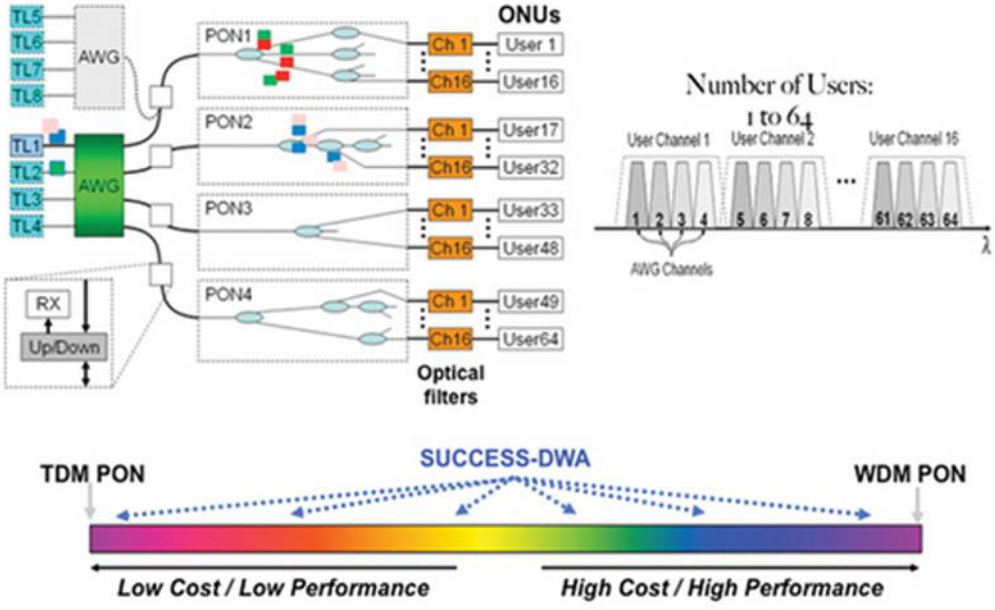
Broadband Optical Access Networks

Leonid G. Kazovsky • Ning Cheng • Wei-Tao Shaw
David Gutierrez • Shing-Wa Wong



WILEY

BROADBAND OPTICAL ACCESS NETWORKS

BROADBAND OPTICAL ACCESS NETWORKS

LEONID G. KAZOVSKY
NING CHENG
WEI-TAO SHAW
DAVID GUTIERREZ
SHING-WA WONG

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data Is Available

Kazovsky, Leonid G.

Broadband optical access networks / Leonid G. Kazovsky, Ning Cheng, Wei-Tao Shaw, David Gutierrez, Shing-Wa Wong.

Includes index.

ISBN 978-0-470-18235-2

Printed in Singapore

oBook ISBN: 978 0470 910931
ePDF ISBN: 978 0470 910924
ePub ISBN: 978 0470 922675

10 9 8 7 6 5 4 3 2 1

CONTENTS

FOREWORD

Broadband optical access networks are crucial to the future development of the Internet. The continuing evolution of high-capacity, low-latency optical access networks will provide users with real-time high-bandwidth access to the Web essential for such emerging trends as immersive video communications and ubiquitous cloud computing. These ultrahigh-speed access networks must be built under challenging economic and environmental imperatives to be "*faster, cheaper, and greener*." This book presents in a clear and illustrative format the technical and scientific concepts that are needed to accomplish the design of new broadband access networks upon which users will surf the wave of the twenty-first-century Internet.

The book is coauthored by Professor Leonid Kazovsky and his graduate students. Professor Kazovsky is a recognized leader and authority in the field and has a long and distinguished track record for making highly timely and significant research contributions within the general area of optical communication systems and optical networks. He has contributed over the last 40 years in the areas of wavelength-division-multiplexed (WDM) and coherent transmission systems for the core network as well as transmission systems and network architectures and technologies at the metro and access levels. This book builds on Professor Kazovsky's research conducted at Bellcore (where he worked in the 1980s), at Stanford University (where he has worked since 1990), and at numerous European research organizations during sabbaticals in the UK, the Netherlands, Italy, Denmark, and (most recently) Sweden. This rich set of influences gives the book and its readers the benefits of broad exposure to diverse research ideas and approaches.

Professor Kazovsky heads the Photonics and Networking Research Laboratory at Stanford University. He and his team of researchers are focusing on broadband optical access networks. They bring their ongoing research results to this unique

book, bridging fundamentals of optical communication and networking system design with technology issues and current standards. Once that foundation is laid, the book delves into current high-capacity research issues, including evolution to WDM optical access, converged hybrid optical/wireless access networks, and implementation issues of broadband optical access. Research ideas generated by Professor Kazovsky's research group have been widely adopted worldwide, including in framework projects of the European Union.

We strongly recommend this book, as it offers timely, accurate, authoritative, and innovative information regarding broadband optical access network design and implementation. We're confident that you will enjoy reading the book and learn much while doing so.

DANIEL KILPER and PETER VETTER
Alcatel-Lucent Bell Labs, Murray Hill, New Jersey
JAMES F. KELLY
Google, Mountain View, California
ALAN WILLNER
USC Viterbi School of Engineering, Los Angeles, California
BISWANATH MUKHERJEE
University of California Davis, Davis, California
ANDERS BERNTSON, GUNNAR JACOBSEN, and MIKHAIL POPOV
Acreo, Stockholm, Sweden

PREFACE

The roots of this book were planted about a decade ago. At that time, I became increasingly convinced that wide-area and metropolitan-area networks, where much of my group's research has been centered at that time, were in good shape. Although research in these fields was (and still is) needed, that's not where the networking bottleneck seemed to be. Rather, the bottleneck was (and still is in many places) in the access networks, which choked users' access to information and services. It was clear to me that the long-term solution to that problem has to involve optical fiber access networks.

That conviction led me to switch the focus of my group's research to optical access networks. In turn, that decision led to a decade of exciting exceptionally interesting research into the many challenges facing modern access networks. These challenges include rapidly increasing demands for larger bandwidth and better quality of service, graceful evolution to more powerful solutions without complete rebuilding of existing infrastructure, enhancing network range and number of users, improving access networks' resilience, simplifying network architecture, finding better control strategies, and solving the problem of fiber/wireless integration. All these problems would have to be solved while maintaining the economic viability of access networks so that operators would be prepared to make the necessary (and huge) investment in fiber and other infrastructure.

Finding solutions for the foregoing problems occupied most of my research group's time and attention for much of the past decade. In the beginning of that decade (and for a long time after that), my group, the Photonics and Networking Research Laboratory (PNRL) at Stanford University, was one of very few (or perhaps even the only) university research group working on fiber access, as many other optical researchers tended to discount optical access issues as trivial. Although that made funding for our

research difficult to find, that position allowed us to make many pioneering contributions widely used and cited today. Later, many other university and industrial research groups entered the field, and several large-scale research efforts were organized, most notably in Europe, where serious research into both passive optical networks (PONs) and active optical networks (AONs) has been conducted over the last several years. Notable European efforts in broadband fiber access include ICT ALPHA (architectures for flexible photonic home and access networks, focused on AON, PON, and technoeconomics), ICT OASE (optical access seamless evolution, focused on PON, technoeconomics, and business models) and ICT SARDANA (focused on PON and optical metropolitan networks). These efforts resulted in extremely fast progress in the field. It was gratifying to see many PNRL research results adopted, used, and developed further by these (and other) efforts, especially in SARDANA.

Many of my colleagues working on optical access research encouraged me over the past few years to integrate results of the PNRL research on optical access networks into a single volume and publish it to ensure the broadest possible dissemination of our results. They feel that our results, when published in a single volume rather than the current combination of conference and journal articles, will further stimulate new research, plant new ideas, and lead to exciting new developments.

For a long while, I was reluctant to do so. The field of broadband fiber access networks is exceptionally broad; in addition, it is still very young and is developing and changing very fast. Thus, writing a comprehensive book on this subject is (nearly) impossible. Eventually, though, a stream of inquiries for additional information about our research convinced me to change my mind, and my research students and myself began the time-consuming process of writing our book.

Our goal was fairly modest: to summarize in one place the research results produced by the PNRL over the past decade or so. The reader should keep this goal in mind. We make no attempt to cover the entire field, just to provide a summary of our research. Even that goal proved to be difficult to achieve, as we are continuing our research as new technologies emerge, so our understanding of the field continues to evolve with time. However, we trust that the reader will consider this book a useful addition to his or her knowledge base of optical access networks.

LEONID KAZOVSKY

Stanford University
Stanford, California

ACKNOWLEDGMENTS

This book is based on research results obtained by our research group, the Photonics and Networking Research Laboratory at Stanford University. Our research on broadband fiber access networks, conducted over a decade or so, required a consistent effort by a large group of exceptionally talented graduate students, postdocs, and visitors. Some of these contributors are co-authors of the book, while others are working in other organizations and on other projects and so were too busy to help with the book-writing process. We are thankful to all of them, however.

Our research on broadband fiber access networks required a sizable team and a substantial amount of experimental, theoretical, and simulation efforts. This would be impossible without the generous and long-term support of our sponsors. We are grateful to our sponsors, who trusted us with the necessary resources. Our main sponsors in that area were, or are, the National Science Foundation under grants 0520291 and 0627085, KDDI Laboratories, Motorola, the Stanford Networking Research Center (no longer in existence), ST Microelectronics, ANDevices, Huawei, Deutsche Telecom, and Alcatel-Lucent Bell Laboratories.

We also thank the many research visitors to our group (mainly postdocs or visiting professors), who helped in a variety of ways, ranging from making research contributions to our book, to providing suggestions and comments on its contents, to taking part in one or more of our broadband access research projects. In particular, we are grateful to Dr. Kyeong Soo (Joseph) Kim of Swansea University; Professor Chunming Qiao of SUNY Buffalo; Dr. Luca Valcarenghi of Scuola Superiore Sant'Anna , Italy; Professor David Larrabeiti of Universidad Carlos III de Madrid, Madrid, Spain; and Dr. Divanilson Campelo of University of Brasilia, Brazil. Many others helped as well; unfortunately, a comprehensive list would be too long to include here.

We are grateful to the challenging, exciting research environment at Stanford University, where the lead author of this book has had the pleasure of working for the past two decades. Without that environment, this book would never have materialized.

Last but not least, we would like to thank our many colleagues all over the world for stimulating discussions, for their friendship, and for their help. We are particularly grateful to Prof. Vincent Chan, MIT; Prof. Alan Willner, USC; Drs. James Kelly and Cedric Lam of Google, Inc.; Prof. Andrea Fumagali, University of Texas; Profs. Ben Yoo and Biswanath Mukherjee, University of California, Davis; Profs. Djan Khoe and Dr. Harm of the Technical University of Eindoven, the Netherlands; Prof. Giancarlo Prati of the Scuola Superiore St. Anna, Pisa, Italy; Prof. Palle Jeppesen of the Danish Technical University, Copenhagen, Denmark; Drs. Gunnar Jacobsen, Mikhail Popov, and Claus Larsen of Acreo, Stockholm, Sweden; Dr. Shu Yamamoto of KDDI, Japan; and Dr. Frank Effenburger of Huawei.

Stanford University
Stanford, California

Leonid Kazovsky
Ning Cheng
Wei-Tao Shaw
David Gutierrez
Shing-Wa Wong

CHAPTER 1

BROADBAND ACCESS TECHNOLOGIES: AN OVERVIEW

In past decades we witnessed the rapid development of global communication infrastructure and the explosive growth of the Internet, accompanied by ever-increasing user bandwidth demands and emerging multimedia applications. These dramatic changes in technologies and market demands, combined with government deregulation and fierce competition among data, telcom, and CATV operators, have scrambled the conventional communication services and created new social and economic challenges and opportunities in the new millennium. To meet those challenges and competitions, current service providers are striving to build new multimedia networks. The most challenging part of current Internet development is the access network. As an integrated part of global communication infrastructure, broadband access networks connect millions of users to the Internet, providing various services, including integrated voice, data, and video. As bandwidth demands for multimedia applications increase continuously, users require broadband and flexible access with higher bandwidth and lower cost. A variety of broadband access technologies are emerging to meet those challenging demands. While broadband communication over power lines and satellites is being developed to catch the market share, DSL (digital subscriber line) and cable modem continue to evolve, allowing telecom and CATV companies to provide high-speed access over copper wires. In the meantime, FTTx and wireless networks have become a very promising access technologies. The convergence of optical and wireless technologies could be the best solution for broadband and mobile access service in the future. As new technology continues to be developed, the future access technology will be more flexible, faster, and cheaper. In this chapter

Broadband Optical Access Networks, First Edition. Leonid G. Kazovsky, Ning Cheng, Wei-Tao Shaw, David Gutierrez, and Shing-Wa Wong.

we discuss current access network scenarios and review current and emerging broad access technologies, including DSL, cable modem, optical, and wireless solutions.

1.1 COMMUNICATION NETWORKS

Since the development of telegraph and telephone networks in the nineteenth century, communication networks have come a long way and evolved into a global infrastructure. More than ever before, communications and information technologies pervade every aspect of our lives: our homes, our workplaces, our schools, and even our bodies. As part of the fundamental infrastructure of our global village, communication networks has enabled many other developments—social, economic, cultural, and political—and has changed significantly how people live, work, and interact.

Today's global communication network is an extremely complicated system and covers a very large geographic area, all over the world and even in outer space. Such a complicated system is built and managed within a hierarchical structure, consisting of local area, access area, metropolitan area, and wide area networks (as shown in Figure 1.1). All the network layers cooperate to achieve the ultimate task: anyone, anywhere, anytime, and any media communications.

Local Area Networks Local area networks (LANs) mainly connect computers and other electronic devices (servers, printers, etc.) within an office, a single building, or a few adjacent buildings. Therefore, the geographical coverage of LANs is very small, spanning from a few meters to a few hundred meters. LANs are generally not a part of public networks but are owned and operated by private organizations. Common

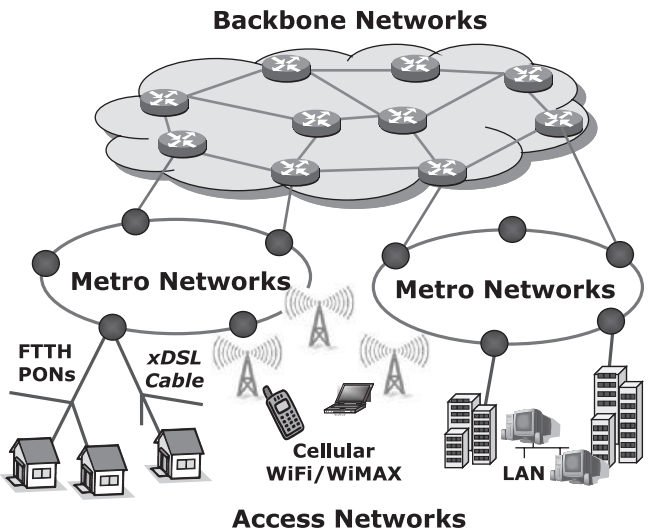


FIGURE 1.1 Hierarchical architecture of global communication infrastructure.

topologies for LANs are bus, ring, star, or tree. The most popular LANs are parts of the Ethernet, supporting a few hundred users with typical bit rates of 10 or 100 Mb/s.

Access Networks The computers and other communication equipment of a private organization are usually connected to a public telecommunication networks through access networks. Access networks bridge end users to service providers through twist pairs (phone line), coaxial cables, or other leased lines (such as OC3 through optical fiber). The typical distance covered by an access network is a few kilometers up to 20 km. For personal users, access networks use DSL or cable modem technology with a transmission rate of a few megabits per second; for business users, networks employ point-to-point fiber links with hundreds of megabits or gigabits per second.

Metropolitan Area Networks Metropolitan area networks (MANs) aggregate the traffic from access networks and transport the data at a higher speed. A typical area covered by a MAN spans a metropolitan area or a small region in the countryside. Its topology is usually a fiber ring connecting multiple central offices, where the transmission data rate is typically 2.5 or 10 Gb/s.

Wide Area Networks Wide area networks (WANs) carry a large amount of traffic among cities, countries, and continents. MAN multiplexes traffic from LANs and transports the aggregated traffic at a much higher data rate, typically tens of gigabits per second or higher using wavelength-division multiplexing (WDM) technology over optical fibers. Whereas a WAN covers the area of a nation or, in some cases, multiple nations, a link or path through a MAN could be as long as a few thousand kilometers. Beyond MANs, submarine links connect continents. Generally, the submarine systems are point-to-point links with a large capacity and an extremely long path, from a few thousand up to 10,000 km. Because these links are designed for ultralong distances and operate under the sea, the design requirements are much more stringent than those of their terrestrial counterparts. Presently, submarine links are deployed across the Pacific and Atlantic oceans. Some shorter submarine links are also widely used in the Mediterranean, Asian Pacific, and African areas.

Service Convergence Historically, communication networks provide mainly three types of service: voice, data, and video (triple play). Voice conversation using plain old telephony is a continuous 3.4-kHz analog signal carried by two-way, point-to-point circuits with a very stringent delay requirement. The standard TV signal is a continuous 6-MHz analog signal usually distributed with point-to-multipoint broadcasting. Data transmission is typically bursty with varying bandwidth and delay requirements. Because the traffic characteristics of voice, data, and video and their corresponding requirements as to quality of service (QoS) are fundamentally different, three major types of networks were developed specifically to render these services in a cost-effective manner: PSTN (public-switched telephone networks) for voice conversation, HFC (hybrid fiber coax) networks for video distribution, and the Internet for data transfer. Although HFC networks are optimized for video broadcasting, the inherent one-way communication is not suitable for bidirectional data or

voice. PSTN adopts circuit switching technology to carry information with specific bandwidth or data rates, such as voice signals. However, circuit-switched networks are not very efficient for carrying bursty data traffic. With packet switching, the Internet can support bursty data transmission, but it is very difficult to meet stringent delay requirements for certain applications. Therefore, no single network can satisfy all the service requirements.

Emerging multimedia applications such as video on demand, e-learning, and interactive gaming require simultaneous transmission of voice, data, and video. Driven by user demands and stiff competition, service providers are moving toward a converged network for multimedia applications, which will utilize Internet protocol (IP) technologies to provide triple-play services. As VoIP (voice over IP) has been developed in the past few years and more recently IP TV has become a mature technology, all network services will converge into an IP-based service platform. Furthermore, the integration of optical and wireless technologies will make quadruple play (voice, data, video, and mobility) a reality in the near future.

1.2 ACCESS TECHNOLOGIES

Emerging multimedia applications continuously fuel the explosive growth of the Internet and gradually pervade every area of our lives, from home to workplace. To provide multimedia service to every home and every user, access networks are built to connect end users to service providers. The link between service providers and end users is often called the *last mile* by service providers, or from an end user's perspective, the *first mile*. Ideally, access networks should be a converged platform capable of supporting a variety of applications and services. Through broadband access networks, integrated voice, data, and video service are provided to end users. However, the reality is that access networks are the weakest links in the current Internet infrastructure. While national information highways (WANs and MANs) have been developed in most parts of the globe, ramps and access routes to these information highways (i.e., the first/last mile) are mostly bike lanes or at best, unpaved roads, causing traffic congestion. Hence, pervasive broadband access should be a national imperative for future Internet development. In this section we review current access scenarios and discuss the last-mile bottleneck and its possible solutions.

1.2.1 Last-Mile Bottleneck

Due to advances in photonic technologies and worldwide deployment of optical fibers, during the last decade the telecommunication industry has experienced an extraordinary increase in transmission capacity in core transport networks. Commercial systems with 1-Tb/s transmission can easily be implemented in the field, and the state-of-the-art fiber optical transmission technology has reached 10 Tb/s in a single fiber. In the meanwhile, at the user end, the drastic improvement in the performance of personal computers and consumer electronic devices has made possible expanding demands of multimedia services, such as video on demand, video conferencing,

TABLE 1.1 Multimedia Applications and Their Bandwidth Requirements

Application	Bandwidth	Latency	Other Requirements
Voice over IP (VoIP)	64 kb/s	200 ms	Protection
Videoconferencing	2 Mb/s	200 ms	Protection
File sharing	3 Mb/s	1 s	
SDTV	4.5 Mb/s/ch	10 s	Multicasting
Interactive gaming	5 Mb/s	200 ms	
Telemedicine	8 Mb/s	50 ms	Protection
Real-time video	10 Mb/s	200 ms	Content distribution
Video on demand	10 Mb/s/ch	10 s	Low packet loss
HDTV	10 Mb/s/ch	10 s	Multicasting
Network-hosted software	25 Mb/s	200 ms	Security

e-learning, interactive games, VoIP, and others. Table 1.1 lists common end-user applications and their bandwidth requirements. As a result of the constantly increasing bandwidth demand, users may require more than 50 Mb/s in the near future. However, the current copper wire technologies bridging users and core networks have reached their fundamental bandwidth limits and become the *first-last-mile bottleneck*. Delays in Web page browsing, data access, and audio/video clip downloading have earned the Internet the nickname "World Wide Wait." How to alleviate this bottleneck has been a very challenging task for service providers.

1.2.2 Access Technologies Compared

For broadband access services, there is strong competition among several technologies: digital subscriber line, hybrid fiber coax, wireless, and FTTx (fiber to the x, x standing for home, curb, neighborhood, office, business, premise, user, etc.). For comparison, Table 1.2 lists the bandwidths (per user) and reaches of these competing technologies. Currently, dominant broadband access technologies are digital

TABLE 1.2 Comparison of Bandwidth and Reach for Popular Access Technologies

Service	Medium	Downstream (Mb/s)	Upstream (Mb/s)	Max Reach (km)
ADSL	Twisted pair	8	0.896	5.5
ADSL2	Twisted pair	15	3.8	5.5
VDSL1	Twisted pair	50	30	1.5
VDSL2	Twisted pair	100	30	0.5
HFC	Coax cable	40	9	25
BPON	Fiber	622	155	20
GPON	Fiber	2488	1244	20
EPON	Fiber	1000	1000	20
Wi-Fi	Free space	54	54	0.1
WiMAX	Free space	134	134	5

subscriber loop and coaxial cable. For conventional ADSL (asymmetric DSL) technology, the bandwidth available is a few Mb/s within the 5.5-km range. Newer VDSL (very high-speed DSL) can provide 50 Mb/s, but the maximum reach is limited to 1.5 km. On the other hand, coaxial cable has a much larger bandwidth than twist pairs, which can be as high as 1 Gb/s. However, due to the broadcast nature of CATV system, current cable modems can provide each user with an average bandwidth of a few Mb/s. While DSL and cable provide wired solutions for broadband access, Wi-Fi (wireless fidelity), and WiMAX (worldwide interoperability for microwave access) provide mobile access in a LAN or MAN network. Even though a nominal bandwidth of Wi-Fi and WiMAX can be relatively higher (54 Mb/s in 100 m for Wi-Fi and 28 Mb/s in 15 km for WiMAX), the reach of such wireless access is very limited and the actual bandwidth provided to users can be much lower, due to the interference in wireless channels. As a LAN technology, the primary use of Wi-Fi is in home and office networking. To reach the central office or service provider, multiple-hop wireless links with WiMAX have to be adopted. An alternative technology that is also under development is MBWA (mobile broadband wireless access, IEEE 802.20), which is very similar to WiMAX (IEEE 802.16e). Compared to the fixed access solutions, the advantages of the wireless technologies are easy deployment and ubiquitous or mobile access, and the disadvantages are unreliable bandwidth provisioning and/or limited access range.

