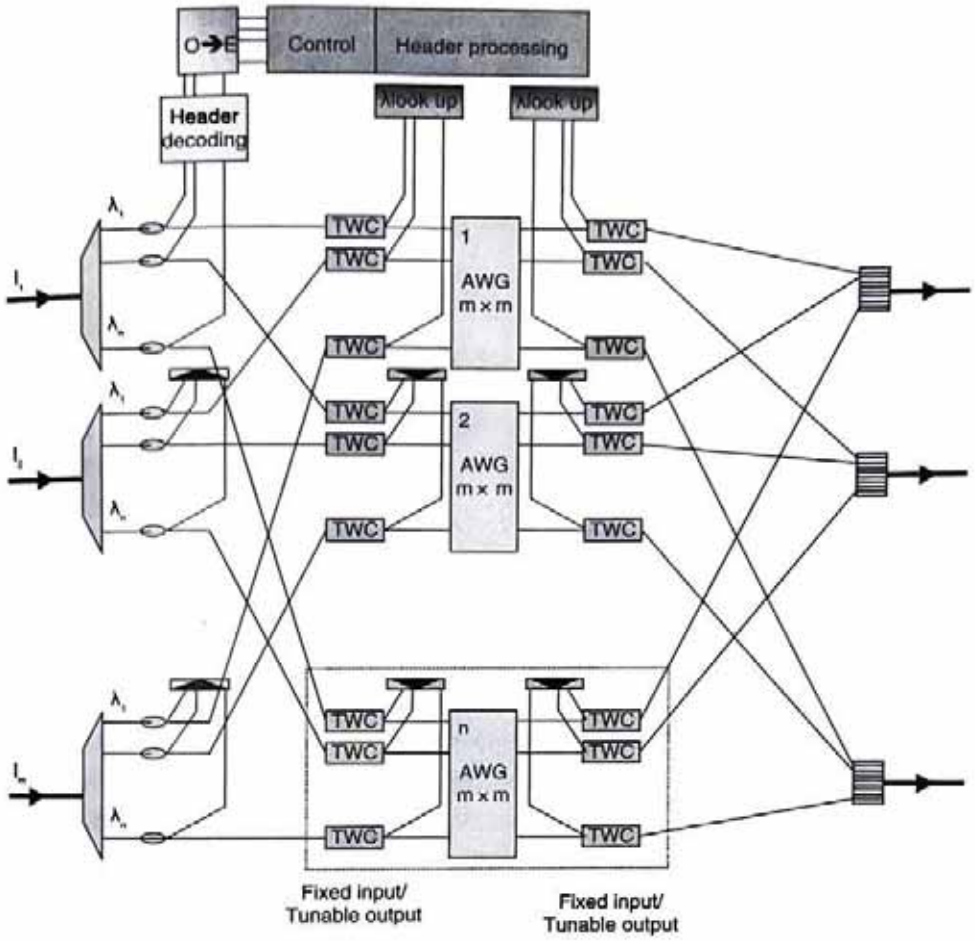


Figure 15.17 Optical packet switch node (this figure may be seen in color on the included CD-ROM).

feasibility of fast packet switching on the physical layer with some extensions to the link layer. In the OPSNET project [32], dynamic switching of asynchronous optical packets at 40 Gbps has been demonstrated in a fully controllable setup able to identify and process the header and route the payload accordingly. In Ref. [32], contention resolution on the wavelength level is also considered on a 10 Gbps packet switching node. A more feasible approach toward the implementation of OPS considers the use of synchronously (slotted) transmitted packets with fixed lengths, but in this case the hardware overhead is on the implementation of the packet synchronizer at the input. However, slotted solutions are attractive for other applications like computer interconnects. More recent approaches take the view of a multiwavelength packet [33], where the payload comprises a number of wavelengths, e.g., 16 wavelengths each supporting 10 Gbps enable a 160-Gbps OPS transmission link to be demonstrated. This approach takes OPS a long way forward in terms of practical realization.

Despite their feasibility limitations, OPS demonstrators assisted the development of numerous ultrafast switching and processing techniques regarding

wavelength conversion, header encoding/decoding and processing, label sw
ping, fast clock extraction, regeneration, and optical contention resolution. Ac
tionally, various switch architectural designs and control protocols have be
proposed, which in combination with the significant technological advances o
the last years, indicate the possible deployment of OPS in future. However,
future relies on advances in photonic integration that will enable cost-effec
subsystems to be constructed, including the all important buffer.

Optical Code Division Multiple Access

OCDMA has been studied as an alternative networking solution able to incre
passively the number of users per wavelength. The advantages of OCDMA ha
been evident for some time through its successful use in wireless networks. I
optical networking, its potential for enhanced security, decentralized control, a
flexibility in bandwidth granularity provides interesting possibilities to solve th
well-known issues in the development of future networks. Additionally, the fea
bility of OCDMA has been assisted by newly developed components able to provi
simple ways of coding and decoding signals in a passive manner, a particular
attractive and cost-effective feature. Figure 15.18 shows a coder/decoder based on
fiber Bragg grating (FBG). Sections of the grating are arranged so that an input pul
(at the Bragg wavelength) is reflected with an appropriate phase so that the outp
from the input pulse is represented by a series of reflections of differing phases. Th
the input pulse is coded into a chip sequence which is dependent on the gratin
design. Gratings have been designed to give up to 511 chips.

The operation of OCDMA relies on each user having a unique code, but sharin
the same bandwidth. User signals are multiplexed at the transmitter on to a comm
wavelength: at the receiver each incoming signal is matched against locally stor

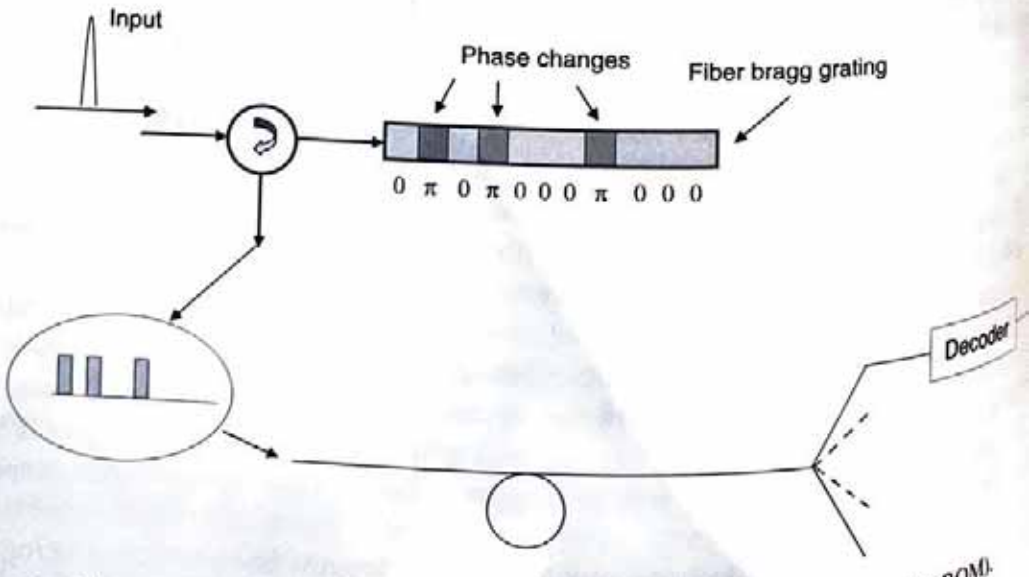


Figure 15.18 OCDMA (this figure may be seen in color on the included CD-ROM).

Dr. Dan Hi

copies of the codes through the process of autocorrelation. The major problem in OCDMA is the interference between user signals, called multiple access interference (MAI). The discrimination achieved between user signals is determined primarily by the set of codes chosen. Essentially good discrimination means long codes, but recall from above that long codes mean a high channel chip rate and hence high channel bandwidths are required, e.g., if each user were to require 1 GBPS and the code length was 50 chips, then the channel bandwidth would be required to be 50 GBPS; codes must be chosen so that the cross-correlation function between codes is low when compared with the peak autocorrelation value. The combination of OCDMA with CWDM technology in access environments is an interesting example of how the technology can help boost the number of users sharing a PON, as recently demonstrated in Ref. [34]. The paper reported the field trial of a 3-WDM × 10 – OCDMA × 10.71 Gbps system over a 111-km field trial. Key aspects of the approach were (a) the use of a multiport encoder/decoder in the central office which can give multiple optical codes in multiple wavelength bands; this device gives good correlation properties to suppress MAI and beat noise; (b) the use of an super-structural fibre Bragg grating (SSFBG) (or tunable transversal filter) at the optical network unit (ONU). The use of the DPSK-OCDMA with balanced detection is seen as a key enabler over conventional on/off keying OCDMA with superior noise performance.