The bandwidth and/or reach of the copper wire and wireless access technology is very limited due to the physical media constraints. To satisfy the future use demand ($>$30 Mb/s), there is a strategic urgency for service providers to deploy FTTx networks. Currently, for cost and deployment reasons, FTTx is competing with other access technologies. Long term, however, only optical fiber can provide the unlimited capacity and performance that will be required by future broadband services. FTTx has long been dubbed as a future-proof technology for the access networks. A number of optical access network architectures have been standardized (APON, BPON, EPON, and GPON), and cost-effective components and devices for FTTx have matured. We are currently witnessing a worldwide deployment of optical access networks and a steady increase in FTTx users.

1.3 DIGITAL SUBSCRIBER LINE

Digital subscriber line (also called *digital subscriber loop*) is a family of access technologies that utilize the telephone line (twisted pair) to provide broadband access service. While the audio signal (voice) carried by a telephony system is limited from 300 to 3400 Hz, the twisted pair connecting the users to the central office is capable of carrying frequencies well beyond the 3.4-kHz upper limit of the telephony system. Depending on the length and the quality of the twisted pair, the upper limit can extend to tens of megahertz. DSL takes advantage of this unused bandwidth and transmits data using multiple-frequency channels. Thus, some types of DSL allow simultaneous use of the telephone and broadband access on the same twisted pair.

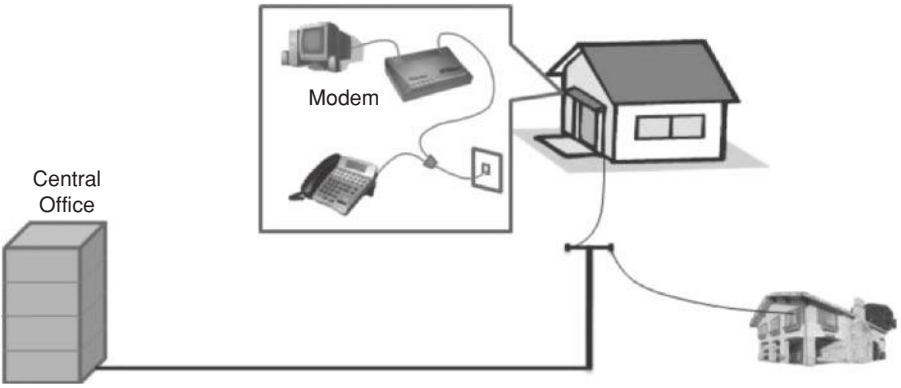


FIGURE 1.2 DSL access networks.

Figure 1.2 shows the typical setup of a DSL configuration. At the central office, a DSLAM (DSL access multiplexer) sends the data to users via downstream channels. At the user side, a DSL modem functions as a modulator/demodulator (i.e., receives data from DSLAM and modulates user data for upstream transmission).

1.3.1 DSL Standards

DSL comes in different flavors, supporting various downstream/upstream bit rates and access distances. DSL standards are defined in ANSI T1, and ITU-T Recommendation G.992/993. Table 1.2 lists various DSL standards and their performance. Collectively, these DSL technologies are referred to as *xDSL*. Two commonly deployed DSL standards are ADSL and VDSL.

As its name suggests, ADSL supports asymmetrical transmission. Since the typical ratio of traffic asymmetry is about 2 : 1 to 3 : 1, ADSL becomes a popular choice for broadband access. In addition, there is more crosstalk from other circuits at the DSLAM end. As the upload signal is weak at the noisy DSLAM end, it makes sense technically to have upstream transmission at a lower bit rate. Depending on the length and quality (such as the signal-to-noise ratio) of the twisted pair, the downstream bit rate can be as high as 10 times the upstream transmission. The maximum reach of ADSL is 5500 m. While ADSL1 can support a downstream bit rate up to 8 Mb/s and an upstream data rate up to 896 kb/s, ADSL2 supports up to 15 Mb/s downstream and 3.8 Mb/s upstream.

To support higher bit rates, the VDSL standard was developed after ADSL. Trading transmission distance for data rate, VDSL can support a much higher data rate but with very limited reach. VDSL1 standards specify data rates of 50 Mb/s for downstream and 30 Mb/s for upstream transmission. The maximum reach of VDSL1 is limited to 1500 m. The newer version of VDSL standards, VDSL2, is an enhancement of

VDSL1, supporting a data rate up to 100 Mb/s (with a transmission distance of 500 m). At 1 km, the bit rate will drop to 50 Mb/s. For reaches longer than 1.6 km, the VDSL2 performance is close to ADSL. Because of its higher data rates and ADSL-like long reach performance, VDSL2 is considered to be a very promising solution for upgrading existing ADSL infrastructure.

ADSL and VDSL are designed for residential subscribers with asymmetric bandwidth demands. For business users, symmetrical connections are generally required. Two symmetrical DSL standards, HDSL and SHDSL, are developed for business customers. While HDSL supports a T1 line data rate at 1.552 Mb/s (including 8 kb/s of overhead) with a reach of about 4000 m, SHDSL can provide a 6.696-Mb/s data rate with a maximum reach of 5500 m. However, HDSL and SHDSL do not support simultaneous telephone service, as most business customers do not have a requirement for a simultaneous voice circuit.

1.3.2 Modulation Methods

DSL uses a DMT (discrete multitone) modulation method. In DMT modulation, complex-to-real inverse discrete Fourier transform is used to partition the available bandwidth of the twisted pair into 256 orthogonal subchannels. DMT is adaptive to the quality of the twisted pair, so all the available bandwidth is fully utilized. The signal-to-noise ratio of each subchannel is monitored continuously. Based on the noise margin and bit error rate, a set of subchannels are selected, and a block of data bits are mapped into subchannels. In each subchannel, QAM (quadrature amplitude modulation) with a 4-kHz symbol rate is used to modulate the bit stream onto a subcarrier, leading to 60 kb/s per channel. Typically, the frequency range between 25 and 160 kHz is used for upstream transmission, and 140 kHz to 1.1 MHz is used for downstream transmission.

1.3.3 Voice over DSL

DSL was designed originally to carry data over phone lines, and DSL signal is separated from voice signal. Recently, new protocols have been proposed to merge voice and data at the circuit level. With advanced coding technologies, a 64-kb/s digitized voice signal can be compressed to 8 kb/s or less, thus allowing more voice channels to be carried over the same phone line. A voice over a DSL (VoDSL) gateway converts and compresses the analog voice signal to digital bit streams, so that calls made over VoDSL are indistinguishable from conventional calls. Usually, 12 to 20 voice channels can be carried over a single DSL line, depending on the transmission distance and the signal quality. A VoDSL system can be integrated into higher-layer protocols such as IP and ATM. Early DSL networks used ATM to ensure QoS, where ATM virtual circuits were used for the voice traffic. ADSL and VDSL networks migrate to packet-based transport, and they use packet-switched based virtual circuits instead of ATM ones.

1.4 HYBRID FIBER COAX

Cable networks were originally developed for a very simple reason: TV signal distribution. Therefore, cable networks are optimized for one-way, point-to-multipoint broadcasting of analog TV signals. As optical communication systems were developed, most cable TV systems have gradually been upgraded to hybrid fiber coax (HFC) networks, eliminating numerous electronic amplifiers along the trunk line. However, before cable access technology can be deployed, a return pass must be implemented for upstream traffic. To support two-way communication, bidirectional amplifiers have to be used in HFC systems, where filters are deployed to split the upstream (forward) and downstream (reverse) signals for separate amplification.

Figure 1.3 presents the network architecture of a typical HFC network. In HFC networks, analog TV signals are carried from the cable headend to distribution nodes using optical fibers, and from the distribution node, coaxial cable drops are deployed to serve 500 to 2000 subscribers. As shown in the figure, an HFC network is a shared medium system with a tree topology. In such a topology, multiple users share the same HFC infrastructure, so medium access control is required in upstream transmission while downstream transmission uses a broadcast scheme. A cable modem deployed at the subscriber end provides data connection to the cable network, while at the headend, the cable modem termination system connects to a variety of data servers and provides service to subscribers.

Compared with the twisted pairs in a telephone system, coaxial cables have a much higher bandwidth (1000 MHz), thus can support a much higher data rate. Depending on the signal-to-noise ratio on the coaxial cable, 40 Mb/s can be delivered to the end users with QAM modulation. For upstream transmission, QPSK can deliver up to a 10-Mb/s data rate. However, as cable systems are shared-medium networks, the bandwidth is thus shared by all the cable modems connected to the network. By contrast, DSL uses dedicated twist pairs for each user, thus no bandwidth sharing for different users. Furthermore, as the transmission bandwidth must be shared by multiple users, medium access control protocol must be deployed to govern upstream transmission. If congestion occurs in a specific channel, the headend must be able to instruct cable modems to tune its receiver to a different channel.

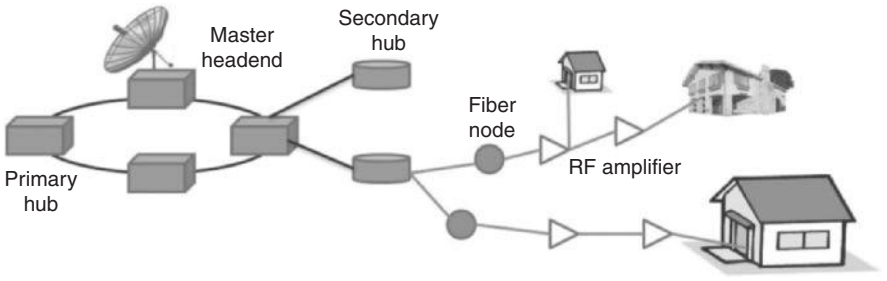


FIGURE 1.3 HFC access networks.

1.4.1 Cable Modem

Cable modems were developed to transport high-speed data to and from end users in an HFC network. Traditional TV broadcasting occupies frequencies up to 1 GHz, with each TV channel occupying 6 MHz of bandwidth (Part 76 in the FCC rules). A cable modem uses two of those 6-MHz channels for data transmission. For upstream transmission, a cable modem sends user data to the headend using a 6-MHz band between 5 and 42 MHz. At the same time, the cable modem must tune its receiver to a 6-MHz band within a 450- to 750-MHz band to receive downstream data. While a QAM modulation scheme is used for downstream data, a QPSK modulation scheme is usually selected for upstream transmission, as it is more immune to the interference resulting from radio broadcasting.

1.4.2 DOCSIS

DOCSIS (Data Over Cable Service Interface Specifications), developed by CableLabs, a consortium of equipment manufactuers, is the current standard for cable access technology. DOCSIS defines the functionalities and properties of cable modems at a subscriber's premises and cable modem termination systems at the headend. As its name suggests, DOCSIS specifies the physical layer characteristics, such as transmission frequency, bit rate, modulation format, and power levels, of cable modem and cable modem termination systems, but also the data link layer protocol, such as frame structure, medium access control, and link security. Three different versions of DOCSIS (1.0/2.0/3.0) was developed during the past decade and were later ratified as ITU-T Recommendation J.112, J.122, and J.222. Although some compromise is needed as cable networks are a shared medium, DOCSIS offers various classes of service with medium access control. Such QoS features in DOCSIS can support applications (such as VoIP) that have stringent delay or bandwidth requirements.

Physical Layer The upstream PMD layer supports two modulation formats: QPSK and 16-QAM, and the downstream PMD layers uses 64-QAM and 256-QAM. The nominal symbol rate is 0.16, 0.32, 0.64, 1.28, 2.56, or 5.12 Mbaud. Therefore, the maximum downstream data rate is about 40 Mb/s and the upstream data rate is about 20 Mb/s. To mitigate the effect of noise and other detrimental channel effects, Reed–Solomon encoding, transmitter equalizer, and variable interleaving schemes are commonly used.

Data Link Layer The DOCSIS data link layer specifies frame structure, MAC, and link security. The frame structure used in HFC networks is very similar to the Ethernet in both the upstream and downstream directions. For the downstream direction, data frames are embedded in 188-byte MPEG-2 (ITU-T H.222.0) packets with a 4-byte header followed by 184 bytes of payload. Downstream uses TDM transmission schemes, synchronous to all modems. In the upstream direction, TDMA or S-CDMA are defined for medium access control. An upstream packet includes physical layer overhead, a unique word, MAC overhead, packet payload, and FEC bytes. MAC

layer specifications also include modem registration, ranging, bandwidth allocation, collision detection and contention resolution, error detection, and data recovery. An access security mechanism in DOCSIS defines a baseline privacy interface, security system interface, and removable security module interface, to ensure information security in HFC networks.

1.5 OPTICAL ACCESS NETWORKS

Due to their ultrahigh bandwidth and low attenuation, optical fibers have been widely deployed for wide area networks and metro area networks. To some extent, multimode fibers were also deployed in office buildings for local area networks. Even though optical fibers are ideal media for high-speed communication systems and networks, the deployment cost was considered prohibitive in the access area, and copper wires still dominate in the current marketplace. However, as discussed in Section 1.2, emerging multimedia applications have created such large bandwidth demands that copper wire technologies have reached their bandwidth limits. Meanwhile, low-cost photonic components and passive optical network architecture have made fiber a very attractive solution. In the past few years, various PON architecture and technologies have been studied by the telecom industry, and a few PON standards have been approved by ITU-T and IEEE. FTTx becomes a mature technology in direct competition with copper wires. In fact, large-scale deployment has started in Asia, North America, and Europe, and millions of subscribers are enjoying the benefit of PON technologies.

1.5.1 Passive Optical Networks

Figure 1.4 illustrates the architecture of a passive optical network. As the name implies, there is no active component between the central office and the user premises. Active devices exist only in the central office and at user premises. From the central office, a standard single-mode optical fiber (feeder fiber) runs to a $1 : N$ passive optical power splitter near the user premises. The output ports of the passive splitter connects to the subscribers through individual single-mode fibers (distribution fibers). The transmission distance in a passive optical networks is limited to 20 km, as specified in current standards. The fibers and passive components between the central office and users premises are commonly called an optical distribution network. The number of users supported by a PON can be anywhere from 2 to 128, depending on the the power budget, but typically, 16, 32, or 64. At the central office, an optical line terminal (OLT) transmits downstream data using 1490-nm wavelength, and the broadcasting video is sent through 1550-nm wavelength. Downstream uses a broadcast and select scheme; that is, the downstream data and video are broadcast to each user with MAC addresses, and the user selects the data packet–based MAC addresses. At the user end, an optical network unit (ONU), also called an optical network terminal (ONT), transmits upstream data at 1310-nm wavelength. To avoid collision, upstream transmission uses a multiple access protocol (i.e., time-division multiple access) to assign time slots to

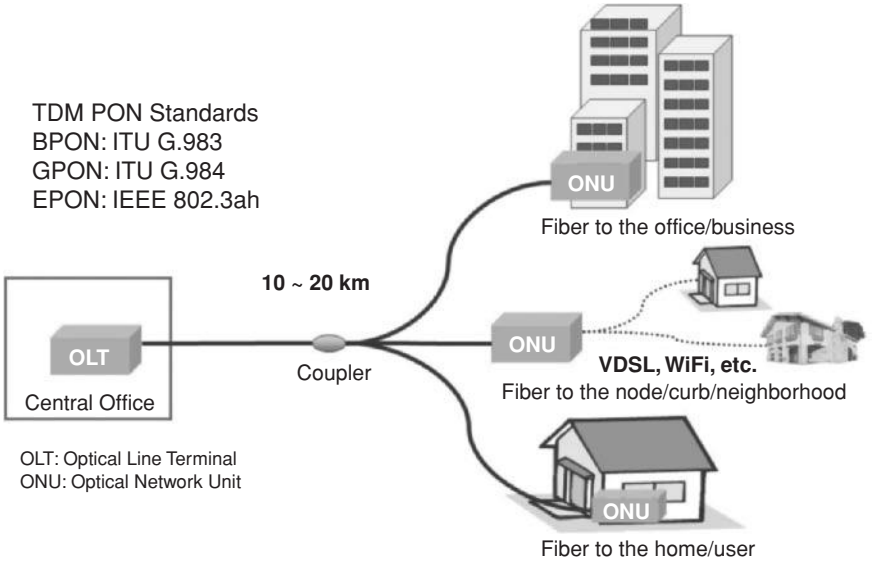


FIGURE 1.4 Passive optical networks.

each user. This type of passive optical network is called TDM PON. The ONU could be located in a home, office, a curbside cabinet, or elsewhere. Thus comes the so-called fiber-to-the-home/office/business/neighborhood/curb/user/premises/node, all of which are commonly referred to as *fiber to the x*. In the case of fiber-to-the-neighborhood/curb/node, twisted pairs are typically deployed to connect end users to the ONUs, thus providing a hybrid fiber/DSL access solution.

1.5.2 PON Standard Development

Early work of passive optical networks started in 1990s, when telecom service providers and system equipment vendors formed the FSAN (full service access networks) working group. The common goal of the FSAN group is to develop truly broadband fiber access networks. Because of the traffic management capabilities and robust QoS support of ATM (asynchronous transfer mode), the first PON standard, APON, is based on ATM and hence referred to as ATM PON. APON supports 622.08 Mb/s for downstream transmission and 155.52 Mb/s for upstream traffic. Downstream voice and data traffic is transmitted using 1490-nm wavelength, and downstream video is transmitted with 1550-nm wavelength. For upstream, user data are transmitted with 1310-nm wavelength. All the user traffic is encapsulated in standard ATM cells, which consists of 5-byte control header and 48-byte user data. APON standard was ratified by ITU-T in 1998 in Recommendation G.983.1. In the early days, APON was most deployed for business applications (e.g., fiber-to-the-office). However, APON networks are largely substituted with higher-bit-rate BPONs and GPONs.

Based on APON, ITU-T further developed BPON standard as specified in a series of recommendations in G.983. BPON is an enhancement of APON, where a higher data rate and detailed control protocols are specified. BPON supports a maximum downstream data rate at 1.2 Gb/s and a maximum upstream data rate at 622 Mb/s. ITU-T G.983 also specifies dynamic bandwidth allocation (DBA), management and control interfaces, and network protection. There has been large-scale deployment of BPON in support of fiber-to-the-premises applications.

The growing demand for higher bandwidth in the access networks stimulated further development of PON standards with higher capacity beyond those of APON and BPON. Starting in 2001, the FSAN group developed a new standard called gigabit PON, which becomes the ITU-T G.984 standard. The GPON physical media–dependent layer supports a maximum downstream/upstream data rate at 2.488 Gb/s, and the transmission convergence layer specifies a GPON frame format, media access control, operation and maintenance procedures, and an encryption method. Based on the ITU-T G.7041 generic framing procedure, GPON adopts GEM (a GPON encapsulation method) to support different layer 2 protocols, such as ATM and Ethernet. The novel GEM encapsulation method is backwardly compatible with APON and BPON and provides better efficiency than do Ethernet frames. Deployment of GPON had taken off in North America and largely replaced older BPONs and more.

While ITU-T rolled out BPON and GPON standards, IEEE Ethernet-in-the-first-mile working group developed a PON standard based on Ethernet. The EPON physical media–dependent layer can support maximum 1.25-Gb/s (effective data rate 1.0 Gb/s) downstream/upstream traffic. EPON encapsulate and transport user data in Ethernet frames. Thus, EPON is a natural extension of the local area networks in the user premises, and connects LANs to the Ethernet-based MAN/WAN infrastructure. Since there is no data fragment or assembly in EPON and its requirement on physical media–dependent layer is more relaxed, EPON equipment is less expensive than GPON. As Ethernet has been used widely in local area networks, EPON becomes a very attractive access technology. Currently, EPON networks have been deployed on a large scale in Japan, serving millions of users.

1.5.3 WDM PONs

As the user bandwidth demands keep increasing, current GPON or EPON will eventually no longer be able to satisfy the bandwidth requirement. There are a few possible solutions. One possibility is to split a single PON into multiple PONs so that each PON supports fewer users and each user gets more bandwidth. Another alternative is to use a higher bit rate, such as 10 Gb/s. In fact, an IEEE 802.3av study group is creating a draft standard on 10-Gb/s EPON. However, both solutions for higher bandwidth (i.e., higher bit rate or fewer users per PON) are not very cost-effective and do not scale very well as the bandwidth demands increase further. In addition, the power distribution of the passive splitter is fixed; that will lead to an uneven power budget for users and limit the transmission distance. Ultimately, WDM PON is the only future proof of technology that can satisfy any bandwidth demands.

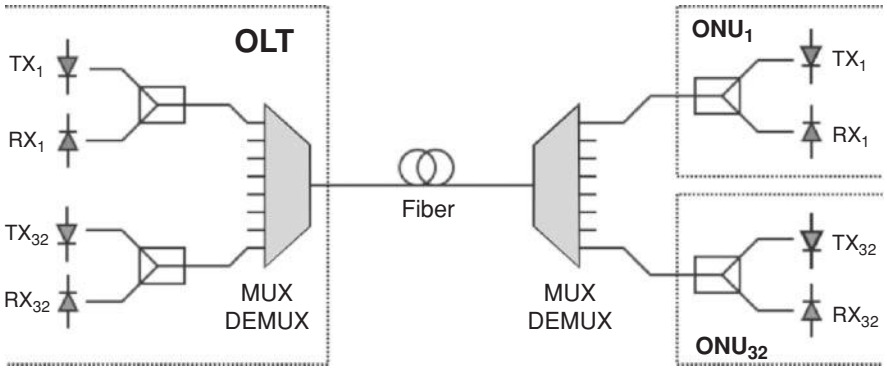


FIGURE 1.5 WDM passive optical networks.