In recent times other interesting possibilities for OCDMA have been reported [35]. For example, the FBGs illustrated in Figure 15.18 have been modified to incorporate the inclusion of tungsten wires at intervals along the grating; by passing current though these wires, the local refractive index can be modified and a reconfigurable coder/decoder can be achieved. Thus a user assigned a particular code on a particular wavelength can alter his code (switching) as required. Finally, the use of optical codes under a different concept has shown the feasibility to implement an OPS node [36]. Here optical codes are used as labels in order to distinguish the different headers of the transmitted packets. After header matching, the autocorrelation peak triggers an electronic controller that shows the output port where the packet should be routed.

In summary, research on OCDMA in recent years has demonstrated that it provides an interesting extension to the usual dimensions of space, time, and wavelength deployed in optical networks. However, it faces challenges if used in areas of the network involving long distances and high bit rates, and so at present most consider the access network the most likely area for deployment. As the access network is very cost sensitive, however, severe demands are placed on the technology to allow it to be competitive with legacy solutions.

15.6 KEY SUBSYSTEMS AND TECHNOLOGIES

The roadmap of Figure 15.10, and the above discussion, broadly outlines a much argued route forward to the future; with switching or transmission aspects relevant to a future global optical network. Generally, the picture is that transmission

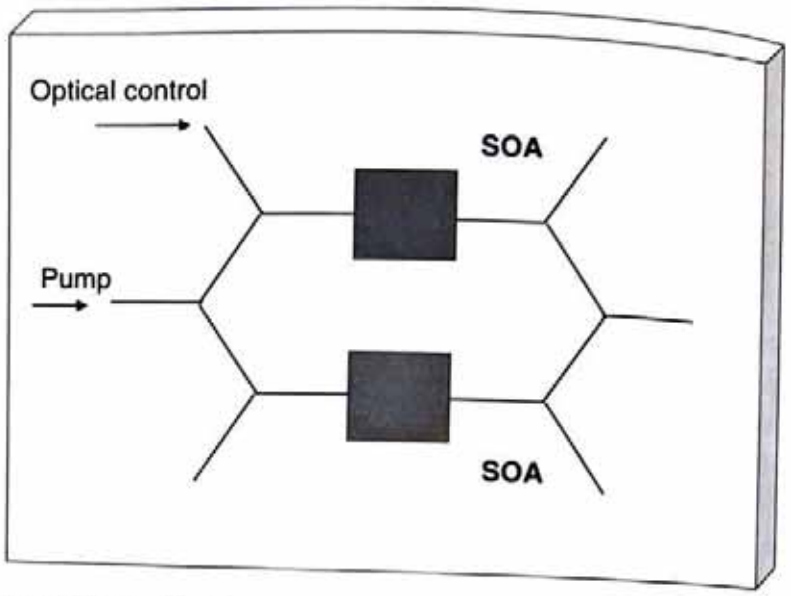
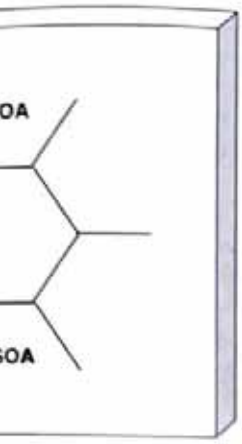


Figure 15.19 Mach Zehnder Interferometer (MZI) (this figure may be seen in color on the inclu CD-ROM).

speeds are still predicted to increase as before (e.g., to 160 Gbps probably usi DPSK), within the context of a more dynamic and granular network, supported b appropriate control plane (e.g., GMPLS with extensions). Many of the key func tional subsystems needed for this future flexible network have already bee described, with experimental demonstrations (e.g., all-optical regeneration a 40 Gbps has been widely reported [37], as has wavelength conversion, fast tunable lasers, tunable dispersion compensators, MEMs-based switch fabrics, etc.), others are still gleams in the eye of the network designer with, as yet, no sure route for realization, e.g., optical memory and multiwavelength optical regeneration.

However, what has become well recognized is that the future of realizing the full potential of optical networks, at affordable cost, lies in the ability to perform good levels of photonic integration. This has not been an easy route to date, but some interesting examples of future approaches are starting to appear. Generally, the integration of photonic components is not straightforward as the substrate technology for various components differs (e.g., lasers are grown on InP whereas filters benefit from silicon technology). Existing photonic integration approaches are based on hybrid integration, where individual components are laid down and interconnected on a suitable substrate (e.g., silicon) (hybrid integration offers full functionality, lower cost, and shorter development times), or on monolithic integration, where a limited set of functions, realizable on a common material, are combined. Photonic integration also faces a challenge from approaches such as Infinera [38], where large number of OEO interfaces are integrated on an IP chip allowing similar functionality in many cases at comparable costs.

Advances in the integration of Mach Zehnder Interferometers (MZI) illustrate the benefits of integration. Figure 15.19 shows the structure of such a device where a pump signal is split between the two interferometer arms and the phase difference between the arms is controlled by an optical signal operating on



gure may be seen in color on the included

(e.g., to 160 Gbps probably using
nd granular network, supported by
xtensions). Many of the key func-
ible network have already been
(e.g., all-optical regeneration at
avelength conversion, fast tunable
-based switch fabrics, etc.), others
ner with, as yet, no sure route for
elength optical regeneration.
is that the future of realizing the
cost, lies in the ability to perform
ot been an easy route to date, but
are starting to appear. Generally,
straightforward as the substrate
, lasers are grown on InP whereas
g photonic integration approaches
ual components are laid down and
con) (hybrid integration offers full
ent times), or on monolithic inte-
zable on a common material, are
hallenge from approaches such as
faces are integrated on an IP chip,
comparable costs.
er Interferometers (MZI) illustrate
s the structure of such a device,
interferometer arms and the phase
an optical signal operating on a

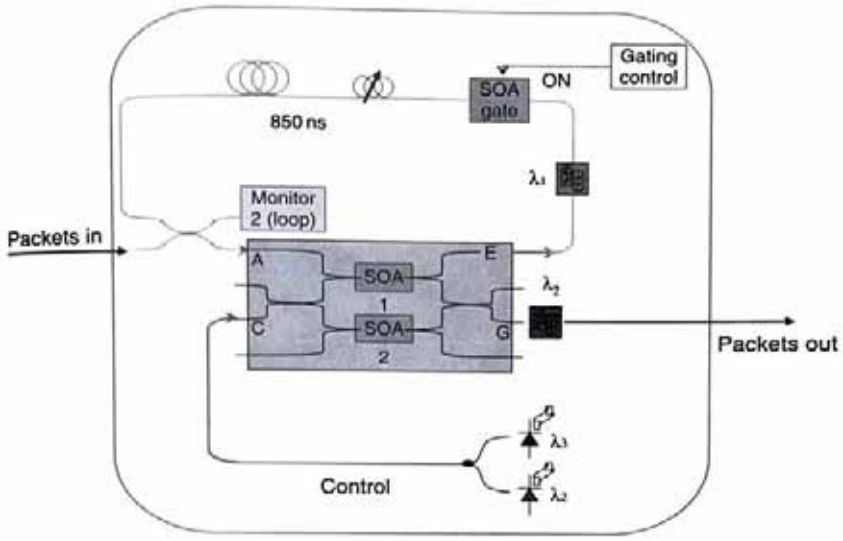


Figure 15.20 MZI buffer (this figure may be seen in color on the included CD-ROM).

nonlinear element, typically a semiconductor optical amplifier (SOA). Integrated MZIs can be used for a variety of functions needed in future systems and cannot be realized by assembling discrete components because of the need for stability in the physical dimensions of the device. Commercial devices are available for operation at 40 Gbps and can be used to realize a variety of functions such as 2R regeneration and wavelength conversion [39]. Research currently shows how, through the use of push-pull configurations where phase changes are applied in both arms, the performance can be significantly enhanced, with up to 160 Gbps operation reported [40]. This configuration has been used in a number of different signal-processing applications such as regeneration, add–drop multiplexing, and time slot interchange switching [41]. Currently, devices are becoming available where multiple MZIs are integrated on to one substrate enabling futuristic functional subsystems to be achieved such as bursty receivers [42]. Figure 15.20 shows the use of such a device to realize an optical buffer, suitable for an OPS system. The data bursts enter the MZI at port A. A control signal comprising a continuous train of return-to-zero (RZ) pulses at the same bit rate as the data is injected at input C; the wavelength of this pulse train can be altered between one of two wavelengths. The data exit the MZI at port E and circulates around the buffer loop, which as shown is set to give a round trip delay of 850 ns—a packet length in this example. Data continue to circulate around the loop until the control wavelength is switched to the other state, in which case the data exit through port G through a bandpass filter. The SOA gate in the upper loop is used (by switching off) to clear the loop for the next input of original data.