Figure 1.5 shows the network architecture of WDM PONs. Transmitters with varying wavelengths will be deployed at the OLT and ONU sides, and a passive wavelength-division multiplexer will be inserted at the distribution node to separate and combine multiple wavelengths. Thus, the fiber distribution network will be kept passive. If the user bandwidth demands are not very large, or in the other words, a small number of users can still share a single wavelength, a passive power splitter following the WDM is used to broadcast the downstream traffic and combine the upstream traffic. In this case, multiple wavelengths separate a single PON into multiple logical TDM PONs. Each PON runs on a different wavelength, and fewer users share the bandwidth of a TDM PON. In addition, since the optical power is split for a smaller number of users, WDM PONs is less subject to optical power budget constraints, leading to long-reach access networks. If a user requires a large amount of bandwidth (e.g., a few gigabits per second), a single wavelength can be provided for this specific user; or in an extreme case, multiple wavelengths, hence a large bandwidth, can be provided to a single user if needed.

In WDM PONs, the equipment and resources at OLT are shared by fewer users, leading to higher cost per user. Hence, WDM PONs are considered much more expensive than TDM PONs. However, to support high-bandwidth applications, there will be a need in the near future to move from TDM access networks to WDM access networks. Currently, the way to migrate from current TDM access networks to WDM access networks in a cost-effective, flexible, and scalable manner is not at all clear. A method to upgrade the access service smoothly and cost-effectively from a current TDM FTTx network to a future WDM FTTx network with a minimum influence on legacy users is the object of intense research. Various approaches to implementing WDM have been and are being explored, and field deployment has begun in Asia (South Korea, to be exact). A number of schemes to incorporate WDM technology into access networks have been studied and tested in experiments, and the WDM FTTx network architecture exhibits certain exceptional features in the WDM implementation in either downstream, upstream, or both directions. As optical

technology becomes cheaper and easier to deploy and end users demand ever-increasing bandwidth, WDM PONs will eventually make the first/last-mile bottleneck history.

1.5.4 Other Types of Optical Access Networks

In addition to the passive optical networks, TDM and WDM PONs, that we have discussed, other types of optical access networks have been developed over the years, including Ethernet over fiber, DOCSIS PON, RF PON, and free-space optical networks. Ethernet over fiber is essentially point-to-point Ethernet built on fiber links. DOCSIS and RF PON is two flavors of PON developed for cable companies. Free-space optical networks is a wireless access solution utilizing optical communication technologies.

Ethernet over Fiber Ethernet over fiber is deployed primarily in point-to-point topology. Typically, dedicated fiber connects a subscriber to the central office, and each subscriber requires two dedicated transceivers (one at the user premises and the other at the central office). This approach requires a large number of fibers and optical transceivers and thus incurs a large cost associated with fiber and equipment. Since each fiber link can run on its full capacity, Ethernet over fiber, which requires gigabit bandwidth, is used primarily for business subscribers. Figure 1.6 shows an

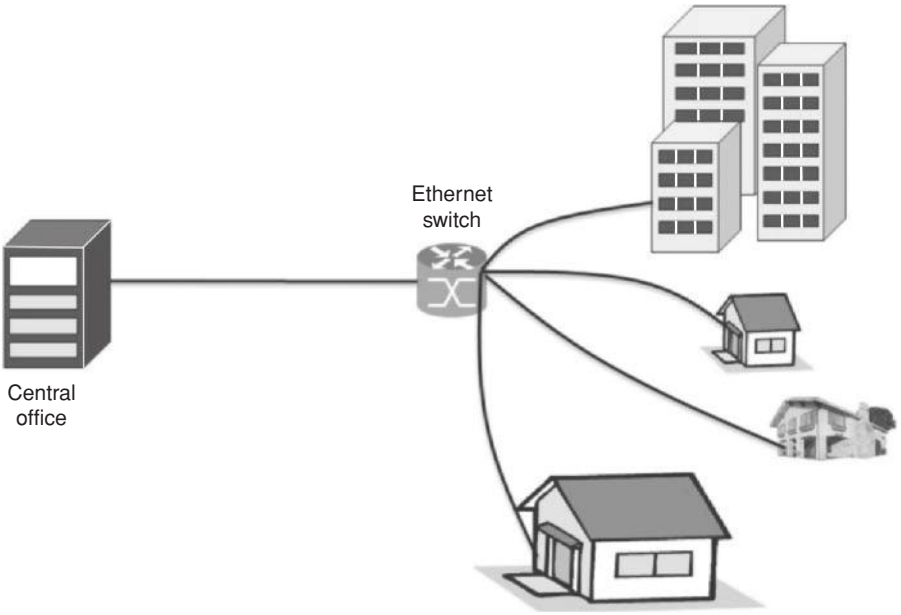


FIGURE 1.6 Point-to-point Ethernet optical access networks.

alternative architecture for Ethernet over fiber. A local Ethernet switch is deployed to the user sites. Individual fiber can then run from the switch to each user, and only a single fiber (bidirectional) or two fibers (unidirectional) connect the Ethernet switch to the central office. This approach reduces the number of fibers run from the central office but requires an active Ethernet switch in the field and requires at least two more transceivers than is the case on the left in the figure.

DOCSIS PON While telecom companies are deploying PONs worldwide on a large scale, MSOs (multisystem operators) need to upgrade their fiber coax systems to compete in FTTx markets. DOCSIS PON, or DPON, is developed to provide a DOCSIS service layer interface on top of PON architecture. DPON implements DOCSIS functionalities, including OAMP (operation, administration, maintenance, and provisioning) on existing PON systems, and thus allow MSOs to use set-top and DOCSIS equipment located in homes and headends over PONs. However, fundamentally, DPON service is based on current EPON or GPON MAC and physical layer standards. Therefore, DPON is just an application running on top of PON systems.

RF PON Radio-frequency PON (RF PON) is another flavor of passive optical networks developed for MSOs. RF PONs support RF video broadcasting signals over optical fibers. As MSOs expand the network footprint and launch new products using additional RF bandwidth, more active RF components are deployed and higher frequencies sometimes require RF electronics change-outs and respacing. As a consequence, HFC networks experience reduced signal quality, lower reliability, and higher operating and maintenance cost. RF PONs are a natural evolution of current HFC networks, as they offer backward compatibility with current RF video broadcasting technologies and provides significant cost reduction in network operation and maintenance.

OCDM PON Optical code-division multiplexing (OCDM) has been demonstrated recently as an alternative multiplexing technique for PONs. Similar to electronic CDMA technology, users in OCDM PONs are assigned orthogonal codes with which each user's data are encoded or decoded into or from optical pulse sequence. OCDM PONs can thus provide asynchronous communications and security against unauthorized users. However, the optical encoders and decoders for OCDM are expensive, and the number of users is limited by interference and noise.

Free-Space Optical Networks Unlike fiber optic communications, *free-space optical communication* (also called *optical wireless communication*) uses atmosphere as the communication medium. This is probably one of the old long-distance communication methods (e.g., smoke signals) used a few thousands years ago. During the past decades, there has been revived interest in free-space communication for satellite and urban environment. Particularly in the access networks, it can used to connect a subscriber directly to a central office. Figure 1.7 shows a typical setup

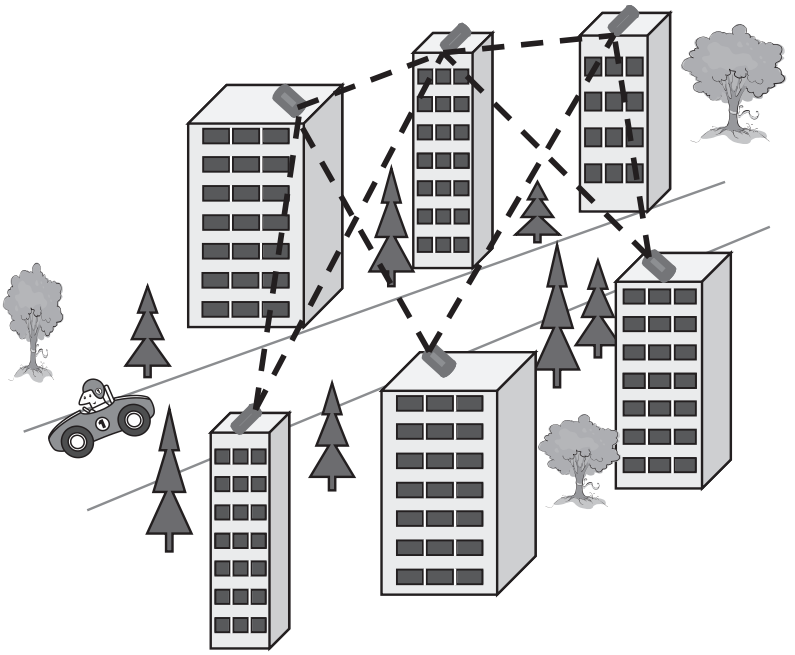


FIGURE 1.7 Free-space optical communications and networks. Point-to-point optical wireless links on the roofs of buildings form a mesh network for broadband access.

for urban free-space optical communication networks. Due to the line-of-sight requirement for free-space optical communications, optical transceivers are usually mounted on the tops of buildings, and telescopes are typically used in the transmitter to improve the alignment of optical links. Multiple point-to-point links can form a mesh network, improving its scalability and reliability. As a wireless technology, the cost of free-space optical communication is very low, about 10% of fiber optic communications, and the high-speed link can be set up and torn down in a couple hours. Compared to other wireless access technologies, it provides a higher data rate, longer reach, and better signal quality. So far, thousands of free-space optical links have been deployed. However, atmosphere is not an ideal transmission medium, due to attenuation and scattering at optical frequency. Turbulence, rain, and dense fog could be very challenging for free-space optical communication. For long-reach links, alignment of optical transmitters and receivers is also difficult, and an adaptive ray-tracking system might be needed for rapid pointing and accurate alignment. Potentially, survivable network topology, transmitter and receiver arrays, and adaptive and equalization technologies could help mitigate the atmospheric effect and alignment problem. Integration with wire line networks such as PONs can greatly improve the reliability and survivability of free-space optical access networks. In the future, we may witness more and more free-space optical networks in urban settings.

1.6 BROADBAND OVER POWER LINES

Ac power lines have long been considered a workable communication medium. For decades, utility companies have used power lines for signaling and control, but they are used primarily for internal management of power grids, household intercoms, and lighting controls. As deregulation of both the telecom and electricity industries was unfolding in the 1990s, broadband access over power lines became a possibility. As power lines reach more residences than does any other medium, significant efforts have been made to develop high-speed access over power lines. A number of solutions have been proposed and tested in the field. Even though DSL or cable currently dominates the broadband access services, and PONs are very promising for the near future, broadband over power lines (BPL) can still claim its part in the current market. For example, in some rural areas, building infrastructure to provide DSL or cable could be very expensive, while power-line communications could easily provide broadband services. Anywhere there is electricity there could be broadband over power lines. In addition, there is a great potential to network all the appliances in a household through the power line, thus providing a smart home solution. However, at present power-line communication technology and its market potential remain to be developed further.

1.6.1 Power-Line Communications

Figure 1.8 shows the topology of the electrical power distribution grid. The three-phase power generated at a power plant enters a transmission substation, where the three-phase power generated by the power generators is converted to extremely high

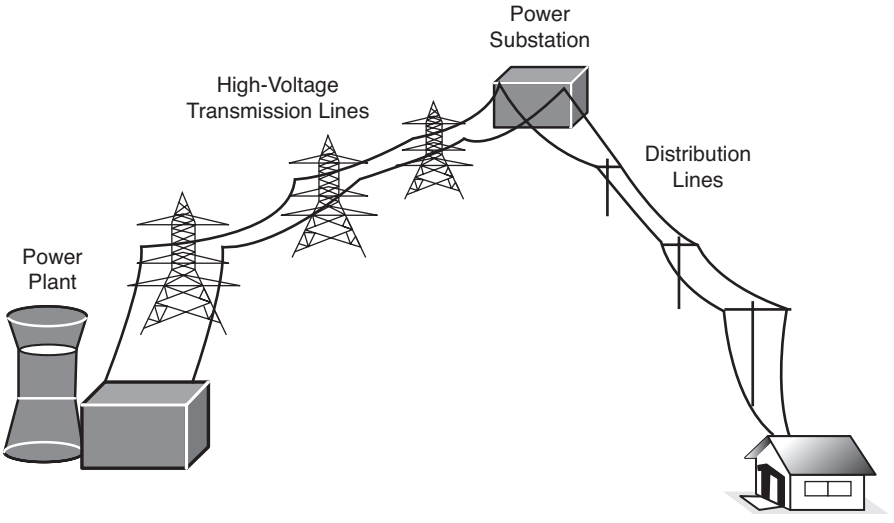


FIGURE 1.8 Electrical power transmission and distribution.

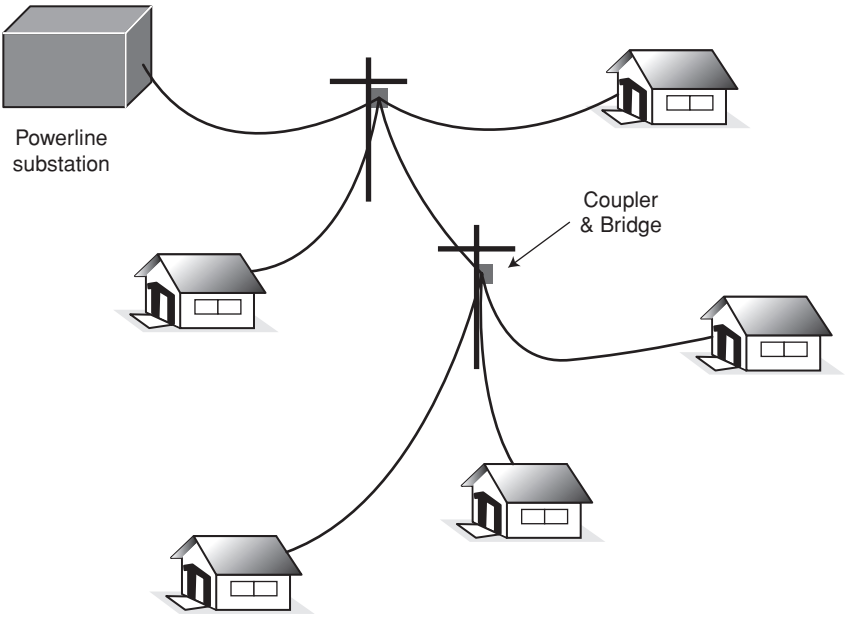


FIGURE 1.9 Broadband power-line communications.

voltages (155 to 765 kV) for long-distance transmission over the grid. Within the transmission grid, many power substations convert the extremely high transmission voltage down to distribution voltages (less than 10 kV), and this medium-voltage electricity is sent through a bus that can split the power in multiple directions. Along the distribution bus, there are regulator banks that regulate the voltage on the line to avoid overshoot or undershoot, and taps that send electricity down the street. At each building or house, there is a transformer drum attached to the electricity pole, reducing the medium voltage (typically, 7.2 kV) to household voltage (110 or 240 V).

Broadband over power lines utilizes the medium-voltage power lines to transmit data to and from each house, as shown in Figure 1.9. Typically, repeaters are installed along the power lines for long-distance data transmission, and some bypass devices allow RF signals to bypass transformers. In the last step of data transmission, the signals can be carried to each house by the power line or, alternatively, using Wi-Fi or other wireless technology for last-mile connection.

1.6.2 BPL Modem

A BPL modem plugs into a common power socket on the wall, sending and receiving data through a power line. On the other end, the BPL modem connects to computers or other network devices by means of Ethernet cables. In some cases, a wireless router can be integrated with a BPL modem. BPL modems transmit at medium to high frequencies, from a few megahertz to tens of megahertz. Typical

data rates supported by a BPL modem range from hundreds of kilobits per second to a few megabits per second. Various modulation schemes can be used for power-line communications, including the older ASK (amplitude shift keying), FSK (frequency shift keying) modulation and newer DMT, DSSS (direct sequence spread spectrum) and OFDM (orthogonal frequency-division multiplexing) technologies. DMT, DSSS, or OFDM modulation is perferred in modern BPL modems, as it is more robust in handling interference and noise. Recent research has demonstrated a gigabit data rate over power lines using microwave frequencies via surface wave propagation. This technology can avoid the interference problems very common in power lines.

1.6.3 Challenges in BPL

BPL is a promising technology, but its development is relatively slow compared with DSL and cable. There are a number of technical challenges that must be overcome. A power line is not a very good medium for data transmission: Various transformers used in the electric grid do not pass RF signals, the numerous sources of signal reflections (impedance mismatches and lack of proper impedance termination) on power lines hinder data transmission, and noise from numerous sources (such as power motors) contaminates the transmission spectrum. Since power lines consist of untwisted and unshielded wire, their long length makes them large antennas emitting RF signals and interfering with other radio communications. Furthermore, a power line is a shared medium limiting the bandwidth delivered to each user and raising security concerns for private communications. All these issues have to be fully addressed before large-scale deployment can be implemented. Fortunately, much progress has been made through intensive research during recent decades. BPL is poised to be a promising technology for entry into the current highly competitive market.

1.7 WIRELESS ACCESS TECHNOLOGIES

Starting with RF communication and broadcasting, wireless communication technologies have had an incredibly powerful effect on the entire world since the beginning of the twentieth century. Nowadays, AM/FM radio and TV broadcasting blanket every continent except Antarctica; wireless cellular networks provide voice communication to hundreds of millions of users; satellites provide video broadcasting and communication links worldwide; and Bluetooth and wireless LANs support mobile services to individuals. Wireless networks are everywhere. The popularity of wireless technologies is due primarily to their mobility, scalability, low cost, and ease of deployment. Wireless technologies will continue to play an important part in our daily lives, and fourth-generation wireless networks will be able to provide quadruple play through seamless integration of a variety of wireless networks, including wireless personal networks, wireless LANs, wireless access networks, cellular wide area networks, and satellite networks. In recent years, a number of wireless technologies have been developed as alternatives to traditional wired access service (DSL, cable, and PONs). Except for free-space optical communications (Section 1.5), most

wireless access networks use RF signals to establish communication links between a central office and subscribers. In this section we discuss various broadband radio access technologies and their characteristics. The choice of radio access technologies depends largely on the applications, required data rate, available frequency spectrum, and transmission distance. Even though wireless access networks cannot compete with wired access technologies in terms of data rate and reliability, they offer flexibility and mobility that no other technologies can provide. Therefore, wireless access networks complement current wired access technologies and will continue to grow in the future.

1.7.1 Wi-Fi Mesh Networks

The Wi-Fi network based on IEEE 802.11 standards was developed in the 1990s for wireless local area networks, where a set of wireless access points function as communication hubs for mobile clients. Because of its flexibility and low deployment cost, Wi-Fi has become an efficient and economical networking option that is widespread in both households and the industrial world, and is a standard feature of laptops, PDAs, and other mobile devices. Now Wi-Fi is available in thousands of public hot spots, millions of campus and corporate facilities, and hundreds of millions of homes. Even though current Wi-Fi networks are limited primarily to point-to-multipoint communications between access points and mobile clients, multiple access points can be interconnected to form a wireless mesh network, as shown in Figure 1.7. The wireless access points establish wireless links among themselves to enable automatic topology discovery and dynamic routing configuration. The wireless links among access points form a wireless backbone referred to as *mesh backhaul*. Multihop wireless communications in mesh backhaul are employed to forward traffic to and from a wired Internet entry point, and each access point may provide point-to-multipoint access to users known as *mesh access*. Therefore, a Wi-Fi mesh network can provide broadband access services in a self-organized, self-configured, and self-healing way, enabling quick deployment and easy maintenance.

Over the years, a set of standards has been specified by the IEEE 802.11 working group, including the most popular 802.11b/g standards. Table 1.3 compares the main attributes of these standards (pp152, 3G Wireless with WiMAX and Wi-Fi). The original 802.11 standard (approved in 1997) supports data rates of 1 or

TABLE 1.3 Comparison of IEEE 802.11 Standards

Parameter	802.11a	802.11b	802.11g	802.11n	802.11y
Operating frequency (GHz)	5	2.4	2.4	2.4 and 5	3.7
Maximum data rate (Mb/s)	54	11	54	248	54
Maximum indoor transmission distance (m)	35	40	40	70	50
Maximum outdoor transmission distance (m)	100	120	120	250	5000

2 Mb/s using FHSS (frequency hopping direct sequence) with GFSK modulation or DSSS (direct sequence spread spectrum) with DBPSK (differential binary-phase shift keying)/DQPSK (differential quadrature-phase shift keying) modulation. In 1999, 802.11b extended the original 802.11 standard to support 5.5- and 11-Mb/s data rates in addition to the original 1- and 2-Mb/s rates. The 802.11b standard uses eight-chip DSSS with a CCK (complementary code keying) modulation scheme at the 2.4-GHz band. Also approved in 1999 by the IEEE, 802.11a operates at bit rates up to 55 Mb/s using OFDM with BPSK, QPSK, 16-QAM, or 64-QAM at the 5-GHz band. In 2003, IEEE ratified a newer standard, IEEE 802.11g, providing a 54-Mb/s data rate at the 2.4-GHz band. The 802.11g standard is back-compatible with 802.11b. The upcoming IEEE 802.11n standard will support a 248-Mb/s data rate operating at the 2.4- and 5-GHz bands. In addition, IEEE 802.11e provides effective QoS support, and IEEE 802.11i supports enhanced security in wireless LANs. Even though Wi-Fi networks based on IEEE 802.11a/g/n can provide data rates over 50 Mb/s, their maximum reach is very limited ($<$ 500 m). For last-mile solution, Wi-Fi mesh networks with multihop paths are necessary. However, due to RF interference, bit rates for multihop wireless communication could be much lower than the maximum data rate of a single wireless link. To support a long reach, IEEE 802.11y is currently under development for 54 Mb/s with a maximum reach of 5 km (outdoor environment).

In wireless networks, interference from different transmitters can be a serious problem limiting the throughput of the entire network. In Wi-Fi networks, MAC layer control uses a contention-based medium access called CSMA/CA (carrier-sense multiple access with collision avoidance) to reduce the interference effect and improve network performance. However, because of the randomness of data packet arrival time and the contentious nature of the MAC layer protocol, the throughput of Wi-Fi networks can be much lower than its maximum capacity.

1.7.2 WiMAX Access Networks

WiMAX access networks, based on IEEE 802.16 standards, can provide wireless broadband Internet access at a relatively low cost. A single base station in WiMAX networks can support data rates up to 75 Mb/s to residential or business users. However, since multiple users are served by a single base station, data payload delivered to end users is likely to 1 Mb/s for residential subscribers and a few Mb/s for business clients. Compared to the transmission distance of a few hundred meters supported by Wi-Fi (802.11a/b/g/n), WiMAX promises wireless access range up to 50 km. Therefore, WiMAX can provide citywide coverage and QoS capabilities, supporting multimedia applications from non-real-time data to real-time voice and video. Furthermore, as an IP-based wireless technology, WiMAX can be integrated seamlessly with other types of wireless or wireline networks.