Recently, novel approaches based on silicon photonic integration (Si Photonics) have been reported by Intel [43] (see also Chapter 11 on "Si Photonics" by Cary Gunn and Thomas L. Koch). Silicon is a well-under material and the basis of low-cost electronics. It is transparent at infrared wa lengths so can be used to guide light, but cannot emit light. Intel are pioneering process whereby a laser chip mounted in a silicon external cavity forms a hyb laser, which together with silicon modulators (realized by MZI configurations) a photodetectors (with Ge doping) shows promise as a future low-cost integrat strategy; current performance is at the 1 Gbps, but 10–40 Gbps is predicted.

There are a number of key functions necessary, at a link and network level, enable efficient operation of the future optical network. Examples are:

Optical switching: To fully advance to transparent networking, OOO switches required, at microsecond (for OBS), millisecond (for OXCs), and nanosecon (for OPS) reconfiguration times. 3-D MEMS switches (ms reconfiguration) now available with port counts to 160×160 [44], and the recent interest in deployment of ROADMs means that the functionality of OOO switches is no available. Fast (ns) switch fabrics (of high dimension) are realized through combination of tunable lasers/wavelength converters together with arra waveguides; it is hard to see this situation changing. Nevertheless, low dimension ns switch modules are available [45], an example based on to internal reflection in a compound semiconductor PLC [46], which forms core of a commercially deployed OPS system (see also Chapter 17 R. Tucker for discussion on key issues for optical switching).

Optical monitoring (see also Chapter 7 on "Optical Performance Monitoring" Alan Willner): As the network performance increases, it is crucial to underst the state of the physical layer so that appropriate routing and remedial actio can be taken. For example, within a network, a variety of bit rates from 10 160 Gbps may exist, with some routes more appropriate to one bit rate th another from the point of view of optical signal-to-noise ratio (OSNR), resulti dispersion channel power, etc. Information should be available through opti monitoring to inform routing decisions at network nodes; this trend now appea in the area of cross-layer routing and is important one for research.

All-optical wavelength conversion and regeneration: All-optical wavelengt conversion and regeneration is desirable (including from a pow and footprint viewpoint as well as cost) for networks operating at speeds >40 Gbps (see also Chapter 20 in Volume A "Nonlinear Optical Circuits for Signal Processing" by Ben Eggleton and Stojan Radic). To achie conversion, the physical properties of a nonlinear element are used perform a logic function between the input signal and a pump. The mos nonlinear elements used are an SOA, electro-absorption modulator (EAM fiber, photonic crystal, and periodically poled $LiNbO_3$ (PPLN) waveguide SOA-based devices, especially quantum dot, and EAMs have the addi advantages of compactness and low-energy requirements to trigg

Dr. Dan Blumenthal

nonlinearities. Fibers have an instantaneous response to pulses but have limited nonlinearity, even in specially designed photonic crystal fibers; hence long lengths are required. PPLN requires intermediate lengths, and very fast conversion (40–160 Gbps) has been demonstrated.

Considerable research has been done in the area of single-wavelength subsystems, and e.g., advanced all-optical regenerative schemes at bit rates beyond 100 Gbps have been recently proposed in Ref. [47], as well as well-reported demonstrations at 40 Gbps [48]. Multiwavelength all-optical regeneration, if feasible, will dramatically decrease the cost of DWDM transmission links. A number of techniques are currently being considered, e.g., (a) through the mechanism of self-phase modulation, which in principle enables operation at speeds >160 Gbps, and (b) based on the inhomogeneously-broadened gain of self-assembled quantum dots in Quantum Dot SOAs [49].

Optical memory: All-optical buffering through fiber delay lines is an approach that requires complex control, and packets are delayed rather than stored. Recently, in the framework of LASOR [14], integrated delay lines have been developed as it has been shown [50] that a small number of low-depth buffers are sufficient for an OPS network. Recent research has focused on the possibility of controlling the velocity of light pulses propagating through a material, and measurements in SOA quantum wells showed controllable delay up to 1 ns; this results in a direct measurement of group velocity of <200 m/s giving a slow down factor up to 1.5×10^6 [51]. However, there is an indication that such phenomenon exhibit a specific delay-bandwidth product which means for high bit rates the achievable delay is low.

15.7 CONCLUSION

The drive toward an optical network, by which we mean a network comprising optical transmission and switching has taken about 20 years from original concept to the start of deployment of limited optical switching. The huge growth in demand for capacity from domestic users, due to the penetration of broadband service, together with the increasing needs of scientific high-end users, has made the economic deployment of advanced optical technology economic in a growing number of application areas. It is also the case that the maturity and ready availability of optical technologies have made it possible for communities to construct their own networks independent of operators, a situation that has greatly changed in the past decade.

New network architectures do not suddenly appear and real networks must evolve carefully, and this chapter discussed the current roadmap for the choice of switching technologies future networks might employ, representing research studies in current and recent years. The main changes represent the recognition that the world has changed to a data rather than voice-centric model; it was shown that between 2000 and 2006 the volume of data generated grew by a factor of 50. Thus initial network changes are to move the current SDH/SONET transmission and

switching technologies to one more amenable to data (NG-SDH) and the changes are now implemented. In conjunction with these near term changes there is a strong move toward carrier-grade Ethernet, in particular 100 G Ethernet on a single channel, and it is likely that this trend will come to dominate. Other immediate changes required are to move toward a more dynamic network where circuits can be established and torn down in minutes rather than hours or days, and such architectures are represented in the standardized ASON architecture, where requests from users activate a control plane which initiates automatic node switching. These two changes represent desired changes in the current circuit-switched network, but the accelerating changes in demand (capacity, quality) require more significant changes. Discussed were:

(1) the move to all IP-based network where the core network comprises an IP router together with an OXC. This is an approach currently being adopted by BT UK.
(2) the move to an application-centric network based on OBS or OPS. Here appropriate network interfaces can assemble bursts representing required QoS at the network edge. The core of the network required optical switching fabrics, ideally in the microsecond switching regime.

From a transmission viewpoint current networks operate mainly at 10 GBPS. While 40-GBPS technology is now available for deployment. As the overall network load increases, studies show that the optimum line rate also increases, so line rates of >100 GBPS must not be discounted.

Finally, it was noted that there have been in recent times significant advances in subsystems and components and in device integration—seen as the key to future optical networking. This integration enables the realization of many of the key functionalities for optical networking to be realized in a manner appropriate to its deployment; examples are all-optical wavelength conversion and optical regeneration.

REFERENCES

[1] H. Shinohara, "Overview of Japanese FTTH Market and NTT Strategy for Entering Full FTTH Era," in Proc. *ECOC 2006*, paper Th1.1.1, Cannes, September 2006.
[2] R. Smith, Grid Computing: A Brief Technology Analysis, CTO Network Library, 2005.
[3] http://www.jb.man.ac.uk/news/evlbi/
[4] Bill St Arnaud, "Future Internet Issues," *OECD ICCP Workshop*, March 2006, Paris, France.
[5] C. Cavazzoni, D. Colle, A. Di Giglio et al., "Evolution of Optical Transport Networks in Europe: The NOBEL Project Vision," in Proc. *Broadband Europe Conf.*, paper T04A.01, Bordeaux, December 2005 (www.bbeurope.org).
[6] N. Larkin, "ASON & GMPLS: the battle for the optical control plane"; http://www.dataconnection.com/network/download/whitepapers/asongmpls.pdf.
[7] Stavdas, S. Sygletos, M. O. Mahony et al., "IST-DAVID: Concept presentation and physical layer modelling of the metropolitan area network," *J. Lightw. Technol.*, 21(2), 372–384, February 20

Dr. Dan

[8] Dimitrios Klonidis, Christina (T.) Politi, Reza Nejabati, Mike J. O. Mahony, and Dimitra Simeonidou; OPSnet: "Design and demonstration of an asynchronous high speed optical packet switch," *IEEE J. Lightw. Technol.*, 23(10), 2914–2925, October 2005.

[9] R. Nejabati, D. Klonidis, G. Zervas et al., "Demonstration of a Complete and Fully Functional End-to-End Asynchronous Optical Packet Switched Network," in Proc. *OFC 2006*, paper OWP5, Anaheim, USA, March 2006.