The salient features of a number of 802.16 standards ratified by IEEE are shown in Table 1.4. The original IEEE 802.16 standrad defines backhaul point-to-point connections with bit rates up to 134 Mb/s using frequencies in the range 10 to 66 GHz, and IEEE 802.16d/e specifies point-to-multipoint wireless access at bit rates

TABLE 1.4 Comparison of IEEE 802.16 Standards

Parameter	802.16	802.16a	802.16e	802.16m
Operating frequency (GHz)	10–66	2–11	2–6	To be determined
Maximum data rate (Mb/s)	134	75	15	1000
Typical cell size (km)	2–5	7–10	2–5	Microcell (to be determined)

up to 75 Mb/s. The newest standard, IEEE 802.16m, supports data rates up to 1 Gb/s but with a much shorter transmission range.

Figure 1.10 shows the architecture of a typical WiMAX network. In WiMAX networks, WiMAX base stations are connected to the wireline networks (usually, optical metro networks) using optical fiber, cable, and microwave high-speed point-to-point links. Theoretically, a base station can cover up to a 50-km radius, but in practice it is usually limited to 10 km. The base station serves a number of subscriber stations (deployed at the locations of residential or business users) using point-to-multipoint links. A WiMAX network can be configured with a star topology or a mesh topology; each has advantages and disadvantages. Whereas star topology can support higher data rates, mesh topology provides a longer reach and faster deployment. The WiMAX MAC layer allocates the uplink and downlink bandwidth to subscribers according to their bandwidth needs. Unlike Wi-Fi networks, WiMAX networks adopt

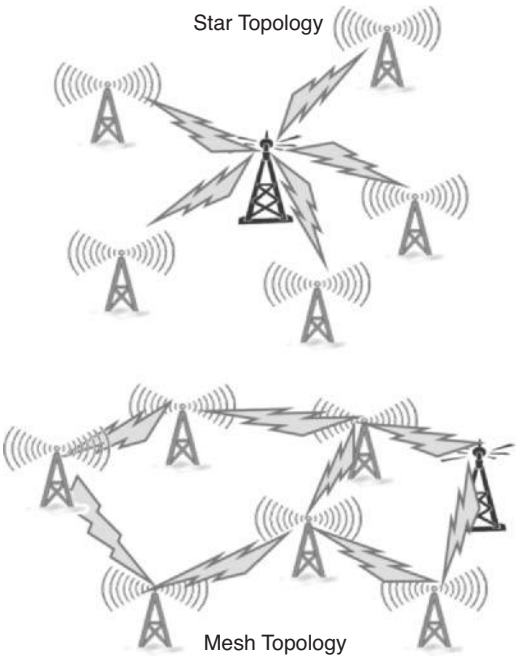


FIGURE 1.10 WiMAX network topology.

scheduled access using a time-division multiplexing technique, but the time slot assigned to each subscriber can vary in length depending on the bandwidth allocated to the subscriber. Because of the scheduling algorithm, WiMAX networks are more bandwidth efficient than are Wi-Fi networks.

1.7.3 Cellular Networks

During the last decade, cellular networks have spread all over the world, evolving from first generation (1G) to 2G and now moving toward 3G and 4G systems. The primary function of cellular networks is to carry voice communications for mobile users. However, as the telecom industry is migrating from voice- to data-centric networks, cellular networks have gradually built up their capacity for multimedia services such as data and video applications. As the first-generation cellular networks, AMPS (the Advanced Mobile Phone System) in North America and ETACS (the Extended Total Access Communication System) in Europe and Asia are analog, circuit-switched systems supporting only voice communications. The second-generation networks began the digital evolution. Digital encoding techniques such as CDMA, GSM, and TDMA pervade the cellular networks, and text messaging service becomes a common application. In addition, GPRS (general packet radio service) adds packet switching in GSM networks for high-speed data transmission (up to 171.2 kb/s), and EDGE (enhanced data rates for GSM evolution) further improved data transmission in GSM networks at bit rates up to 473.6 kb/s. The third-generation cellular networks based on UMTS (the universal mobile telecommunication system) or WCDM (wideband code-division multiple access) provide data service with bit rates above 144 kb/s. The emerging fourth-generation network will be an IP-based mobile system combining multiple radio access technologies, such as Bluetooth and wireless LAN, into an integrated network. The data rates supported by 4G networks could be as high as 100 Mb/s, thus providing truly broadband and ubiquitous access services.

Figure 1.11 illustrates the configuration of a typical cellular network, consisting of a base station controller, mobile switching center, base station transceiver, and mobile devices. To use the radio spectrum efficiently, the area covered by the cellular

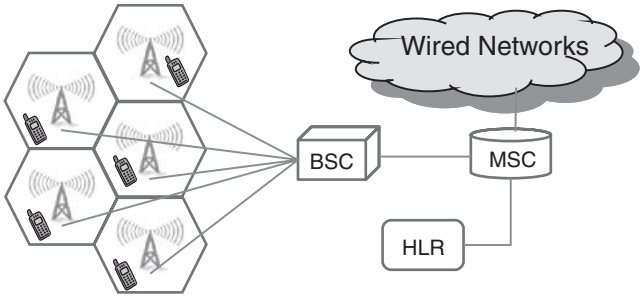


FIGURE 1.11 Cellular network architecture. MSC, mobile switching center; BSC, base station controller; HLR, home location register.

network is divided into small cells. Frequencies are reused in nonadjacent cells. Each cell has a base station that transmits and receives signals from the mobile devices within the cell. A group of base stations are connected to a base station controller. A group of base station controllers are in turn connected to a mobile switching center via microwave or fiber optic links. The base station controller controls communications between a group of base stations and a single mobile switching center. The mobile switching center connects to the public-switched telephone networks, which switch calls to other mobile stations or wired telephones. To provide data service, the mobile switching center is also connected to the Internet through edge routers.

Low-data-rate and incompatible technologies in current cellular networks (2G or 2.5G) present a serious problem for emerging multimedia applications. Hence, 3G networks have been developed to provide data rates over 1 Mb/s with a compatible radio interface among countries. However, economic concerns cast a doubt over large-scale deployment of 3G networks. Meanwhile, 4G technologies have emerged as a promising approach for mobile data service with a faster data rate than 3G. Despite all the efforts taken with developing data-centric cellular networks, broadband multimedia service over cellular networks still lags behind Wi-Fi and WiMAX networks in term of available bandwidth and network capacity.

1.7.4 Satellite Systems

Satellites have played an important role in providing digital communication links all over the world for a few decades. Originally developed for long-distance and intercontinental communications, satellites are also used for video broadcasting. Due to the development of VSATs (very small aperture terminals), satellite direct-to-home video broadcasting has been widely accepted since the mid-1990s. So far, satellite links have reached about 100 million homes, and widespread use of satellites for broadband access has become a reality. Satellite systems can cover a wide geographic area and support a variety of broadband applications, making it a very attractive broadband access solution. In fact, large corporate users have utilized satellite networks to establish wide area data networks to serve geographically dispersed corporate offices since the 1980s. A special type of satellite network called a *global positioning system* (GPS) has found popular applications for both military and civil navigation. Satellite operators for video broadcasting are dashing forward for broadband Internet access and multimedia applications.

Orbiting around the Earth, a satellite serves as a repeater and establishes a wireless link between any two users on the Earth. A satellite receives signals from Earth stations on an uplink, amplifies those signals, and then transmits them on a downlink with a different frequency. The first-generation satellites operate in the C-band, with a 4-GHz downlink and a 6-GHz uplink. However, large dish antennas have to be used for ground stations to improve receiver sensitivity and reduce microwave beamwidth. As bandwidth demands increase, the Ku-band (12/14 GHz for downlink/uplink) and Ka-band (20/30 GHz for downlink/uplink) were allocated by the U.S. Federal Communications Commission (FCC) for satellite communications. Using higher frequencies can support a higher data rate and permit the use of smaller-aperture antennas.

Recently, a higher-frequency satellite band, a V-band with a 40-GHz downlink and a 50-GHz uplink, has been approved by the ITU-T. With a V-band satellite, over 2 GHz of bandwidth is available, but atmospheric and rain attenuation become more severe at the V-band than at a lower-frequency band.

Modern communication satellites typically use a geostationary orbit with an orbital period matching the rotation period of the Earth. At a geostationary orbit, a single satellite can cover a huge geographical area (roughly 40% of the surface of the Earth). Since a geostationary orbit has a radius about 42,164 km, a long signal delay (about a 0.25 s round-trip delay) and large signal attenuation are unavoidable. Alternatively, low Earth orbit (200 to 2000 km orbital altitude) or medium Earth orbit (2000 to 3000 km orbital altitude) can be used with shorter delays and lower power attenuation. However, the coverage area of a low/medium-Earth-orbiting satellite is much smaller.

A satellite Earth station typically consists of a satellite modem, a transceiver, and an antenna. A parabolic reflector antenna is commonly used to transmit and receive satellite signals. A satellite transceiver includes low-noise frequency converters and power amplifiers made from microwave monolithic integrated circuits. A satellite modem performs data encoding and modulation. Since satellite links are mostly power limited, complicated encoding and modulation schemes are commonly used to trade bandwidth for better performance.

A set of open standards called DVB (digital video broadcasting) has been developed for digital television, including DVB-S, DVB-S2, and DVB-SH for satellite video broadcasting. However, these DVB standards specify that only point-to-multipoint one-way communication links be used for video broadcasting. With the growing demand for broadband access, two standards have been proposed to support two-way broadband communication links over satellites: DVB-RCS (return channel system) and DOCSIS-S. In DVB-RCS, the forward link (from the service provider to subscribers) is completely compatible with DVB-S. In other words, the forward link can be used for video broadcasting or Internet access. In addition, a return channel is specified for sending user data to the service provider, where ATM-like packets are used for data transmission. DOCSIS-S is an adaption of the DOCSIS standard for satellites. In DOCSIS-S, QPSK or SPSK, combined with turbo coding, is implemented for satellite links, and IP encapsulation is used for data transmission, resulting in more efficient bandwidth utilization and about 10% less overhead than with DVB-RCS.

1.7.5 LMDS and MMDS Systems

Local multipoint distribution service (LMDS), developed by the IEEE 802.16.1 working group, is a last-mile point-to-multipoint wireless access technology. Figure 1.12 shows the network architecture of LMDS systems. In the United States, a 1.3-GHz band between 28 and 31 GHz has been allocated for LMDS, whereas in Europe, LMDS may use different frequency bands from 24 to 43.5 GHz. LMDS can transmit 34 to 38 Mb/s of data covering the range 3 to 5 km. Therefore, multiple cells are

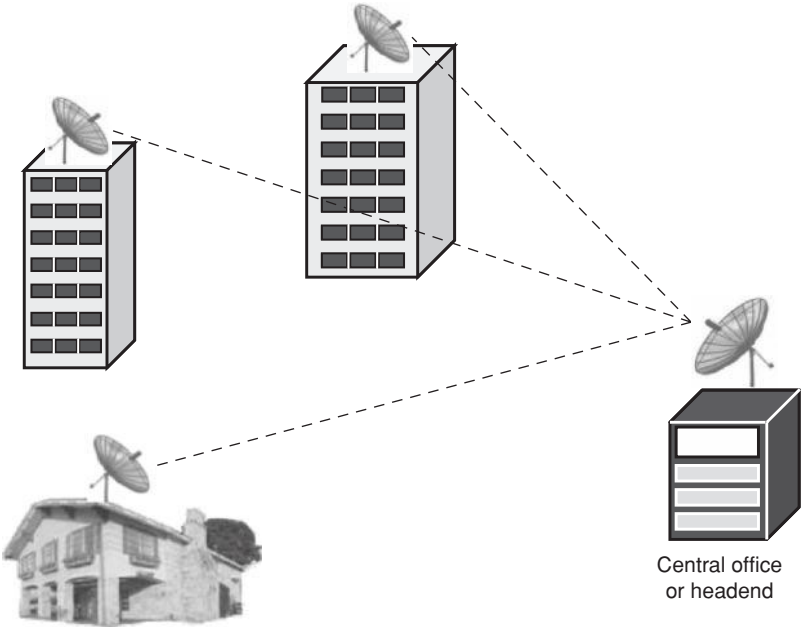


FIGURE 1.12 LMDS architecture.

usually required to serve a metropolitan area. In each cell, a base station with multiple transceivers mounted on the roof of a tall building or on a tall pole transmits users data in a point-to-multipoint mode. A return link from the user to the base station is achieved by a point-to-point link. Even though the physical layer is different from that of the wired cable networks, LMDS has adopted DOCSIS specifications.

Multichannel multipoint distribution service (MMDS), also known as *wireless cable*, was developed in the 1970s as an alternative to cable TV broadcasting. It can support 31 analog channels (6 MHz each) in a 200-MHz frequency band from 2.5 to 2.7 GHz. An MMDS system can also be used as a general-purpose broadband access network. MMDS has been deployed for wireless high-speed Internet access in rural areas where other types of broadband access are unavailable or prohibitively expensive. Figure 1.13 shows the architecture of an MMDS system. In an MMDS system, analog video signals or QAM/OFDM data signals are broadcast from microwave towers at the headend. At the user premises, rooftop antennas pick up the broadcast signal and downconvert it to cable channel frequencies. A gateway device is used to rout various signals to in-home network devices. Overall, the architecture of MMDS resembles that of LMDS. Similarly, MMDS systems have adopted DOCSIS specifications. DOCSIS modified for wireless broadband is commonly referred to as DOCSIS+. MMDS systems can provide a data rate of over 10 Mb/s to a single user. MMDS broadcasts can transmit signal power up to 30 W and cover a diameter of about 50 km, much more than LMDS systems.

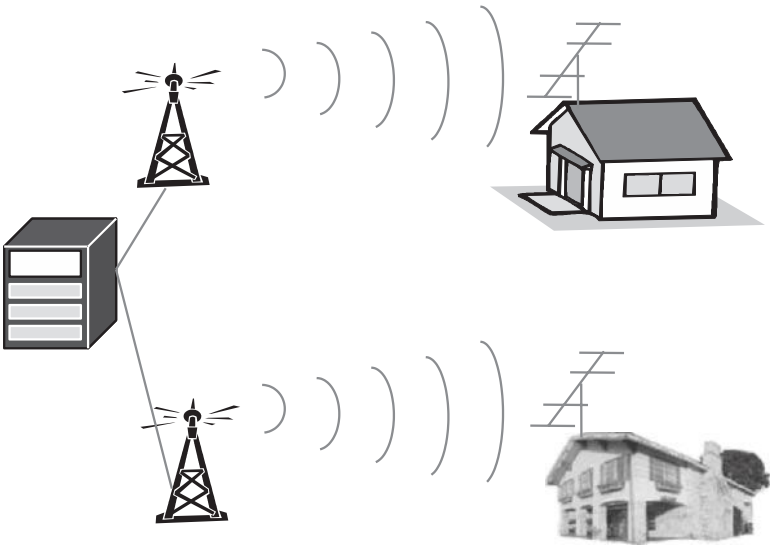


FIGURE 1.13 MMDS architecture.

In the past, even though DSL and cable dominated the broadband access market, LMDS and MMDS showed some promising aspects as wireless solutions. However, for both technical and marketing reasons, LMDS and MMDS systems were never widely adopted for broadband access. Now as WiMAX standards are developing, LMDS and MMDS are surpassed by WiMAX in both technical merit and market potential. Therefore, LMDS and MMDS may become obsolete in the near future.

1.8 BROADBAND SERVICES AND EMERGING TECHNOLOGIES

Broadband access technologies have shown explosive growth in large-scale deployment during the past decade. As of 2007, there are over 300 millions of broadband subscribers worldwide. In the United States alone, broadband Internet access has penetrated over half of U.S. households, reaching 66 million subscribers in 2007. The number of broadband subscribers will continue to grow in the years to come.

Today's broadband applications are mostly driven by Internet users—hundreds of millions of end users generating terabits per second of Internet traffic, and others by the entertainment industry—online or broadcasting video and music consuming a large portion of Internet traffic. The huge bandwidth demands impose a great pressure on broadband access, the technology bridging the gap between home and Internet backbone. Although it presents a great technical challenge, broadband access will

lead to a great opportunity to develop new applications and services. In previous sections we have presented various broadband access technologies. Among these, DSL and cable are the dominant wireline access networks, and cellular networks and Wi-Fi hot spots represent the most widely deployed wireless access technologies. On the other hand, two versions of PON networks, GPON and EPON, become the most promising solution for future broadband services. In this section we review current market demands and driving forces for broadband access, discuss challenging issues in broadband access services, and present possible solutions for future broadband access technologies.

1.8.1 Broadband Access Services

The existing pressure from broadband services has led to heated competitions in access technologies and may shatter the current landscape of the telecom industry. There are a few driving forces behind this huge wave of broadband deployment. Multimedia applications, user bandwidth demands, industrial competitions, economical factors, and government regulation all play very important roles in the march toward "broadband for all" society. Multimedia applications create huge market demands for broadband access. Video services such as IP TV, video on demand, and videoconferencing, particularly, have become killer applications for bandwidth explosion in access networks. In addition to HD and standard TV broadcasting, video has consumed more than 30% of current Internet traffic, and this percentage keeps increasing. It is predicted that video traffic will consist of more than 50% of total Internet traffic. Because of peer-to-peer Web traffic (including video sharing) and other bandwidth-hungry applications (e.g., interactive games), user bandwidth demands are increasing rapidly, rendering broadband DSL and cable Internet access "slow speed." To meet user bandwidth demands and to support multimedia applications, both telecom and cable industries are deploying next-generation broadband technologies, including passive optical networks. Government deregulation, particularly local loop unbundling, has created heated competition in the access segment between telecom operators and MSOs and between ILECs (incumbent local exchange carriers) and IXCs (interexchange carriers). In addition, economical reasons for reducing OPEX (operating expenses) and increasing revenue create a big incentive for large-scale deployment of passive optical networks. Driving by market demands and technical innovations, broadband access networks will continue to evolve in the next few years.

Even though DSL and cable access have come to dominate broadband access in many countries, the broadband service currently offered by service providers is just an extension of existing technologies that provide data service. Telecom operators developed DSL to offer data service over phone lines, and MSOs added bidirectional transmission in HFC networks to support data transmission. *Triple play* has become a buzzword for service providers, but it is very difficult for DSL and cable access to offer triple play, due to their limited bandwidth. In addition to bandwidth, next-generation broadband access requires much more.

As communication networks is evolving toward anywhere, anytime, and any medium communications, future broadband subscribers may requires integrated access services over a unified interface with good end-to-end quality of service at a low service fee. Integrated services over the broadband access networks must be able to provide triple play or even quadruple play (voice, data, video, and mobility). Bandwidth over 100 Mb/s per user might be necessary to support triple play. Furthermore, good end-to-end quality of service is essential for many applications, and the broadband service must be dependable and available all the time. For real-time voice and video, there are tight constraints in delay and jitter. It might not be possible for a single technology to meet these requirements, but future broadband services must be available through an intelligent interface that is transparent to subscribers. No matter what service a subscriber might need, a single user interface will provide it with good QoS support. Finally, the access segment is always cost sensitive; service providers must be able to provide broadband for all service at a price comparable to that of current DSL or cable access.

1.8.2 Emerging Technologies

As discussed earlier, fiber is the only medium that can provide unlimited bandwidth for broadband access services. In the past, economics and lack of killer applications hindered its deployment in the access segment. As the optical technologies become mature, optical components are much less expensive and the fiber deployment cost continues to drop. It is now economically feasible to massively deploy passive optical networks. In the meantime, killer applications such as video on demand have emerged, demanding broadband access service that can be only supported by optical fibers in terms of bandwidth and reach. Since IP over WDM optical networks has been widely deployed in both WAN and MAN, it is not only reasonable but also necessary to deploy optical fibers to the user premises. In fact, different flavors of TDM PONs (mostly BPON and EPON) have been deployed on a large scale. Curently, Japan and Korea have taken the lead in FTTx deployment, and fibers have reached a large percentage of households, serving millions of users. In Japan, there are about 26 million broadband subscribers. Among these, 33% were using FTTH connections in 2007. Even though the majority of subscribers (about 50%) continue to use DSL, the market share of DSL has begun to shrink and FTTH continues to grow. In the first quarter of 2007, FTTH subscribers increased by 860,000 while DSL lost 220,000 users. In the United States, rapid deployment of passive optical networks began in 2005. In 2007, 2 million homes in the United States had fiber connections. FTTH networks will continue to grow all over the world. In Chapter 2 the fundamental of optical communications and the physical technologies for passive optical networks are discussed in detail. Then in Chapter 3, current TDM PON standards are reviewed and compared.

Next-generation optical access networks will evolve to higher bit rates and multiple wavelengths. Currently, 10-Gb/s PONs are being discussed by standard bodies (IEEE, FSAN, and ITU-T). 10-Gb/s downstream deployment include upstream bit rates of 1.25 or 2.5 Gb/s, and symmetric 10-Gb/s PONs will enter the market in the next few

years. Eventually, optical access networks will adopt WDM technologies. The advantage of WDM PONs are higher bandwidth, flexible data format, and better security. However, point-to-point WDM access is relatively expensive despite the quick drop in the cost of optical components. Furthermore, a few issues, including protection and restoration and colorless ONUs, need to be addressed before WDM PONs become wide available. In the short term, hybrid TDM/WDM can provide an evolutionary approach to upgrade TDM to WDM PONs in a scalable and cost-effective manner. Next-generation optical access networks—higher-bit-rate, multiple-wavelength PONs—are the focus of Chapter 4.

Even though passive optical networks can satisfy any user bandwidth demands for triple play (voice, data, and video), their fixed infrastructure and limited coverage cannot fulfill the requirement of ubiquitous and flexible access for emerging multimedia applications. Due to recent advances in wireless technologies, wireless access networks such as Wi-Fi (IEEE 802.11) and WiMAX (IEEE 802.16) become a promising solution to serve the growing number of wireless subscribers interested in high-quality video streams and other multi-media applications using handheld mobile devices. In the future, convergence of optical and wireless technologies is inevitable in the access segment for quadruple play (voice, data, video, and mobility). However, as the traffic behavior and channel quality of these two technologies are far from each other, seamlessly integrating passive optical networks and wireless mesh networks present a very challenging task that demands further investigation. In Chapter 5 we present the challenging issues and possible solutions for hybrid optical and wireless networks.

1.9 SUMMARY

In this chapter we provide a brief overview of the architecture of communication networks and show that current Internet bottlenecks are due to the lack of high-speed access technologies. Then various broadband access technologies are discussed and the major features of DSL, HFC, PON, BPL, Wi-Fi, WiMAX, celluar and satellite networks, and LMDS and MMDS are presented in detail.