[10] J. Ryan, "An Outlook for Optical Executatives," in Proc. *OSA Executative Program*, Anaheim, USA, March 2006.www.osa.org/partner/network/program/Ryan%20John.ppt

[11] B. Fabianek, "Optical Networking Testbeds in Europe," in Proc. *OFC 2006*, paper OWU1, Anaheim, USA, March 2006.

[12] Kenichi-Kitayama, Tetsuya Miki, Toshio Morioka et al., "Photonic network r&d activities in Japan-current activities and future prospects," *J. Lightw. Technol.*, 23(10), 3404–3418, October 2005.

[13] D. T. Neilson, D. Stiliadis, and P. Bernasconi, "Ultra-High Capacity Optical Ip Routers for the Networks of Tomorrow: Iris Project," in Proc. *ECOC 2005*, paper We 1.1.4, Glasgow, UK, September 2005.

[14] D. J. Blumenthal, "LASOR (Label Switched Optical Router): Architecture and Underlying Integration Technologies," in Proc. *ECOC 2005*, paper No XX, Glasgow, UK, September 2005.

[15] http://www.nlr.net/ess/index.html.

[16] P. Freeman, "GENI: Global Environment for Networking Innovations," in Proc. *The Future of the Internet*, OECD, Paris, France, March 2006 (http://www.oecd.org/dataoecd/43/63/36274169.pdf.)

[17] http://www.glif.is/

[18] J. Zhou, R. Cadeddu, E. Casaccia et al., "Crosstalk in multiwavelength optical cross-connect networks," *J. Lightw. Technol.*, 14, 1423–1435, June 1996.

[19] R. E. Wagner et al., "Monet: Multiwavelength optical networking," *J. Lightw. Technol.*, 14, 1349–1355, June 1996.

[20] T. Freeman, "DWDM platform set for 40 G deployment," *Fibre Systems*, 3(1), 23, January/February 2006.

[21] L. Zong, P. Ji, T. Wang et al., "Study on Wavelength Cross-Connect Realised with Wavelength Selective Switches," in Proc. *OFC 2006*, paper NThC3, Anaheim, USA, March 2006.

[22] E. Hernandez-Valencia and H. Menendez, "Exploiting Carrier Ethernet to Deliver Profitable New Services, in Proc. *OFC 2006*, paper NTUD1, Anaheim, USA, March 2006.

[23] D. Simeonidou, R. Nejabati, B. St. Arnaud et al., "Optical Network Infrastructure for Grid," http://www.ggf.org/gf/docs/?final

[24] K. Kitayama, X. Wang, and N. Wada, "A Solution Path to Gigabit-Symmetric FTTH: OCDMA over WDM PON," in Proc. *Broadband Europe Conf.*, paper T01A.02, Bordeaux, December 2005. (www.bbeurope.org)

[25] R. C. Mendez, P. Toliver, S. Galli et al., "Network applications of cascaded code translation for wdm-comaptible spectrally phase encoded optical cdma," *J. Lightw. Technol.*, 23(10), 3129–3231, October 2006.

[26] Esther Le Rouzic and Stephanie Gosselin "160 Gb/s Optical networking: A prospective techno-economic analysis," *IEE J. Lightw. Technol.*, 23(10), 3024–3033, October 2005.

[27] http://www.matissenetworkks.com.

[28] C. Qiao and M. Yoo, "Optical burst switching – A new paradigm for an optical internet," *J. High Speed Netw.*, Special Issue on Optical Networks, 8(1), 69–84, 1999.

[29] A. Sahara, R. Kasahara, E. Yamazaki et al., "The Demonstration of Congestion Controlled Optical Burst Switching Network Utilizing Two-Way Signaling Field Trial in Jgn Ii Testbed," in Proc. *OFC 2005*, paper OFA7, Anaheim, CA, USA, March 2005.

[30] Dimitra Simeonidou, Reza Nejabati, Georgios Zervas et al., "Dynamic optical network architectures and technologies for existing and emerging grid services," *J. Lightw. Technol. Member*, 23(10), 3347–3357, October 2005.

[31] J. Gripp, D. Stiliadis, J. E. Simsarian et al., "IRIS optical packet router, [Invited]," *J. Opt. Netw.*, 5, 589–597, 2006.