DSL can offer data rates over 10 Mb/s within a short distance. Efforts to develop next-generation DSL focus on increasing data rates and transmission distance. With DSL, voice and data can be supported by a single phone line. In the past decade, DSL has become one of the dominant broadband access technologies worldwide. However, TV broadcasting and IP TV are still a technical challenge for DSL networks. Using VDSL and advanced data-compressing techniques, video over IP may be offered in the near future.

Traditionally, HFC networks offer analog TV broadcasting. With the development of cable modem, broadband Internet access can be provided to subscribers. Cable is currently the archrival of DSL, claiming a large portion of market share. However, the bandwidth of coax cable access is still limited to about 10 Mb/s because of hundreds of users sharing the same cable. Currently, MSOs have added VoIP and digital TV over HFC networks. Further development will extend the cable plant beyond 1 GHz, and higher data rates will be the main focus of HFC networks.

BPL is developed as an alternative for DSL and HFC networks. The data rate provided by BPL can reach only a few megabits per second. As power lines reach more residences than any other wired medium, BPL is considered a feasible access technology for rural areas that have no DSL and HFC access. However, technical problems such as noise and interference have hindered large-scale development of BPL.

As optical fibers can provide essentially unlimited bandwidth, PONs are considered the most promising wired access technologies for the future. Current TDM PONs support data rates of tens of megabits per second for a single user, and large-scale deployment of TDM PONs has begun in Asia and North America. As the user bandwidth demands are ever increasing, WDM PONs will be developed as the ultimate solution for broadband access and satisfy the bandwidth requirements of any broadband access services. However, the high deployment cost make them presently a less attractive solution. Current efforts on PON development include bit-rate enhancement and service overlay on hybrid TDM/WDM PONs.

In addition to wired broadband solutions, many wireless technologies have been developed to provide broadband Internet access, such as free-space optical communications, Wi-Fi, WiMAX, and cellular and satellite networks. Free-space optical communications can support gigabit per second data rates and a transmission distance of a few kilometers, but the atmospheric effects impose severe constraints for free-space optical communications. Wi-Fi is widely used for wireless local area networks with transmission distances up to a few hundred meters. With a mesh topology, Wi-Fi networks can support extended reach and broadband Internet access. WiMAX can support wireless access over 10 km, but it requires higher transmitted power and the data rate is lower than in Wi-Fi networks. Cellular networks are used primarily for mobile voice communication. With digital encoding technologies, data transmission service can be added over cellular networks but with a very limited data rate (up to a few Mb/s). Further development of 3G and 4G is expected to support much higher data rates over 10 Mb/s. Satellite networks are used primarily for direct video distribution, but data service over satellite can cover a large geographical area. The main disadvantage of satellite communication is that of very limited data rates. The advantages of wireless technologies are mobility, scalability, low cost, and ease of deployment. Except for free-space optical communication, other wireless technologies uses RF or microwave frequencies. The bit rate–distance product of RF technologies is very limited, and the network capacity and reliability are also much lower than these of wired access networks. In the future, optical and wireless technology may be combined in hybrid optical and wireless access networks, leveraging their complementary characteristics to provide quadruple-play service.

In summary, emerging multimedia applications demand broadband access networks, and a number of wired and wireless broadband access technologies have been developed over the past decade. The long-term perspective of broadband access technology lies in the convergence of optical and wireless technologies. By solving the last-mile bottleneck problem, a variety of new applications will be made possible. One day the dream of broadband access networks for anyone, anywhere, anytime, any medium communications will become a reality.

REFERENCES

1. DSL standards: ITU-T G.992.1/G.992.2 for ADSL, G.992.3/G.992.4 for ADSL2, G.993.1 for VDSL, G.993.2 for VDSL2, G.991.1 for HDSL, and 991.2 for SHDSL.
2. DOCSIS standards: ITU-T J.112 for DOCSIS 1.0, J.122 for DOCSIS 2.0, and J.222 for DOCSIS 3.0.
3. PON standards: ITU-T G.983 for APON and BPON, G.984 for GPON, IEEE 802.3ah for EPON, and 802.3av for 10GEPON.
4. BPL standards: X10 and IEEE P1675/1775/1901.
5. Wi-Fi standard: IEEE 802.11.
6. WiMAX standard: IEEE 802.16.

CHAPTER 2

OPTICAL COMMUNICATIONS: COMPONENTS AND SYSTEMS

Optical communications make use of light waves, very high frequency (100 terahertz) electromagnetic waves, for information transmission. Modern optical communications were begun in the 1960s, when lasers were invented as a coherent light source. Since then, the rapid development of photonic technologies has made possible optical communication links with a capacity of terabits per second and a transmission distance of many thousands of kilometers. The explosive growth of optical communication technology in the past decades has revolutionized the telecom industry and created a global communication infrastructure with optical networks.

A typical fiber optic communication system consists of an optical transmitter, optical fiber, and an optical receiver. The optical transmitter converts the information-carrying electronic signal to an optical signal, which are then sent through a long length of optical fiber. At the receiver end, an optical detector converts the optical signal back to an electronic signal so that the information is recovered and delivered to the destination. In this chapter we focus on the fundamentals of optical communication technologies. The emphasis is on the enabling technologies and physical-layer design for passive optical networks. First, key components of optical communication systems are discussed, including the main characteristics and performance features of optical fibers, transmitters, receivers, amplifiers, and various passive components. Then optical link design and system performance of optical communications are presented. At the end of this chapter, we discuss burst mode transmission, a unique optical transmission technique used in passive optical networks, and its related technologies.

Broadband Optical Access Networks, First Edition. Leonid G. Kazovsky, Ning Cheng, Wei-Tao Shaw, David Gutierrez, and Shing-Wa Wong.

2.1 OPTICAL FIBERS

Optical fiber is a cylindrical waveguide made of dielectric materials such as glass or plastic. Because of its waveguide structure, optical fiber confines a light wave in its core and guides optical signals along its axis. Because of its high bandwidth and low attenuation, optical fiber is widely used as the transmission channel for optical communication systems, carrying high-bit-rate optical signals over long distance. When doped with rare earth elements, optical fiber can serve as an optical amplifier, boosting optical signal power for long-haul transmission. In addition to telecommunications, optical fiber is also used in illumination, imaging, and sensor applications. In this section we focus on its principles and characteristics as an optical transmission medium.

2.1.1 Fiber Structure

Optical fiber used in optical communications consists of a cylindrical dielectric core surrounded by dielectric cladding. A polymer buffer coating is commonly used to enhance its mechanical strength and protect it from environmental effects. Figure 2.1 shows the structure of an optical fiber. Both the core and the cladding are made of silica (SiO_2) or plastics (e.g., PMMA). To guide light waves along the fiber, the refractive index of the core, n_1, must be larger than the refractive index of the cladding, n_2. Typically, the refractive index difference is very small (only a few percent), depending on the desired characteristics of the fiber. There are two different types of fibers: single mode and multimode. Multimode fibers have either 50-, 62.5-, and 85-μm core diameters and a 125-μm cladding diameter. The large diameter of the fiber core facilitates optical power coupling in and out of fiber. Single-mode fibers, on the other hand, have a much smaller core diameter, typically 5 to 8 μm.

2.1.1.1 Multimode Fibers Multimode fiber is used for short-distance transmission, up to a few kilometers. Multimode fibers are commonly deployed for local area networks (e.g., gigabit Ethernet) used in offices, buildings, medical facilities, or campus complexes. Because of its large core diameter, multimode fiber can carry many light rays simultaneously, each propagating at a different angle. Depending on the refractive index profile, multimode fiber can be categorized as step- or graded-index fiber.

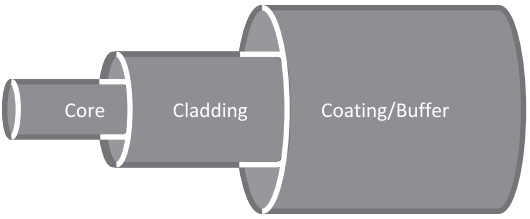


FIGURE 2.1 Optical fiber structure.

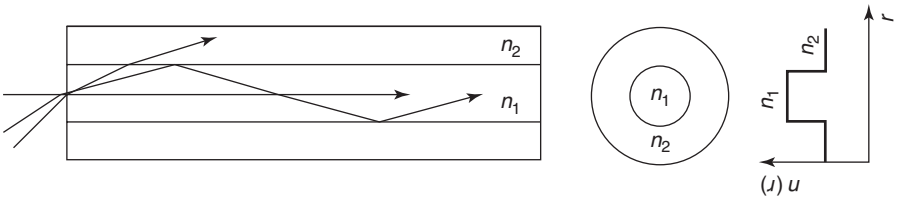


FIGURE 2.2 Step-index fiber.

Step-Index Fibers The refractive index profile of a step-index fiber is shown in Figure 2.2 Light transmission in step-index multimode fibers is based on total internal reflection. Since the core index is larger than that of the cladding, total internal reflection occurs when light rays are incident to the core–cladding interface with an incident angle larger than the critical angle θ_c. The critical angle for total internal reflection is given by $\sin\theta_c = n_2/n_1$. Because of total internal reflection, light rays can propagate along the fiber core in a zigzag way if the fiber is straight without bending. Alternatively, light rays can travel in a straight line parallel to the fiber axis. Since light rays travels in different paths along the fiber, they may experience different propagation delays. Suppose that all the light rays are coincident at the input end of the multimode fiber. As they propagate through the fiber, some travel in a straight line while others may travel in zigzag paths. These rays disperse in time at the output of the fiber. The shortest ray path is a straight line equal to the fiber length L, while the longest path occurs for rays with an incident angle at critical angle θ_c and is of length $L/\sin\theta_c$. Therefore, the propagation delay between two rays taking the shortest and longest paths is

$$\Delta T = \frac{L}{v}\left(\frac{1}{\sin\theta_c} - 1\right) = \frac{L}{v}\frac{n_1}{n_2}\Delta, \tag{2.1}$$

where $\Delta = (n_1 - n_2)/n_1$ and v is the group velocity of light rays in the fiber.

Since light rays traveling with different paths are referred to as different fiber modes (the concept of fiber mode is discussed further in Section 2.1.2), the dispersion in time, that is, the difference in the propagation delay for different light rays, is called *modal dispersion*. Because of modal dispersion, a short optical pulse will be broadened as it propagates in a multimode fiber. The pulse broadening will introduce intersymbol interference, and hence limit the bit rate of optical communication systems. In multimode fibers, modal dispersion is one of the most important limiting factors for high-speed optical communications. Step-index multimode fibers are usually designed with a very small index difference. With $\Delta = 2 \times 10^{-3}$, a step-index multimode fiber can support optical transmission with a 100-Mb/s data rate over a 1-km distance. Hence, step-index multimode fibers are used only for low-data-rate applications over short distances.

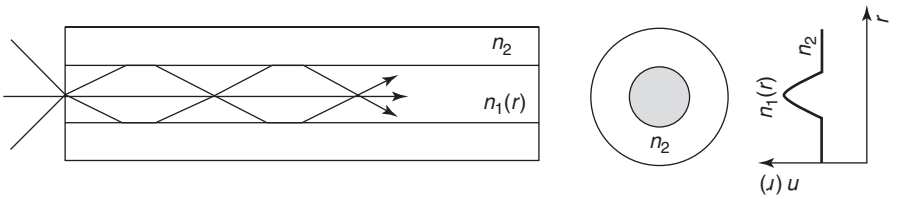


FIGURE 2.3 Graded-index fiber.

Graded-Index Fibers In graded-index fibers, the refractive index of the core gradually decreases from the center of the core to the cladding, as shown in Figure 2.3. When a light ray propagates in a straight line in the center of the core, it takes the shortest path but has the lowest group velocity. Oblique rays will follow serpentine traces due to the graded-index profile. A large portion of their paths has a smaller refractive index, and hence they have a larger group velocity. Therefore, in a graded-index fiber with a suitable index profile, all light rays will arrive at the fiber output at the same time. Detailed analysis found that minimum modal dispersion is

$$\Delta T = \frac{1}{8}\frac{n_1 L}{c}\Delta^2, \tag{2.2}$$

where c is the speed of light in vacuum. This minimum modal dispersion can be achieved with an index profile given by

$$n(r) = n_1\left[1 - \Delta\,(r/a)^{\alpha}\right], \tag{2.3}$$

where $\alpha = 2\,(1 - \Delta)$. Since Δ is very small, the optimum index profile is approximately parabolic. In practice, graded-index multimode fibers can support data rates at 1 Gb/s over 10 km. To further improve the data rates and transmission distance, single-mode fibers have to be deployed.

2.1.1.2 Single-Mode Fibers Single-mode fiber has a very small core diameter, comparable to the wavelength of an optical signal, and only one fundamental fiber mode exists in single-mode fibers. As a consequence, there is no modal dispersion in single-mode fibers, and very high data rates (e.g., 10 Gb/s) can be supported by single-mode fibers over a distance of tens of kilometers. Single-mode fibers are used widely in telecommunication applications, creating modern optical networks with a total capacity above 10 Tb/s. However, there are disadvantages to using single-mode fiber rather than multimode fiber. The small core diameter of single-mode fiber makes it very difficult to couple light in and out of a single-mode fiber, and it puts more constraints on the tolerance of connectors and splices used to connect single-mode fibers.

Due to their superior performance, single-mode fibers are widely deployed for wide area networks and metropolitan area networks. Millions of miles of single-mode fibers have been laid worldwide in the ground, under the sea, or over the air. Passive optical networks also use single-mode fibers in their fiber plants to support high data rates ($\geq$ 1 Gb/s) and long-distance transmission (up to 20 km). In the remainder of the chapter we focus on the properties and transmission characteristics of single-mode fibers.

2.1.2 Fiber Mode

Fiber mode is the optical field distribution in an optical fiber that satisfies certain boundary conditions imposed by the physical structure of the optical fiber. This distribution does not change with optical signal propagation along the optical fiber (except as an attenuation factor). Like all other electromagnetic phenomena, the optical field distribution in an optical fiber is governed by Maxwell's equations. For homogeneous dielectric media (e.g., optical fiber), Maxwell's equations can be written as

$$\nabla \times \mathbf{E}(\mathbf{r}, t) = -\mu \frac{\partial \mathbf{H}(\mathbf{r}, t)}{\partial t}, \tag{2.4}$$

$$\nabla \times \mathbf{H}(\mathbf{r}, t) = \epsilon \frac{\partial \mathbf{E}(\mathbf{r}, t)}{\partial t}, \tag{2.5}$$

$$\epsilon \nabla \cdot \mathbf{E}(\mathbf{r}, t) = 0, \tag{2.6}$$

$$\mu \nabla \cdot \mathbf{H}(\mathbf{r}, t) = 0, \tag{2.7}$$

where $\mathbf{E}$ and $\mathbf{H}$ are the electric and magnetic field vectors, respectively, and ϵ and μ are the permittivity and permeability of the dielectric material.

Assume that a monochromatic light wave propagates along an optical fiber in the $+z$ direction. The electric and magnetic fields of the light wave are given by

$$\mathbf{E}(\mathbf{r}, t) = \left[\hat{x} E_x(x, y) + \hat{y} E_y(x, y) + \hat{z} E_z(x, y)\right] e^{-\gamma z} e^{j\omega t}, \tag{2.8}$$

$$\mathbf{H}(\mathbf{r}, t) = \left[\hat{x} H_x(x, y) + \hat{y} H_y(x, y) + \hat{z} H_z(x, y)\right] e^{-\gamma z} e^{j\omega t}, \tag{2.9}$$

where E_x, E_y, and E_z are the components of the electric field distribution, and H_x, H_y, and H_z are the components of the electric field distribution in the cross section of the optical fiber. $(\hat{x})$, $(\hat{y})$, and $(\hat{z})$ are the unit vectors in the x, y, and z directions, respectively. γ is the propagation constant of the light wave, given by

$$\gamma = \alpha + j\beta, \tag{2.10}$$

where α is the field attenuation coefficient and β is the phase propagation constant. Substituting eqs. (2.8) and (2.9) into Maxwell's equations yields two wave propagation equations,

$$\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + \kappa^2 E_z = 0, \tag{2.11}$$

$$\frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} + \kappa^2 H_z = 0, \tag{2.12}$$

where $\kappa^2 = \gamma^2 + \omega^2 \mu\epsilon$. The other four field components can be found by

$$E_x = -\frac{\gamma}{\kappa^2}\frac{\partial E_z}{\partial x} - \frac{j\omega\mu}{\kappa^2}\frac{\partial H_z}{\partial y}, \tag{2.13}$$

$$E_y = -\frac{\gamma}{\kappa^2}\frac{\partial E_z}{\partial y} + \frac{j\omega\mu}{\kappa^2}\frac{\partial H_z}{\partial x}, \tag{2.14}$$

$$H_z = \frac{j\omega\mu}{\kappa^2}\frac{\partial E_z}{\partial y} - \frac{\gamma}{\kappa^2}\frac{\partial H_z}{\partial x}, \tag{2.15}$$

$$H_y = -\frac{j\omega\mu}{\kappa^2}\frac{\partial E_z}{\partial x} - \frac{\gamma}{\kappa^2}\frac{\partial H_z}{\partial y}. \tag{2.16}$$

The equations above are written in rectangular coordinates. As an optical fiber is cylindrical, it is more convenient to work with cylindrical coordinates. In a cylindrical coordinates, the field components in the cross section of the optical fiber can be written

$$E_r = E_x \cos\phi + E_y \sin\phi,$$

$$E_\phi = -E_x \sin\phi + E_y \cos\phi,$$

$$H_r = H_x \cos\phi + H_y \sin\phi,$$

$$H_\phi = -H_x \sin\phi + H_y \cos\phi.$$

The wave propagation equations, eqs. (2.11) and (2.12), written in coordinates, become

$$\frac{\partial^2 E_z}{\partial r^2} + \frac{1}{r}\frac{\partial E_z}{\partial r} + \frac{1}{r^2}\frac{\partial^2 E_z}{\partial \phi^2} + \kappa^2 E_z = 0, \tag{2.17}$$

$$\frac{\partial^2 H_z}{\partial r^2} + \frac{1}{r}\frac{\partial H_z}{\partial r} + \frac{1}{r^2}\frac{\partial^2 H_z}{\partial \phi^2} + \kappa^2 H_z = 0. \tag{2.18}$$

Solving the wave propagation, one can find the field distribution in the core and cladding of an optical fiber.

In the core, $r < a$, the field components are

$$E_z(r,\phi) = AJ_\nu(pr)\,e^{j\nu\phi}, \tag{2.19}$$

$$H_z(r,\phi) = BJ_\nu(pr)\,e^{j\nu\phi}, \tag{2.20}$$

$$E_r = -\frac{j}{p^2}\left(A\beta pJ'_\nu(pr) + \frac{jB\nu\omega\mu_1}{r}J_\nu(pr)\right)e^{j\nu\phi}, \tag{2.21}$$

$$E_\phi = -\frac{j}{p^2}\left(\frac{jA\nu\beta}{r}J_\nu(pr) - Bp\omega\mu_1 J'_\nu(pr)\right)e^{j\nu\phi}, \tag{2.22}$$

$$H_r = -\frac{j}{p^2}\left(Bp\beta J'_\nu(pr) - \frac{jA\nu\omega\epsilon_1}{r}J_\nu(pr)\right)e^{j\nu\phi}, \tag{2.23}$$

$$H_\phi = -\frac{j}{p^2}\left(\frac{jB\nu\beta}{r}J_\nu(pr) + Ap\omega\epsilon_1 J'_\nu(pr)\right)e^{j\nu\phi}, \tag{2.24}$$

and in the cladding, $r > a$, the field components are

$$E_z(r,\phi) = CK_\nu(qr)\,e^{j\nu\phi}, \tag{2.25}$$

$$H_z(r,\phi) = DK_\nu(qr)\,e^{j\nu\phi}, \tag{2.26}$$

$$E_r = -\frac{j}{q^2}\left(C\beta qK'_\nu(qr) + \frac{jD\nu\omega\mu_2}{r}K_\nu(qr)\right)e^{j\nu\phi}, \tag{2.27}$$

$$E_\phi = -\frac{j}{q^2}\left(\frac{jC\nu\beta}{r}K_\nu(qr) - Dq\omega\mu_2 K'_\nu(qr)\right)e^{j\nu\phi}, \tag{2.28}$$

$$H_r = -\frac{j}{q^2}\left(Dq\beta K'_\nu(qr) - \frac{jC\nu\omega\epsilon_2}{r}K_\nu(qr)\right)e^{j\nu\phi}, \tag{2.29}$$

$$H_\phi = -\frac{j}{q^2}\left(\frac{jD\nu\beta}{r}K_\nu(qr) + Cq\omega\epsilon_2 K'_\nu(qr)\right)e^{j\nu\phi}, \tag{2.30}$$

where ν is an integer, J_ν a Bessel function of the first kind of order ν, and K_ν a modified Bessel function of the second kind of order ν. J'_ν and K'_ν represent, respectively, the derivatives of J_ν and K_ν. ϵ_1 and ϵ_2 are the permittivities of the core and cladding, respectively. μ_1 and μ_2 are the permeabilities of the core and cladding,

respectively. The parameters p and q are given by

$$p^2 = n_1^2 k_0^2 - \beta^2, \tag{2.31a}$$

$$q^2 = \beta^2 - n_2^2 k_0^2. \tag{2.31b}$$

The wave number k_0 is defined as $k_0 = 2\pi/\lambda$, and the propagation constant β satisfies $n_2 k_0 < \beta < n_1 k_0$. The constants β, A, B, C, and D can be determined by applying the boundary conditions for the two tangential components of the electric and magnetic fields at the core–cladding interface, $r = a$. The following four equations in unknowns A, B, C, D, and β are derived from the boundary conditions:

$$\begin{bmatrix} J_\nu(pa) & 0 & -K_\nu(qa) & 0 \\ \dfrac{\beta\nu}{p^2 a} J_\nu(pa) & \dfrac{j\omega\mu_1}{p} J'_\nu(pa) & \dfrac{\beta\nu}{q^2 a} K_\nu(qa) & \dfrac{j\omega\mu_2}{q} K'_\nu(qa) \\ 0 & J_\nu(pa) & 0 & -K_\nu(qa) \\ -\dfrac{j\omega\epsilon_1}{p} J'_\nu(pa) & \dfrac{\beta\nu}{p^2 a} J_\nu(pa) & -\dfrac{-j\omega\epsilon_2}{q} K'_\nu(qa) & \dfrac{\beta\nu}{q^2 a} K_\nu(qa) \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

These homogeneous equations have a nontrivial solution if the determinant of the coefficient matrix vanishes. Thus we have the characteristic equation

$$\left(\frac{J'_\nu(pa)}{pJ_\nu(pa)} + \frac{K'_\nu(qa)}{qJ_\nu(qa)} \right) \left(\frac{n_1^2 J'_\nu(pa)}{pJ_\nu(pa)} + \frac{n_2^2 K'_\nu(qa)}{qJ_\nu(qa)} \right) k_0^2 = \left(\frac{\beta\nu}{a} \right)^2 \left(\frac{1}{p^2} + \frac{1}{q^2} \right)^2. \tag{2.32}$$

For a given ν and ω only a finite number of values β can be found which satisfy the characteristic equation and $n_2 k_0 < \beta < n_1 k_0$. A fiber mode is determined uniquely by its propagation constant β. Having found β, we then have

$$\frac{B}{A} = \frac{j\beta\nu}{\omega\mu} \left[\frac{1}{(pa)^2} + \frac{1}{(qa)^2} \right] \left[\frac{J'_\nu(pa)}{paJ_\nu(pa)} + \frac{K'_\nu(qa)}{qaK_\nu(qa)} \right],$$

$$\frac{C}{A} = \frac{J_\nu(pa)}{K_\nu(qa)}, \qquad \frac{D}{A} = \frac{C}{A}\frac{B}{A}.$$

The only undetermined coefficient A can be found by the power constraint of the light wave.