[32] Dimitrios Klonidis, Christina (T.) Politi, Reza Nejabati, Mike J. O. Mahony et al., "OPS Design and demonstration of an asynchronous high speed optical packet switch," *IEEE J. Lightw. Technol.*, 23(10), 2914–2925, October 2005.
[33] H. Furukawa, N. Wada, H. Harai et al., "Field Trial of IP over 160 Gbit/s Colored = Optical-Packet Switching Network with transient Response Suppressed EDFAand 320 Gbit/s toughput Optical Packet Switch Demonstrator," in Proc. *OFC 2007*, PDL, Anaheim, USA, March 2007.
[34] K. Kitayama, X. Wang, and N. Wada, "A Solution Path to Gigabit-Symmetric FTTH: OCDMA over WDM PON," in Proc. *Broadband Europe Conf.*, paper T01A.02, Bordeaux, December 20(www.bbeurope.org)
[35] C. Tian, Z. Zhang, M. Ibsen et al., "Demonstration of a 16 channel code-reconfigurable OCDMA DWDM system," in Proc. *OFC 2007*, paper OMO2, Anaheim, USA, March 2007.
[36] X. Wang and N. Wada, "Demonstration of OCDMA Traffic over Optical Packet Switched Networks with PLC and SSFBG En/decoders for Time Domain OC Processing," in Proc. *ECOC 2005*, paper PDL We1.4.5, Glasgow, UK, September 2005.
[37] K. Stubkjaer, "Semiconductor optical amplifier-based all-optical gates for high speed optical processing, *IEEE J. Sel. Top. Quantum Electron.*, 6(6), 1428, November 2000.
[38] R. Nagarajan, C. Joyner, and R. Schneider, "Large-scale photonic integrated circuits," *IEEE J. Sel. Top. Quantum Electron.*, 11(1), January/February 2005.
[39] G. Maxwell, R. McDougall, R. Harmon et al., "WDM – enabled, 40 GB/s Hybrid Integrated all optical Regenerator," in Proc. *ECOC 2005*, paper PDL 4.2.2, Glasgow, UK, September 2005.
[40] S. Nakamura, Y. Ueno, and K. Tajima, "168 Gbit/s all-optical wavelength conversion with symmetric-Mach-Zehnder-type switch," *IEEE Photon. Technol. Lett.*, 13, 1091, 2001.
[41] S. Pau, Y. Jianjun, K. Kojima et al., "160-Gb/s all-optical MEMS time-slot switch for OTDM and WDM applications," *IEEE Photon. Technol. Lett.*, 14(10), 1460–1462, October 2002.
[42] Dimitrios Petrantonakis, George T. Kanellos, Panagiotis Zakynthinos et al., "A 40 Gb/s 3R Burst Mode Receiver with 4 Integrated MZI Switches," in Proc. *OFC 2006*, paper PDP 25, Anaheim, USA, March 2006.
[43] P. Koonath, T. Indukuri, and B. Jalali, "Monolithic 3-D silicon photonics," *J. Lightw. Technol.*, 24(4), 1796–1804, April 2006.
[44] http://www.glimmerglass.com
[45] R. Varrazza, I. B. Djordjevic, and Y. Siyuan, "Active vertical-coupler-based optical crossbar switch matrix for optical packet-switching applications," *J. Lightw. Technol.*, 22(9), 2034–2042, 2004.
[46] http://www.yokogawa.com
[47] Yong Liu, Eduward Tangdiongga, Zhonggui Li et al., Eindhoven Univ. of Technology, The Netherlands, Aston Univ., UK. "Error-Free 320 Gb/s SOA-Based Wavelength Conversion Using Optical Filtering," in Proc. *OFC 2006*, paper PDP 28, Anaheim, USA, March 2006.
[48] O. Lecrec et al., "Optical regeneration at 40 Gb/s and beyond," *J. Lightw. Technol.*, 21(11), 2779–2790, 2003.
[49] T. Akiyama et al., "Application of Spectral-Hole Burning in the Inhomogeneously Broadened Gain of Self-Assembled Quantum Dots to a Multi-Channel Nonlinear Optical Device," *Photon. Technol. Lett.*, 12(10), 1301–1303, 2000.
[50] N. Beheshti, Y. Ganjali, R. Rajaduray et al., "Buffer Sizing in All-Optical Packet Switches," in Proc. *OFC 2006*, paper OThF8, Anaheim, USA, March 2006.
[51] C. Chang Hasnain, "Controlling Light Speed in Semiconductors," in Proc. *ECOC 2005*, paper Mo3.1.3, Glasgow, UK, September 2005.