Once the propagation constant β and field amplitudes are known, the electromagnetic field inside the fiber core is

$$E_z(r,\phi,z,t) = AJ_\nu(pr)\,e^{j\nu\phi}e^{-\alpha z}e^{j(\omega t-\beta z)}, \tag{2.33}$$

$$E_r = -\frac{j}{p^2}\left(A\beta pJ'_\nu(pr) + \frac{jB\nu\omega\mu_1}{r}J_\nu(pr)\right)e^{j\nu\phi}e^{-\alpha z}e^{j(\omega t-\beta z)}, \tag{2.34}$$

$$E_\phi = -\frac{j}{p^2}\left(\frac{jA\nu\beta}{r}J_\nu(pr) - Bp\omega\mu_1 J'_\nu(pr)\right)e^{j\nu\phi}e^{-\alpha z}e^{j(\omega t-\beta z)}, \tag{2.35}$$

$$H_z(r,\phi,z,t) = BJ_\nu(pr)\,e^{j\nu\phi}e^{-\alpha z}e^{j(\omega t-\beta z)}, \tag{2.36}$$

$$H_r = -\frac{j}{p^2}\left(Bp\beta J'_\nu(pr) - \frac{jA\nu\omega\epsilon_1}{r}J_\nu(pr)\right)e^{j\nu\phi}e^{-\alpha z}e^{j(\omega t-\beta z)}, \tag{2.37}$$

$$H_\phi = -\frac{j}{p^2}\left(\frac{jB\nu\beta}{r}J_\nu(pr) + Ap\omega\epsilon_1 J'_\nu(pr)\right)e^{j\nu\phi}e^{-\alpha z}e^{j(\omega t-\beta z)}; \tag{2.38}$$

and in the cladding

$$E_z(r,\phi,z,t) = CJ_\nu(qr)\,e^{j\nu\phi}e^{-\alpha z}e^{j(\omega t-\beta z)}, \tag{2.39}$$

$$E_r = -\frac{j}{q^2}\left(C\beta qK'_\nu(qr) + \frac{jD\nu\omega\mu_2}{r}K_\nu(qr)\right)e^{j\nu\phi}e^{-\alpha z}e^{j(\omega t-\beta z)}, \tag{2.40}$$

$$E_\phi = -\frac{j}{q^2}\left(\frac{jC\nu\beta}{r}K_\nu(qr) - Dq\omega\mu_2 K'_\nu(qr)\right)e^{j\nu\phi}e^{-\alpha z}e^{j(\omega t-\beta z)}, \tag{2.41}$$

$$H_z(r,\phi,z,t) = DJ_\nu(qr)\,e^{j\nu\phi}e^{-\alpha z}e^{j(\omega t-\beta z)}, \tag{2.42}$$

$$H_r = -\frac{j}{q^2}\left(Dq\beta K'_\nu(qr) - \frac{jC\nu\omega\epsilon_2}{r}K_\nu(qr)\right)e^{j\nu\phi}e^{-\alpha z}e^{j(\omega t-\beta z)}, \tag{2.43}$$

$$H_\phi = -\frac{j}{q^2}\left(\frac{jD\nu\beta}{r}K_\nu(qr) + Cq\omega\epsilon_2 K'_\nu(qr)\right)e^{j\nu\phi}e^{-\alpha z}e^{j(\omega t-\beta z)}. \tag{2.44}$$

Except for an attenuation factor $\exp(-\alpha z)$ and a phase factor $\exp[j(\omega t - \beta z)]$, the field distribution in an optical fiber does not change when the monochromatic light wave propagates along the fiber. However, the field distribution depends slightly on the frequency (wavelength) of the light wave. The field distribution in the core is described by a Bessel function of the first kind, J_ν, a sinusoidal-like function. On the other hand, the field distribution in the cladding is described by a modified Bessel function K_ν, which is an exponential-like decaying function as the radius r becomes large. Therefore, the optical field is mostly confined in the fiber core and decays in the cladding. Because of this confinement, the energy of an optical signal is mostly

concentrated in the fiber core, and the optical fiber becomes an effective waveguide at optical frequencies. Optical signal can propagate along an optical fiber over a long distance with small attenuation.

Solving the characteristic equation (2.32), we find that two sets of modes exist in the fiber, whose characteristic equations are, for HE modes,

$$\frac{J_{\nu-1}(pa)}{paJ_\nu(pa)} = -\frac{n_1^2+n_2^2}{2n_1^2}\frac{K'_\nu(qa)}{qaK_\nu(qa)} + \frac{\nu}{pa} + R, \tag{2.45}$$

and for EH modes,

$$\frac{J_{\nu+1}(pa)}{paJ_\nu(pa)} = \frac{n_1^2+n_2^2}{2n_1^2}\frac{K'_\nu(qa)}{qaK_\nu(qa)} + \frac{\nu}{pa} + R, \tag{2.46}$$

where

$$R = \left[\left(\frac{n_1^2-n_2^2}{2n_1^2}\right)^2\left(\frac{K'_\nu(qa)}{qaK_\nu(qa)}\right)^2 + \left(\frac{\beta\nu}{n_1k_0}\right)^2\left(\frac{1}{p^2a^2}+\frac{1}{q^2a^2}\right)^2\right]^{1/2}.$$

The characteristic equations can be solved numerically to find the propagation constant β. In general, both E_z and H_z are nonzero (except in the case of $\nu = 0$). Therefore, these fiber modes are referred to as *hybrid modes*, and the solutions are denoted by $\mathrm{EH}_{\nu m}$ or $\mathrm{HE}_{\nu m}$, depending on whether E_z or H_z dominates. When $\nu = 0$, there is no ϕ dependence for the field distribution; that is, the field components are radially symmetric. In this case, one of the longitudinal field components (either E_z or H_z) is zero, so we have transverse (TE or TM) modes. For the TE mode, the characteristic equation becomes

$$\frac{J'_0(pa)}{pJ_0(pa)} + \frac{K'_0(qa)}{qK_0(qa)} = 0. \tag{2.47}$$

In this case, $A = C = 0$ (i.e., $E_z = 0$). The only nonvanishing field components are E_r, H_r, and H_z. For the TM mode, the characteristic equation is

$$n_1^2\frac{J'_0(pa)}{pJ_0(pa)} + n_2^2\frac{K'_0(qa)}{qK_0(qa)} = 0. \tag{2.48}$$

Now, $B = D = 0$, and we have TM modes with nonvanishing field components H_r, E_r, and E_z. The characteristic equation can be plotted as a function of pa. Since $n_1 \approx n_2$, the solutions of the characteristic equation for TE0m are very close to those of TM0m.

Single-Mode Condition A fiber mode is referred to as being cut off when it is no longer bound to the fiber core; in other words, the field does not decay in the cladding.

The cutoffs for the various modes are founded by solving the characteristic equation in the limit $q \to 0$. By a proper choice of the core radius a, core refractive index n_1, and cladding refractive index n_2, so that

$$V = k_0 a\sqrt{n_1^2 - n_2^2} \approx \frac{2\pi a}{\lambda} n_1 \sqrt{2\Delta} \leq 2.405, \tag{2.49}$$

TE_0m, TM_0m, and all other higher-order modes are cut off for a light wave with wavelength λ. However, the HE_{11} mode is the fundamental mode of optical fibers and has no cutoff. Hence, single-mode fibers are designed to satisfy the cutoff condition in eq. (2.49) and supports on the HE_{11} mode. Typically, telecommunication fibers operate in single mode for wavelength $\lambda > 1.2$ μm. These fibers are usually designed with a core radius of 3 to 5 μm and $\Delta = 3$ to 5×10^{-3}.

The field distribution for a single-mode fiber is often approximated by a Gaussian distribution

$$E_z(r, \phi) = A \exp\left(-\frac{r^2}{w^2}\right), \tag{2.50}$$

where w is the field radius (sometimes called the *spot size*). $2w$ is commonly known as the *mode field diameter*. For $1.2 < V < 2.4$, the field radius can be approximated by

$$w/a \approx 0.65 + 1.619V^{-3/2} + 2.879V^{-6}. \tag{2.51}$$

Hence, the field radius decreases as V increases. The fraction of optical power contained in the fiber core is referred to as the *confinement factor*. For weakly guiding fiber (small Δ), the confinement factor is given by

$$\Gamma = \frac{P_{\text{core}}}{P_{\text{total}}} = \frac{\int_0^a |E|^2 r\, dr}{\int_0^\infty |E|^2 r\, dr} = \left(1 - \frac{p^2}{V^2}\right)\left(1 - \frac{J_v^2(pa)}{J_{v+1}(pa)\, J_{v-1}(pa)}\right). \tag{2.52}$$

Using Gaussian approximation for the field distribution, the confinement factor can be found as

$$\Gamma = 1 - \exp\left(-\frac{2a^2}{w^2}\right). \tag{2.53}$$

As V increases, the mode field diameter decreases, and the optical power is more confined in the core, providing better guidance for optical signals. However, to maintain single-mode operation, V must be less than 2.405. Therefore, most telecom single-mode fibers are designed with $2 < V < 2.4$.

2.1.3 Fiber Loss

Various loss mechanisms attenuate optical signal propagating along the fiber. Since optical receivers need a minimum amount of optical power to recover the information transmitted, fiber loss limits the maximum distance that an optical signal can be transmitted. Therefore, fiber loss is a fundamental limiting factor in long-haul optical communication systems. Even though modern optical networks use multiple stages of optical amplifiers to compensate for the fiber loss, optical amplifiers add amplified spontaneous noise. There is a limit on the maximum reach that an optical communication system can support no matter how many optical amplifiers are used. In passive optical networks, optical amplifiers are not commonly used, for reasons of economics. Current TDM PONs are essentially power limited, with fiber loss playing a significant role in the total power budget constraints.

When an optical signal propagates along an optical fiber, its power attenuation is described by

$$dP/dz = \alpha_f P, \tag{2.54}$$

where α_f is the power attenuation coefficient of the optical fiber. If at the transmitter side an optical power signal with power P_{in} is launched into an optical fiber of length L, after propagating through an optical fiber, the optical power received at the other end of the fiber is

$$P_{\text{out}} = P_{\text{in}} \exp\left(-\alpha_f z\right). \tag{2.55}$$

Therefore, the optical power is attenuated exponentially in an optical fiber. Modern optical communication only became possible when low-attenuation fibers were fabricated in 1970s.

Figure 2.4 shows the fiber loss of standard single-mode fiber as a function of wavelength. There are three wavelength regions for local loss minima: 0.85, 1.30, and 1.55 μm. Historically, when low-loss optical fibers were developed (but the impurity concentration was still higher than modern fibers), the first window (0.85 μm) was opened, matching the wavelength of AlGaAs semiconductor lasers developed in the earlier years. This window has a loss at about 2 dB/km, which is too high for modern telecommunication applications but is still widely used for short-distance data communications (e.g., gigabit Ethernet). As the fiber fabrication process was improved further in the late 1970s, lower-loss windows become available at 1.30 and 1.55 μm. Because of its low loss (in addition, low dispersion at 1.31 μm), these two windows are preferred for long-reach high-data-rate optical communications. The losses are 0.5 and 0.2 dB/km for wavelengths at 1.30 and 1.55 μm, respectively. In current TDM PON standards, the wavelength at 1.31 μm is used for upstream transmission, while the wavelength at 1.55 μm band (1490 nm) is used for downstream transmission.

Three major factors contribute to the fiber loss spectrum: material absorption, Rayleigh scattering, and waveguide imperfections. In practice, fiber connectors and

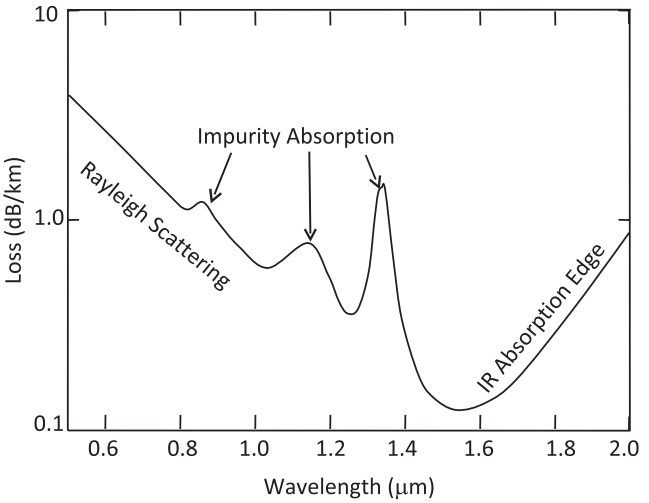


FIGURE 2.4 Optical fiber loss spectrum.

splice and nonlinear effects (e.g., stimulated Raman–Brillouin scattering) also contribute to power loss when an optical signal propagates along the fiber.

Material absorption results from both fused silica (SiO_2 from which fiber is made) and impurities such as water and transition metals (e.g., Fe, Cu, and Ni). In fused silica, electronic resonance creates strong absorption peaks in the ultraviolet region, and molecular vibration absorbs infrared wavelengths. The tails of these resonance peaks extend into the three windows of optical communications. This intrinsic material loss sets the fundamental limit on fiber loss, typically less than 0.03 dB/km from 1.3 to 1.6 μm. Much of the improvement in reducing fiber loss in the early years came from material purification. The transition metals can lead to significant absorption in wavelengths from 0.8 to 1.6 μm. To achieve a loss below 1 dB/km, the concentration of transition metal impurities has to be reduced to 1 part per billion. Residue water vapor or, to be more exact, OH ions have absorption resonance peaks around 1.39, 1.24, and 0.95 μm. Development of a new fiber fabrication process has lead to dry fibers, in which the OH ion absorption peaks are eliminated. This type of fiber opens the entire 1.3 to 1.60-μm wavelength region for high-speed long-haul optical communications. Dopants used to change the refractive index of the fiber core or cladding, such GeO_2, P_2O_5, and B_2O_3, also cause the loss of optical power.

Rayleigh scattering results from the refractive index fluctuation in fiber materials. During the fabrication process, silica preform is melted, drawn into fibers, and then cooled down. The small fluctuations in refractive index are on a scale smaller than the wavelength of the optical signal, but these random changes still lead to light scattering. The Rayleigh scattering is inverse proportional to the optical wavelength:

$$\alpha_R = \frac{C_R}{\lambda^4}, \tag{2.56}$$

where the parameter C_R is about 0.7 to 0.9 dB/km·μm^4. Therefore, Rayleigh scattering is more significant for shorter wavelengths, and it is a dominant factor in single-mode fibers for wavelengths from 0.8 to 1.6 μm.

Ideally, optical fiber is a perfect cylinder with constant core diameter and a smooth core–cladding interface. However, waveguide imperfections such as bending and core radius variation will lead to optical power loss. Coupling between guiding modes and cladding mode occurs when there are refractive index inhomogeneities at the core–cladding interface. Small fiber bends may lead to a part of optical energy being scattered into the cladding layer. To reduce the fiber loss resulting from waveguide imperfection, fiber diameter is monitored closely during fiber drawing to ensure that the variations in fiber diameter are less than 1%. In addition, fiber installation and cabling are designed to minimize the random axial distortions.

2.1.4 Fiber Dispersion

In addition to fiber loss, dispersion of optical pulse is another detrimental effect of optical transmission in optical fibers. Dispersion leads to pulse broadening and intersymbol interference. As a consequence, long-distance high-bit-rate communication systems are limited by fiber dispersion. There are basically three types of dispersion in optical fibers: modal dispersion, intramodal dispersion, and polarization mode dispersion. Modal dispersion, or intermodal dispersion, is the dominant effect when it is present in multimodal fiber. As discussed in Section 2.1.2, each mode in a multimode fiber has a characteristic group velocity and corresponding propagation delay. The modal dispersion is the difference between the longest and shortest propagation delays. In single-mode fiber, modal dispersion vanishes as the energy of optical signals is coupled into a single mode. *Intramodal dispersion*, or *chromatic dispersion*, is the difference in propagation delay for different spectral components of an optical signal. Two sources of chromatic dispersion in single-mode fibers are material dispersion and waveguide dispersion. Polarization mode dispersion is due to birefringence in optical fibers, which leads to different group velocity for two orthogonal polarization modes.

2.1.4.1 Modal Dispersion Modal dispersion is related to the group velocity and propagation delay of different modes. The group velocity of an optical signal is defined as

$$v_g = \frac{d\omega}{d\beta}, \tag{2.57}$$

and the propagation delay in an optical fiber with length L is given by

$$\tau = \frac{L}{v_g} = L\frac{d\beta}{d\omega}. \tag{2.58}$$

For weakly guiding fibers ($\Delta \ll 1$), the propagation constant for a mode denoted by νm can be written as

$$\beta_{\nu m} = \frac{\omega n_2}{c}\left(1 + \Delta b_{\nu m}\right), \tag{2.59}$$

where b is the normalized propagation constant given by

$$b_{\nu m} = \frac{\beta_{\nu m}/k_0 - n_2}{n_1 - n_2}. \tag{2.60}$$

Therefore, the propagation delay for mode νm is

$$\frac{\tau}{L} = \frac{d\beta_{\nu m}}{d\omega} = \frac{1}{c}\left(n_2 + \omega\frac{dn_2}{d\omega}\right)\left(1 + b_{\nu m}\Delta\right) + \frac{\omega n_2}{c}\frac{d\left(b_{\nu m}\Delta\right)}{d\omega}. \tag{2.61}$$

In this equation, the first term on the right-hand side represents material dispersion, and the second term describes the excess delay due to the waveguide. Modal dispersion is part of waveguide dispersion contained in the second term, and it is the dominant waveguide dispersion in multimode fibers. Modal dispersion is determined by the difference in propagation delay between the fastest and slowest modes.

The modal dispersion for an optical fiber with unit length can be rewritten as

$$\tau_w = \frac{n_2\Delta}{c}\frac{d\left(Vb_{\nu m}\right)}{dV} = \frac{n_2\Delta}{c}\left[1 - \frac{p^2}{q^2}\left(1 - \frac{2K_\nu^2\left(qa\right)}{K_{\nu-1}\left(qa\right)K_{\nu+1}\left(qa\right)}\right)\right]. \tag{2.62}$$

Away from cutoff ($\nu \gg 1$), the term $d\left(Vb_{\nu m}\right)/dV$ is approximately $2\left(\nu - 1\right)/\nu$. Obviously, different modes ν give different propagation delays. For $\nu = 0$, 1, and 2, these modes travel fastest, and $d(Vb_{\nu m})/dV \approx 1$; for $\nu = \nu_{\max}$, the mode travels slowest. Therefore, for an optical fiber with unit length, the difference in propagation delays for the fastest and slowest modes is given by

$$\delta\tau_w = \frac{n_1 - n_2}{c}\left(1 - \frac{2}{\nu_{\max}}\right). \tag{2.63}$$

Since the largest mode number that can be supported in an optical fiber is approximately $\nu_{\max} \approx 2V/\pi$, the modal dispersion in a multimode step-index fiber can be expressed as

$$\delta\tau_w = \frac{n_1 - n_2}{c}\left(1 - \frac{\pi}{V}\right). \tag{2.64}$$

Modal dispersion is a multipath phenomenon, independent of the characteristics of the optical signal transmitted. The total modal dispersion is found by multiplying the dispersion per unit length by the total length of the fiber. In practice, when fiber is

longer than a critical length called the *coupling length*, the total modal dispersion increases as the square root of the fiber length. The reason for this reduction in total dispersion is mode coupling. The modal dispersion can be reduced by making Δ small. Small Δ leads to smaller V, reducing the number of modes that can be supported in the fiber. Hence, the longest propagation delay is reduced with essentially no change in the shortest.

2.1.4.2 Intramodal Dispersion Modal dispersion can be eliminated effectively by reducing the number of modes to one. Single-mode fibers carry only the lowest-order mode, the HE_11 mode. The cutoff condition for other modes (e.g., the TE_{01} or TM_{01} mode) is $V < 2.405$. In single-mode fibers, intramodal dispersion becomes significant. Intramodal dispersion results from different frequency components of an optical signal having a different group velocity in an optical fiber. Hence, intramodal dispersion is also called *group velocity dispersion*. It includes mainly material dispersion and waveguide dispersion. Material dispersion results when the dielectric constant and index of refraction vary with frequency (wavelength). Waveguide dispersion is the result of the waveguide structure of an optical fiber, where the propagation parameters depend on the waveguide structure.

If an optical signal with a spectrum width $\delta\omega$ propagates along an optical fiber of length L, group velocity dispersion leads to pulse distortion because different spectrum components travel with different group velocity and arrive at the fiber output with different propagation delays. The difference in propagation delay for different spectrum components can be found by

$$\delta\tau = \frac{d\tau}{d\omega} = \frac{d}{d\omega}\left(\frac{L}{v_g}\right)\delta\omega, \tag{2.65}$$

from which we have

$$\delta\tau = L\frac{d^2\beta}{d\omega^2}\Delta\omega, \tag{2.66}$$

where $\beta_2 = d^2\beta/d\omega^2$ is known as the *group velocity dispersion parameter*. For two signals separated by a unit wavelength, the difference in group delay in an optical fiber with unit length is defined as the *fiber dispersion parameter*, given by

$$D = \frac{d}{d\lambda}\left(\frac{1}{v_g}\right) = -\frac{2\pi c}{\lambda^2}\beta_2. \tag{2.67}$$

Since

$$\frac{d\beta}{d\omega} = \frac{1}{c}\left(n_2 + \omega\frac{dn_2}{d\omega}\right)(1 + b\Delta) + \frac{\omega n_2}{c}\frac{d\,(b\Delta)}{d\omega}, \tag{2.68}$$

we can find the group velocity dispersion parameter as

$$\frac{d^2\beta}{d\omega^2} = \frac{1}{c}\frac{dN_2}{d\omega}\left(1+\Delta\frac{d(Vb)}{dV}\right) + \frac{1}{\omega c}\frac{\Delta N_2}{n_2}V\frac{d^2(Vb)}{dV^2} + \frac{N_2}{c}\frac{d\Delta}{d\omega}\frac{d^2(V^2b)}{dV^2}, \tag{2.69}$$

where N_2 is the group index, defined as

$$N_2 = n_2 + \omega\frac{dn_2}{d\omega}. \tag{2.70}$$

Based on the group velocity dispersion parameter β_2, the fiber dispersion parameter can be determined as

$$D = D_m + D_w + D_d, \tag{2.71}$$

where the material dispersion parameter is given by

$$D_m = \frac{\omega}{\lambda c}\frac{dN_2}{d\omega}\left(1+\Delta\frac{d(Vb)}{dV}\right) = -\frac{\lambda}{c}\frac{d^2n_2}{d\lambda^2}\left(1+\Delta\frac{d(Vb)}{dV}\right), \tag{2.72}$$

the waveguide dispersion parameter is

$$D_w = -\frac{1}{\lambda c}\frac{\Delta N_2^2}{n_2}\left(V\frac{d^2(Vb)}{dV^2}\right) = -\frac{N_1 - N_2}{\lambda c}\left(V\frac{d^2(Vb)}{dV^2}\right), \tag{2.73}$$

and the differential material dispersion parameter is

$$D_d = -\frac{\omega N_2}{\lambda c}\frac{d\Delta}{d\omega}\frac{d^2\left(V^2b\right)}{dV^2}. \tag{2.74}$$

For the material dispersion parameter (2.72), the term $\Delta \cdot d(Vb)$ is small, and thus the material dispersion is determined largely by the wavelength dependence of n_2. For standard single-mode fibers, the material dispersion is negative for smaller wavelength ($< 1.3\mu$m) and positive for longer wavelength. For the waveguide dispersion parameter (2.73), the term in parentheses is due to the effect of the waveguide on the propagation constant. The waveguide dispersion for standard single-mode fibers is negative. For the differential material dispersion parameter, $d\Delta/d\omega$ is very small, and hence the differential material dispersion parameter in standard single-mode fibers is usually neglected.

Figure 2.5 shows the dispersion parameter of the standard single-mode fiber at various wavelengths, as well the material dispersion parameter and the waveguide dispersion parameter. Since the negative waveguide dispersion balances the positive materials dispersion, the zero-dispersion wavelength of the standard single-mode

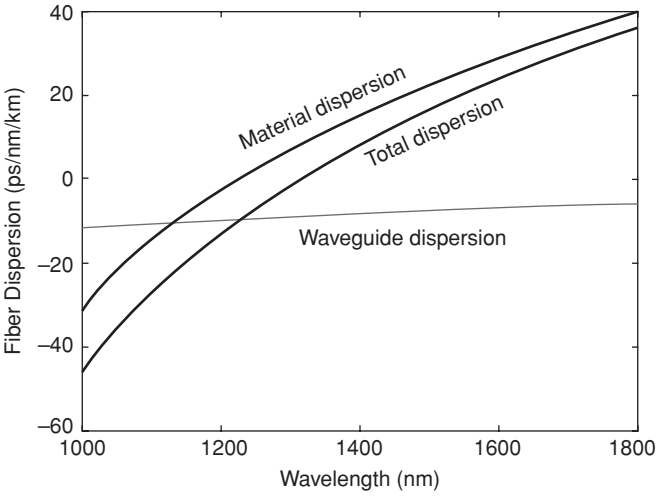


FIGURE 2.5 Dispersion coefficient of standard single-mode fiber.

fiber is around 1300 nm. The dispersion parameter of standard single-mode fiber in the 1.55-μm wavelength region is about 17 ps/km·nm. As the waveguide dispersion D_w depends on fiber parameters such as the core radius a and the refractive index difference Δ, with proper design to make the absolute value of waveguide dispersion larger, it is possible to shift the zero-dispersion wavelength into the 1550-nm region with lowest fiber loss. This type of fiber is called *dispersion-shifted fiber*. Furthermore, by using multiple cladding layers and a more complex refractive index profile, it is possible to achieve a low total dispersion over a wide range of wavelength. Such fibers are called *dispersion-flattened fibers*.

When an optical communication system operates in wavelengths away from zero-dispersion wavelength, pulse distortion is determined primarily by the first-order dispersion parameter D. However, when the system operates at the zero-dispersion wavelength, then higher-order dispersion becomes significant. The dominant higher-order dispersive effects are dominated by the dispersion slope, given by

$$S = \frac{dD}{d\lambda} = \left(\frac{2\pi c}{\lambda^2}\right)^2 \frac{d^3\beta}{d\omega^3} + \frac{2\pi c}{\lambda^3}\frac{d^2\beta}{d\omega^2}. \tag{2.75}$$

The dispersion slope is also called the *differential-dispersion parameter* or *second-order dispersion parameter*. At zero-dispersion wavelength, $d^2\beta/d\omega^2 = 0$, so the second-order dispersion is proportional to $d^3\beta/d\omega^3$. For a source with spectral width $\delta\lambda$, the effective value of the dispersion parameter becomes $D_{\text{eff}} = S\delta\lambda$.

2.1.4.3 Polarization Mode Dispersion For an ideal optical fiber with cylindrical symmetry, optical signals with two orthogonally polarized modes have the same propagation constant. However, due to irregularities in fiber geometry and

nonuniform stress in cabling of optical fibers, optical fibers can never be perfectly cylindrical. As a consequence, the mode indices and propagation constants associated with two orthogonally polarized optical signals exhibit a slight difference. If an input optical pulse excites both polarization modes, the pulse will become broader at the fiber output since the two polarization components disperse along the fiber because of their different group velocities. This phenomenon is known as *polarization mode dispersion*. Similar to group velocity dispersion, pulse broadening can be estimated from the time delay δT between the two polarization components, given by

$$\delta T = \left| \frac{L}{v_{gx}} - \frac{L}{v_{gy}} \right| = L \left| \frac{d\beta_x}{d\omega} - \frac{d\beta_y}{d\omega} \right|, \tag{2.76}$$

where the subscripts x and y denote the two orthogonal polarization modes. Polarization mode dispersion can be characterized by

$$\delta\beta_1 = \left| \frac{d\beta_x}{d\omega} - \frac{d\beta_y}{d\omega} \right|. \tag{2.77}$$

However, $\delta\beta_1$ cannot be used directly to estimate polarization mode dispersion because of random mode coupling. In practice, polarization mode dispersion is usually characterized by the root-mean-square value of δT,

$$\sigma_T^2 = \frac{1}{2} (\delta\beta_1)^2 h^2 \left[\frac{2L}{h} - 1 + \exp\left(-\frac{2L}{h} \right) \right], \tag{2.78}$$

where h is the decorrelation length, with typical values in the range 1 to 10 m. For $L \gg h$,

$$\sigma_T \approx \delta\beta_1 \sqrt{hL} = D_p \sqrt{L}, \tag{2.79}$$

where D_p is the polarization mode dispersion parameter. Typical values of D_p are in the range 0.1 to 1 ps/$\sqrt{\text{km}}$. Because of its square-root dependence on the fiber length, pulse broadening induced by polarization mode dispersion is relatively small compared with the group velocity dispersion. However, polarization mode dispersion can become a dominant factor for high-speed fiber optic communication systems operating over a long distance near the zero-dispersion wavelength or with dispersion compensation.

2.1.5 Nonlinear Effects

Even though silica is not a highly nonlinear medium, nonlinear processes can be observed in optical fibers at relatively lower power levels (10 mW) because of two important characteristics of single-mode fibers: a small spot size (which leads to high intensities at low powers) and extremely low loss (which results in long effective

length). Although some optical signal processing applications such as optical switching and optical regeneration take advantage of nonlinear effects, nonlinear effects in optical fibers are generally detrimental for long-distance optical transmission. There are two principal nonlinear effects in optical fibers: nonlinear scattering and nonlinear refraction. *Nonlinear scattering*, arising from the interaction of photons and phonons in fiber materials, could lead to loss of optical power and limit optical power levels that can be transmitted through an optical fiber. *Nonlinear refraction*, on the other hand, results from the interaction between photons and bounded electrons in SiO_2. Nonlinear refraction includes self-phase modulation, cross-phase modulation, and four-wave mixing effects.

2.1.5.1 Nonlinear Scattering There are two major nonlinear scattering effects: stimulated Brillouin scattering and stimulated Raman scattering. In stimulated Brillouin scattering, a photon loses energy to a lower-energy photon and an acoustic phonon is generated in the scattering process. In stimulated Raman scattering, on the other hand, a photon loses its energy to a lower-energy photon and creates an optical photon in the silica. Therefore, both stimulated Brillouin and Raman scattering generate scattered light with a longer wavelength. If a signal is present with a longer wavelength, nonlinear scattering will amplify this signal while the original signal loses power. Stimulated Raman gain spectrum is very broad, extending to 30 THz. The gain spectrum of stimulated Brillouin scattering is very narrow, with a gain bandwidth of less than 100 MHz. In single-mode fibers, stimulated Brillouin scattering occurs only in the backward direction, whereas stimulated Raman scattering happens in both directions, but forward direction dominates in stimulated Raman scattering. The effect of stimulated Brillouin and Raman scattering is very small when the optical power level of an optical signal is low. However, if the optical power is above a threshold, significant nonlinear scattering could happen. For stimulated Brillouin scattering, the threshold is about 5 mW for CW (continuous wave) light. With modulated signal, the threshold can be increased to 10 mW or more. For stimulated Raman scattering, its threshold is about 600 mW in the 1.55-μm region. As optical powers in optical communication systems are usually less than 10 mW, stimulated Raman scattering is generally not a limiting factor. However, for multiple channels, the Raman effect can introduce crosstalk between channels. On the positive side, stimulated Raman scattering has been used to amplify an optical signal when a strong pump signal with shorter wavelength is present in the fiber.

2.1.5.2 Nonlinear Refraction At high powers, the refractive index of silica fibers increases with optical intensity, due to nonlinear refraction. Since the propagation constant β depends on fiber parameters such as Δ, it is also dependent on optical power, due to nonlinear refraction. In practice, as the intensity (or power) of an optical signal varies as a function of time, so does the refractive index of the silica fiber. When an optical signal propagates along an optical fiber, the phase of the optical signal is proportional to $\exp(-j\beta z)$. A time-varying refractive index of silica fiber will lead to a time-varying propagation constant β and the phase of the optical signal. A time-varying optical phase introduces frequency chirp and broadens the

optical signal spectrum. When combined with fiber group velocity dispersion, large-frequency chirp can lead to significant pulse distortion. If the nonlinear refraction is induced by an optical signal itself, it is called *self-phase modulation*. In a WDM system, nonlinear refraction is caused by multiple optical signals, and the nonlinear refraction effect is known as *cross-phase modulation*.

Another nonlinear refraction effect in optical fibers is four-wave mixing. In WDM optical communication systems, two optical signals can create an interference pattern (i.e., intensity variation) in the fiber. Through nonlinear refraction, the intensity variation gives rise to refractive index grating in the fiber. If a third signal is also present, it will be scattered to another wavelength. With four-wave mixing, a significant amount of optical power may be transferred to its neighboring channels. Such power transfer leads to power loss for a specific channel, and crosstalk in another channel. To generate strong four-wave mixing, phase matching is needed for those four optical signals. In contrast to self-phase and cross-phase modulation, which are significant only for high-speed systems, four-wave mixing is independent of the bit rate. However, four-wave mixing depends on the channel spacing and fiber group velocity dispersion. Increasing the channel spacing or increasing group velocity dispersion reduces the four-wave mixing effect, because the phase of optical signals cannot be well matched in these cases.

2.1.6 Light-Wave Propagation in Optical Fibers

Propagation of an optical signal in single-mode fibers can be described by the nonlinear Schrödinger equation, given by

$$\frac{\partial A(z,t)}{\partial z} + \frac{\alpha_f}{2}A(z,t) + \beta_1\frac{\partial A(z,t)}{\partial t} + j\frac{1}{2}\beta_2\frac{\partial^2 A(z,t)}{\partial t^2} - \frac{1}{6}\beta_3\frac{\partial^3 A(z,t)}{\partial t^3} = j\gamma\left|A^2\right|A(z,t), \tag{2.80}$$

where $A(z,t)$ is the slowly varying envelope of the optical field and γ is the nonlinear coefficient with a typical value of 1 to 5 W^{-1}/km. $\beta_m = d^m\beta/d\omega^m$ represents different orders of fiber dispersion. The propagation equation is a nonlinear differential equation that does not generally have analytical solutions. A numerical approach is often necessary for understanding of the propagation of an optical signal in optical fibers. Two generally used numerical methods are the split-step Fourier transform method and the finite-difference method.

If the optical power of an optical signal is relatively low (< 5 mW), the nonlinear effect of optical fibers can be neglected. In this case, the optical fiber is essentially a dispersive medium with a transfer function given by

$$H(f) = \exp\left\{-\frac{1}{2}\alpha_f L + \frac{j\pi\lambda_0^2 DLf^2}{c}\left[1 - \frac{\lambda_0}{3c}\left(2 + \frac{\lambda_0}{D}\frac{dD}{d\lambda}\right)\right]\right\}, \tag{2.81}$$

where $dD/D\lambda$ is the dispersion slope. If the optical wavelength is far away from the zero-dispersion wavelength, second-order dispersion (dispersion slope) can be neglected, and the fiber transfer function can be simplified as

$$H(f) = \exp\left(-\frac{1}{2}\alpha_f L + \frac{j\pi\lambda_0^2 DLf^2}{c}\right). \tag{2.82}$$

Neglecting nonlinear effect, an optical fiber introduce an attenuation factor and a nonlinear phase factor.

2.2 OPTICAL TRANSMITTERS

Optical transmitters convert electrical signals to optical signals and launch the optical signals into optical fibers for optical transmission. A key component in optical transmitters is an optical source. Semiconductor lasers are commonly used in optical transmitters because of their narrow spectral width, high-speed modulation, compact size, and good reliability. Passive optical networks usually utilize direct modulation of semiconductor lasers for optical transmission up to 10 Gb/s. However, for transmitters with direction modulation, the wide spectrum of optical signals transmitted, combined with fiber dispersion, limits the optical transmission distance at high data rates. In high-performance optical communication systems, optical modulators are used to convert electrical signals to optical signals, while semiconductor lasers serve as CW (continuous wave) light sources. In this section we review operation principles and characteristics of semiconductor lasers and optical modulators. Furthermore, the design of optical transmitters is discussed, with emphasis on the system performance of optical transmissions.

2.2.1 Semiconductor Lasers

Semiconductor lasers were invented in 1962 with pulsed operation at liquid-nitrogen temperature (77 K). In the 1970s, CW semiconductor lasers operating at room temperature became available and widely used in optical communication systems with direct modulation. Since then, various types of semiconductor lasers have been developed with improved performance and reliability, including FP (Fabry–Perot) lasers, DFB (distributed feedback) lasers, DBR (distributed Bragg reflector) lasers, and VCSELs (vertical cavity surface-emitting lasers). In addition, tunable semiconductor lasers have been demonstrated for optical networks with improved performance.

2.2.1.1 Principle of Operation The operation of semiconductor lasers is based on stimulated emission. Figure 2.6 shows the band diagram of a semiconductor with a direct bandgap (i.e., the minimum of conduction band is coincident with the maximum of the valence band). When the energy of a photon is larger than the bandgap of the semiconductor material, carriers (electrons and holes) in the semiconductor can interact with photons in a number of ways. An electron in the valence band can

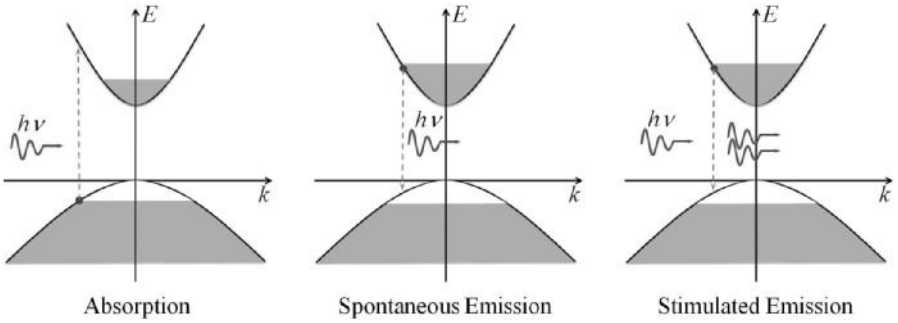


FIGURE 2.6 Optical processes in a direct bandgap semiconductor: absorption, spontaneous emission, and stimulated emission. In absorption, the loss of photon energy generates an electron–hole pair. In spontaneous emission, the recombination of an electron–hole pair generates a photon. In stimulated emission, a photon stimulates the recombination of an electron–hole pair, and another photon with the same energy and direction as the incident photon is generated.

absorb the energy of a photon and end up in the conduction band. Meanwhile, a hole is generated in the valence band. This is the normal case of absorption in any semiconductor materials. The excited electron in the conduction band will eventually combine with a hole in the valence band and emit a photon when the transition happens. This is called *spontaneous emission*. The photon emitted by spontaneous emission has a random direction, but with an energy equal to the transition energy (from conduction band to valence band) of the electron. A third process, *stimulated emission*, happens when a photon stimulates the transition of an electron in the conduction band to the valence band and an additional photon is emitted with the same energy and direction as the incident photon. Through stimulated emission, optical signals can be amplified as more photons are generated in the process. For a semiconductor material in equilibrium, there are more electrons in the valence band and more holes in the conduction band. Therefore, the absorption process dominates both spontaneous and stimulated emission, and the semiconductor material is absorptive for optical signals with photon energy larger than its bandgap. For optical amplification to happen, the stimulated emission process must be stronger than the absorption process. This can be achieved in nonequilibrium states with population inversion. In semiconductor materials, population inversion requires that the probability of occupation for electrons in the bottom of the conduction band is larger than the probability of occupation for electrons in the valence band. In other words, there are a large number of electrons in the bottom of the conduction band and a large number of holes in the valence band. When a photon with appropriate energy (larger than the bandgap energy) is incident in the semiconductor, there is a higher probability that the photon will stimulate the recombination of electrons in the conduction band and holes in the valence band than when the photon energy is absorbed by an electron in the valence band. To achieve population inversion in semiconductor materials, external energy

must be injected into semiconductor materials to excite electrons from the valence band to the conduction band. Optical energy or electrical current is commonly used as excitation for semiconductor materials. Through stimulated emission, semiconductor lasers convert electrical energy (electrical current) to optical energy.

Essentially, a semiconductor is an optical cavity resonator with a semiconductor gain medium. The gain in the semiconductor material compensates for the losses in the cavity, so that an optical wave can be bounced back and forth inside the cavity without loss of energy. The optical radiation modes of a laser are determined by the properties of the cavity, such as the structure, dimension, and refractive index of the cavity. The simplest cavity is a Fabry–Perot etalon (interferometer), which is generally used in semiconductor lasers. Fabry–Perot lasers consist of two plane reflectors or mirrors (semiconductor lasers generally use the cleaved end facets as the reflection mirror) with a plane electromagnetic wave propagating along the axis normal to the mirrors. The electromagnetic wave inside the laser cavity can be written as

$$E(t, x) = A \exp[(g - \alpha_s) x/2] \exp[j(\omega t - \beta x)], \tag{2.83}$$

where g is the power gain coefficient due to the gain medium inside the cavity, α_s is the power attenuation constant, and β is the propagation constant given by $\beta = 2\pi n_s/\lambda$.

Assume that the mirrors have reflection coefficients r_1 and r_2 and are separated by a distance L (the length of the cavity). The electromagnetic wave will be reflected back and forth inside the cavity. If a wave starts at one mirror ($x = 0$) as

$$E(t, 0) = A \exp(j\omega t), \tag{2.84}$$

after one round trip in the cavity the electromagnetic wave becomes

$$E'(t, x) = A r_1 r_2 \exp[(g - \alpha_s) x/2] \exp[j(\omega t - \beta x)]. \tag{2.85}$$

The condition for steady-state oscillation is that the complex amplitude, that is, the magnitude and phase, of the return wave must be equal to the original amplitude and the return wave must be equal to the original amplitude and phase. This gives

$$r_1 r_2 \exp[(g - \alpha_s) L] = 1 \qquad \text{and} \qquad \exp(j2\beta L) = 1. \tag{2.86}$$

The first equation will enable us to determine which modes have sufficient gain for oscillation to be sustained and to calculate the amplitudes of these modes. The second equation will enable us to find the resonant frequencies of the Fabry–Perot optical cavity resonator.

Gain To sustain the laser oscillation, the gain must satisfy

$$r_1 r_2 \exp[(g - \alpha_s) L] \geq 1; \tag{2.87}$$

in other words,

$$g \geq \alpha_s + \frac{1}{L} \ln \frac{1}{r_1 r_2} = \alpha_s + \frac{1}{2L} \ln \frac{1}{R_1 R_2}, \tag{2.88}$$

where $R_1 = r_1^2$ and $R_2 = r_2^2$ are the power reflection coefficient of the mirrors. The first term on the right side of the equation, α_s, includes all of the distributed losses, such as scattering and absorption. The second term represents the losses at the mirrors, a lumped loss that is averaged over the length $2L$. If If the gain is equal to or greater than the total loss, oscillations will be initiated. As the amplitude increases, nonlinear (saturation) effects will reduce the gain. The stable amplitude of oscillation is the amplitude for which the gain has been reduced so that it exactly matches the total loss.

Longitudinal Modes The condition for phase matching after a round trip gives

$$\exp(j2\beta L) = 1; \tag{2.89}$$

that is,

$$2\beta L = 2m\pi, \tag{2.90}$$

where m is an integer. Since $\beta = 2\pi n_s/\lambda$, we find that the laser wavelength is given by

$$\lambda = \frac{2nL}{m}, \tag{2.91}$$

where n is the refractive index of the semiconductor. Equation (2.91) gives the relationship between the laser wavelength and the laser cavity: The optical length of the cavity must be an integral number of half wavelengths. Different values of m correspond to different longitudinal modes.

For semiconductor lasers, the gain, g, is a function of frequency. There will be a band of frequencies in FP lasers for which the gain is greater than the loss. Within that band, there will usually be several, or many, frequencies that satisfy the oscillation condition. At each of these frequencies, laser oscillation will occur. These frequencies, with their intensity and other characteristics, are called *modes of oscillation* of the laser. For FP lasers, the loss is independent of the frequency (or mode independent). The mode closest to the gain peak becomes the dominant mode. Under ideal conditions, the other modes would not reach the oscillation condition (threshold) since their gains are always less than those of the dominant modes. However, in practice the difference in the gain is very small and the neighboring modes on each side of the dominant mode have a significant portion of the total power. The spectrum of the FP lasers usually shows several significant side modes, and the spectrum width is about a few nanometers. Each mode emitted by FP lasers

propagates in an optical fiber at a slightly different speed, due to group velocity dispersion, and pulse broadening occurs for optical communication systems using FP lasers. The multimode nature of an FP laser limits its transmission distance operating in 1.55 μm. Significant improvement in transmission distance and/or data rate can be achieved with single-longitudinal-mode lasers.

Single-Longitudinal-Mode Lasers In contrast with FP lasers, whose losses are mode independent, single-longitudinal-mode lasers are designed such that cavity losses are different for different longitudinal modes of the cavity. Because of the mode-dependent loss, the longitudinal mode with the smallest cavity loss reaches threshold first and becomes the dominant mode. Other neighboring modes are discriminated by their higher losses, which prevent their buildup. Single-mode operation of semiconductor lasers is usually characterized by the side-mode suppression ratio (SMSR),

$$\text{SMSR} = \frac{P_{\text{mm}}}{P_{\text{sm}}}, \tag{2.92}$$

where P_{mm} is the power of the main lasing mode and P_{sm} is the power of the most significant side mode. There are several techniques to achieve single-mode operation for semiconductor lasers. DFB lasers, DBR lasers, VCSELs, and couple-cavity semiconductor lasers all operate in single longitudinal mode.

The feedback in distributed feedback (DFB) lasers is not localized at the facets but is distributed throughout the cavity length. This is achieved through an internal built-in grating for a periodic variation of the mode index. Feedback occurs by means of Bragg diffraction, where the forward- and backward-propagating waves are tightly coupled with each other. Mode selectivity in DFB lasers results from the Bragg condition, and strong coupling occurs only for wavelength $\lambda_B = 2\bar{n}\Lambda$, where Λ is the grating period and $\bar{n}$ is the average mode index. Similarly, DBR lasers use Bragg gratings as end mirrors, whose reflectivity is maximum for λ_B. The cavity loss of DFB or DBR lasers is therefore minimum for the longitudinal mode closest to λ_B and increases substantially for other longitudinal modes.

The light is emitted in a direction normal to the active-layer plane, so the active region is very short and does not provide much gain. Two high-reflectivity (>99.5%) DBR mirrors are grown expitaxially on both sides of the active layer to form a microcavity. Because of the high-reflectivity DBR mirrors, loss of the cavity is relatively small, so that VCSELs can lase with a relatively small gain. As VCSELs have an extremely small cavity length (1 μm), the mode spacing for VCSELs is much larger than the gain bandwidth, so there is only one longitudinal mode in the entire gain spectrum.

In couple-cavity semiconductor lasers, single-longitudinal-mode operation is realized by coupling the light to an external cavity. A portion of the reflected light from the external cavity is fed back to the laser cavity. The in-phase feedback occurs only for those laser modes whose wavelength nearly coincides with one of the longitudinal

modes of the external cavity. In effect, the effective reflectivity of the laser facet facing the external cavity becomes wavelength dependent and leads to mode-dependent losses. The longitudinal modes that are closest to the gain peak and have the lowest cavity loss become the dominant modes.

Tunable Semiconductor Lasers Tunable light sources provide flexibility and reconfigurability for network provisioning, minimize production cost, and reduce the backup stock required. Commonly used options for tunable lasers are external cavity lasers, multisection DFB/DBR lasers, and tunable VCSELs. Due to the cost concerns, it is desirable that tunable lasers can be directly modulated. An external cavity laser is usually tuned by changing the characteristics of the external cavity, which consists of a grating or FP etalon. The tuning ranges of external cavity lasers are extremely wide, covering a few hundred nanometers. However, the long cavity length prevents high-speed modulation, so external cavity lasers are not suitable for fiber optic communications. The tuning speed and stability are also issues with external cavity lasers. Traditional DFB lasers can support high-speed direct modulation and be thermally tuned over a few nanometers. However, the tuning speed is limited to the millisecond range. Multisection DFB/DBR lasers usually consist of three or more sections: an active (gain) section, a phase control section, and a Bragg section. Wavelength tuning is achieved by adjusting the currents in the phase-control and Bragg sections. Using multisection DFB/DBR lasers, the tuning speed can reach nanoseconds by current injection, and tuning ranges over tens of nanometers can be achieved. Some multisection DFB/DBR lasers with sampled gratings can be tuned over 100 nm. The disadvantages of multisection DFB/DBR lasers are mode hopping and complicated electronic control.

Tunable VCSELs use a MEMS (microelectromechanic system) structure that changes the cavity length through electrostatic control. The tuning speed can be a few microseconds and the tuning range can reach 10 to 20 nm. VCSELs have a potential for low-cost mass production because of simple one-step epitaxy and on-chip testing. However, the development of long-wavelength VCSELs has been hindered by unsatisfactory optical and thermal properties of InP-based group III to V materials. New design using different materials and dielectric mirrors has lead to the successful development of 1.3- and 1.5-m VCSELs. As the fabrication method matures, VCSELs will be a strong candidate for access networks.

2.2.1.2 Semiconductor Laser Characteristics

A semiconductor laser is an EO (electrical–optical) converter that converts electrical energy into optical energy. Its operation characteristics are governed by the interaction of electrons and photons inside the laser cavity. Based on laser rate equations, we describe the major properties of semiconductor lasers, including their steady-state output, small-signal frequency response, and large-signal modulation characteristics.

Rate Equations Assume that a semiconductor laser operates in a single longitudinal mode above the threshold condition. The photon density and carrier density

inside the active region can be described by two coupled rate equations. The carrier rate equation is given by

$$\frac{dN(t)}{dt} = \frac{J(t)}{qV} - \frac{N(t)}{\tau_n} - g_0 (N(t) - N_0) \frac{1}{1 + \epsilon S(t)} S(t), \tag{2.93}$$

where $N(t)$ is the carrier density, $S(t)$ the photon density, $J(t)$ the injection current, q the elementary charge, τ_n the carrier lifetime, g_0 the gain slope constant, N_0 the carrier density at transparency for which the net gain is zero, and ϵ the gain compression factor. The first term on the right-hand side is the rate of carrier injection, the second term denotes spontaneous emission, and the third term represents stimulated emission. The rate equation for photon density can be written

$$\frac{dS(t)}{dt} = \Gamma g_0 (N(t) - N_0) \frac{1}{1 + \epsilon S(t)} S(t) - \frac{S(t)}{\tau_p} - \frac{\Gamma \beta N(t)}{\tau_n}, \tag{2.94}$$

where τ_p is the photon lifetime, Γ the mode confinement factor, and β the fraction of spontaneous emission coupled into the laser mode. The first term on the right-hand side denotes the increase in photon density due to stimulated emission, the second term is the rate of photon loss because of radiation (laser output) and absorption, and the third term is the spontaneous emission coupled into the lasing mode.

Steady State Under steady-state conditions, the rates of changes in electron density and photon density are zero. If we neglect the spontaneous emission and the gain saturation factor for stimulated emission, we have

$$\frac{\bar{J}}{qV} - g_0 \left(\bar{N} - N_0\right) \bar{S} - \frac{\bar{N}}{\tau_n} = 0, \tag{2.95}$$

$$\Gamma g_0 \left(\bar{N} - N_0\right) (\bar{S}) - \frac{\bar{S}}{\tau_p} = 0, \tag{2.96}$$

where the overbars denote steady-state quantities. For a current such that $\bar{J} = \Gamma g_0 \left(\bar{N} - N_0\right) \tau_p < 1$, the gain is less than the loss, and the laser threshold has not been reached. Below threshold, the photon density $\bar{S} = 0$ if spontaneous emission is neglected. The carrier density is $\bar{N} = \tau_n \bar{J}/qV$ (i.e., the carrier density increases linearly with the injection current).

The laser threshold is reached at an injection current for which $\Gamma g_0 \left(\bar{N} - N_0\right) \tau_p = 1$. From eq. (2.96) we have

$$\bar{N} = N_0 + \frac{1}{\Gamma g_0 \tau_p} \equiv N_t h. \tag{2.97}$$

Thus, at steady state above the laser threshold, the carrier density is clamped at a threshold level $N_t h$. The gain is also clamped at

$$\bar{g} = g_0 \left(N_t h - N_0 \right) = \frac{1}{\Gamma \tau_p}. \tag{2.98}$$

The threshold current is obtained from eq. (2.97) by setting the photon density $\bar{S} = 0$:

$$J_t h = \frac{q V N_t h}{\tau_n} = \frac{q V}{\tau_n} \left(N_0 + \frac{1}{\Gamma g_0 \tau_p} \right). \tag{2.99}$$

For an injection current $\bar{J} > J_t h$, lasing starts in the semiconductor cavity. The photon density is given by

$$\bar{S} = \frac{1}{g_0 \left(\bar{N} - N_0 \right)} \left(\frac{\bar{J}}{qV} - \frac{\bar{N}}{\tau_n} \right) = \frac{\Gamma \tau_p}{qV} \left(\bar{J} - \frac{q V N_t h}{\tau_n} \right) = \frac{\Gamma \tau_p}{qV} \left(\bar{J} - I_t h \right). \tag{2.100}$$

The output optical power of the semiconductor laser is related to the photon density inside the laser cavity:

$$\bar{P} = \frac{1}{2} \eta_i nt \left(v_g \alpha_m \right) \hbar \omega \bar{S} V \propto \left(\bar{J} - I_t h \right), \tag{2.101}$$

where $\alpha_m = -\ln \left(R_1 R_2 \right) / 2L$ is the loss associated with the cavity mirrors and $\eta_i nt$ is the internal quantum efficiency, which indicates the fraction of injected electrons that are converted into photons through stimulated emission. Above threshold, $\eta_i nt$ is close to 100% for most semiconductor lasers. The factor of 1/2 makes P the power emitted from each facet with equal reflectivities. The optical output power of a semiconductor laser is linearly proportional to the injection current.

Modulation Dynamics The small-signal transfer function of semiconductor lasers can be obtained by applying a current of the form

$$J \left(t \right) = \bar{J} + J_m e^{j \omega t}, \tag{2.102}$$

where $\bar{J}$ is a dc bias current and the modulation current is small compared to the bias current, $J_m \ll \bar{J}$. Under first-order approximation, the electron density and photon density can be written as

$$N \left(t \right) = \bar{N} + N_m e^{j \omega t}, \tag{2.103}$$

$$S \left(t \right) = \bar{S} + S_m e^{j \omega t}. \tag{2.104}$$

Substituting eqs. (2.103) and (2.104) into the coupled rate equations and using

$$\frac{1}{1+\epsilon S} \approx 1-\epsilon S \qquad \text{for} \quad \epsilon S \ll 1,$$

we can find the small-signal frequency response given by

$$H(j\omega) = \frac{S_m}{J_m} = \frac{X}{(j\omega)^2 + j\omega Y + Z}, \tag{2.105}$$

where

$$X = \frac{1}{qV}\left[\Gamma g_0 \left(1-\epsilon \bar{S}\right)\bar{S} + \frac{\Gamma\beta}{\tau_n}\right],$$

$$Y = g_0\left(1-\epsilon\bar{S}\right)\bar{S} - \Gamma g_0\left(\bar{N}-N_0\right)\left(1-2\epsilon\bar{S}\right) + \frac{1}{\tau_n} + \frac{1}{\tau_p},$$

$$Z = \frac{1}{\tau_n} g_0\left(1-\epsilon\bar{S}\right) + \frac{1}{\tau_n}\Gamma g_0(\beta-1)\left(\bar{N}-N_0\right)\left(1-2\epsilon\bar{S}\right) + \frac{1}{\tau_n\tau_p}.$$

Due to the interaction of photons and electrons, the response of a semiconductor laser to an injection current resembles a second-order system with damping oscillation of the form

$$\frac{H(j\omega)}{H(0)} = \frac{\omega_n^2}{(j\omega)^2 + 2\xi\omega_n(j\omega) + \omega_n^2}, \tag{2.106}$$

where the natural frequency is given by

$$\omega_n = \sqrt{Z}, \tag{2.107}$$

and the damping ratio is

$$\xi = \frac{Y}{2\sqrt{Z}}. \tag{2.108}$$

The 3-dB bandwidth of the small-signal response is

$$\mathrm{BW}_{3\mathrm{dB}} = \omega_n\left[\left(1-3\xi^2\right) + \sqrt{4\xi^4 - 4\xi^2 + 2}\right]^{1/2}. \tag{2.109}$$

Thus, the 3-dB bandwidth is proportional to the natural frequency ω_n and decreases with an increase in the sampling factor ξ. In practice, the 3-dB bandwidth of

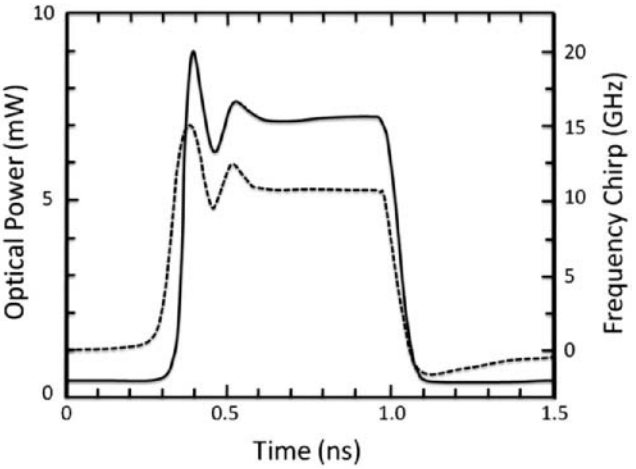


FIGURE 2.7 Optical output power and frequency chirp of a semiconductor laser under large-signal modulation at 1.25 Gb/s. The solid line represents the laser output power, and the dashed line denotes the frequency chirp of the output signal. Due to the interaction between photon and electrons, the response of the semiconductor laser exhibits overshoot and ringing. Th frequency chirp of the semiconductor laser is caused by the variation in carrier density and the temperature effect, leading to significant spectrum broadening.

semiconductor lasers can be approximated by

$$\mathrm{BW}_{3\mathrm{dB}} \propto \left(\bar{J} - J_t h\right)^{1/2} \propto P^{1/2}; \tag{2.110}$$

that is, the 3-dB bandwidth increases with an increase in bias current. For large-signal modulation, the rate equations of semiconductor lasers have to be solved numerically. Figure 2.7 shows the pulse response of semiconductor lasers. Due to the electron–photon interaction, the output power of a semiconductor laser shows relaxed oscillation behavior, as demonstrated by overshoot and ringing in large-signal modulation.

Frequency Chirp From Kramers–Kronig relations we know that whenever the optical gain of a semiconductor material changes, its refractive index changes as well. As a consequence, amplitude modulation in semiconductor lasers is always accompanied by phase modulation because of the carrier-induced changes in the mode index n. Two fundamental effects contribute to the frequency chirp of semiconductor lasers: the changes in temperature and the change in carrier concentration. Temperature affects the refractive index (increases with increased temperature) and the optical length of the laser cavity (the product of the two, nL, determines the laser mode frequency). Temperature also changes the semiconductor bandgap (decreases with increased temperature) and thus the gain spectrum. Carrier concentration affects both the refractive index (decreases with increased carrier density due to plasma-induced

refractive index change) and the gain coefficient (proportional to carrier density). In DFB lasers, the lasing frequency is determined by the mode of the distributed feedback structure. The gain spectrum and bandgap change do not affect the laser wavelength; so the dominant factor here is the refractive index change.

The phase modulation associated with intensity modulation in semiconductor lasers can be described by

$$\frac{d\phi}{dt} = \frac{1}{2}\alpha_c \left[\Gamma g_0 \left(N\left(t\right) - N_0 \right) - \frac{1}{\tau_p} \right], \tag{2.111}$$

where α_c is the linewidth enhancement factor (also called the α parameter), as it leads to an enhancement of spectral width associated with a single longitudinal mode. For a single-mode laser, neglecting the gain compression factor ϵ and the spontaneous emission factor β, the rate equation for photon density becomes

$$\frac{dS}{dt} = \Gamma g_0 \left(N - N_0 \right) S - \frac{S}{\tau_p}. \tag{2.112}$$

Therefore,

$$\frac{d\phi}{dt} = \frac{1}{2}\alpha_c \frac{d \ln S\left(t\right)}{dt}. \tag{2.113}$$

Since the output power of the semiconductor laser is proportional to the photon density, $P(t) \propto S(t)$, we have $d\left(\ln P(t)\right) = d\left(\ln S(t)\right)$. The frequency chirp associated with intensity modulation can be written

$$\Delta f = \frac{1}{2\pi}\frac{d\phi}{dt} = \frac{1}{4\pi}\alpha_c \frac{d\left(\ln P\left(t\right)\right)}{dt}. \tag{2.114}$$

The linewidth enhancement factor can thus be determined by

$$\alpha_c = \frac{2d\phi\left(t\right)}{d\left(\ln P\left(t\right)\right)}. \tag{2.115}$$

The linewidth enhancement factor is in the range 4 to 8, and 1 to 2 for quantum well lasers. Under pulse operation, the chirp of the semiconductor lasers causes the mode wavelength to vary during the pulse. The wavelength becomes shorter (positive frequency chirp) during the leading edge of the pulse and longer (negative frequency chirp) during the trailing edge, as shown in Figure 2.7 The frequency chirp from a directly modulated semiconductor laser causes significant broadening in the spectrum of the output optical signals. When combined with group velocity dispersion of an optical fiber, the frequency chirp could cause significant pulse broadening and thus degrade the system performance. At the 1.55-μm region, the dispersion coefficient of standard-mode fibers is positive, so higher-frequency spectral components of an optical signal travel faster than do lower-frequency spectral components. Because of

the laser chirp, the leading edge (positive-frequency chirp) of an optical pulse will travel faster than the trailing edge (negative-frequency chirp). Therefore, frequency chirp leads to pulse broadening in standard single-mode fibers.

2.2.2 Optical Modulators

Direct modulation of semiconductor lasers provides a compact and cost-effective means of converting electrical signals to optical signals. However, there are some limitations on direction modulation. The modulation speed is dependent on the dynamics of electron–photon interaction in semiconductor lasers, limiting the bit rate of direction modulation up to 10 Gb/s (some specially designed semiconductor lasers can be modulated up to 40 Gb/s, but they are rarely used in practice). The frequency chirp associated with intensity modulation broadens the spectrum of the optical signal; when combined with fiber group velocity dispersion, laser chirp limits the transmission distance. For high-bit-rate long-reach communication systems, external modulation is most commonly used. With external modulation, semiconductor lasers are used as a light source operating with fixed injection current and constant optical output power, and an optical modulator is used to modulated the CW light. Optical modulators can operate with bit rates as high as 100 Gb/s and produce less chirp than semiconductor lasers. Therefore, optical transmitters with external modulation provide better performance than direct modulation, and such transmitters may be used in high-bit-rate (e.g., 10 Gb/s) or long-reach (>20 km) PONs.

Light modulation can be achieved by changes in amplitude, phase, polarization, or frequency of the incident light waves through changing the refractive index, absorption coefficient, or the direction of the light transmitted in the external modulator. Among these, frequency modulation is very difficult to create, due to the great energy needed, and is rarely used in practice (only in coherent communication systems). Polarization of the light wave is relatively difficult to control during its propagation in common fibers (unless polarization-maintaining fiber is used), so polarization modulation is not used in commercial systems. Phase modulation is used primarily in advanced systems (e.g., DPSK systems) for high-speed long-haul optical communications. Most practical communication systems use an intensity modulation–direct detection (IM-DD) scheme. To achieve intensity modulation, the electroabsorption and electrorefraction effects are commonly used. The electroabsorption effect changes the absorption of the device and thus results in amplitude modulation (intensity modulation). Semiconductor electroabsorption modulators have been widely used in optical communication systems. The *electrorefraction effect* (also called the *electrooptic effect*) changes the index of refraction and leads to phase modulation; it can be combined with a Mach–Zehnder interferometer or a directional coupler to achieve intensity modulation. A lithium niobate modulator, the most commonly used optical modulator, is based on the electrorefraction effect.

2.2.2.1 Lithium Niobate Modulators Lithium niobate is a nonlinear optical crystal with a $3m$ point-group symmetry. The refractive indices of $LiNbO_3$ have a uniaxial form ($n_o = n_x = n_y = 2.297$ and $n_e = n_z = 2.208$), and its linear electrooptic

coefficient is given by

$$\mathbf{r} = \begin{bmatrix} 0 & -r_{22} & r_{13} \\ 0 & r_{22} & r_{13} \\ 0 & 0 & r_{33} \\ 0 & r_{51} & 0 \\ r_{51} & 0 & 0 \\ -r_{22} & 0 & 0 \end{bmatrix}, \tag{2.116}$$

where $r_{13} = 8.6 \times 10^{-12}$ m/V, $r_{22} = 3.4 \times 10^{-12}$ m/V, $r_{33} = 30.8 \times 10^{-12}$ m/V, and $r_{51} = 28.0 \times 10^{-12}$ m/V. Since r_{33} is the largest coefficient, an applied electric field along the z direction will be the most efficient for optical modulation. Therefore, for $E_x = E_y = 0$, the index ellipsoid of the lithium niobate crystal becomes

$$x^2\left(\frac{1}{n_o^2} + r_{13}E_z\right) + y^2\left(\frac{1}{n_o^2} + r_{13}E_z\right) + z^2\left(\frac{1}{n_e^2} + r_{33}E_z\right) = 1. \tag{2.117}$$

Due to electrorefraction, the new refractive indices are now given by

$$n_x = n_y \approx n_o - \frac{1}{2}n_o^3 r_{13}E_z, \tag{2.118}$$

$$n_z \approx n_e - \frac{1}{2}n_e^3 r_{33}E_z. \tag{2.119}$$

Therefore, under an electrical field in the z direction, the lithium niobate crystal remains uniaxial and the optical axis remains unchanged, but the index ellipsoid is deformed by the electric field. Light propagating along z will experience the same phase change regardless of its polarization state. However, light propagating along the x or y direction will experience a different phase change, depending on its polarization. For an X-cut lithium niobate crystal, as shown in the Figure 2.8, two electrodes are placed symmetrically on both sides of the waveguide such that the bias field is along the z direction. An incident optical signal with TE polarization will transmit as

$$\mathbf{E}_{\text{TE}} = \hat{z}E_0 e^{-jk_0\left(n_e - n_e^3 r_{33}E_z/2\right)y}. \tag{2.120}$$

Similarly, for TM polarization,

$$\mathbf{E}_{\text{TM}} = \hat{x}E_0 e^{-jk_0\left(n_o - n_o^3 r_{13}E_z/2\right)y}. \tag{2.121}$$

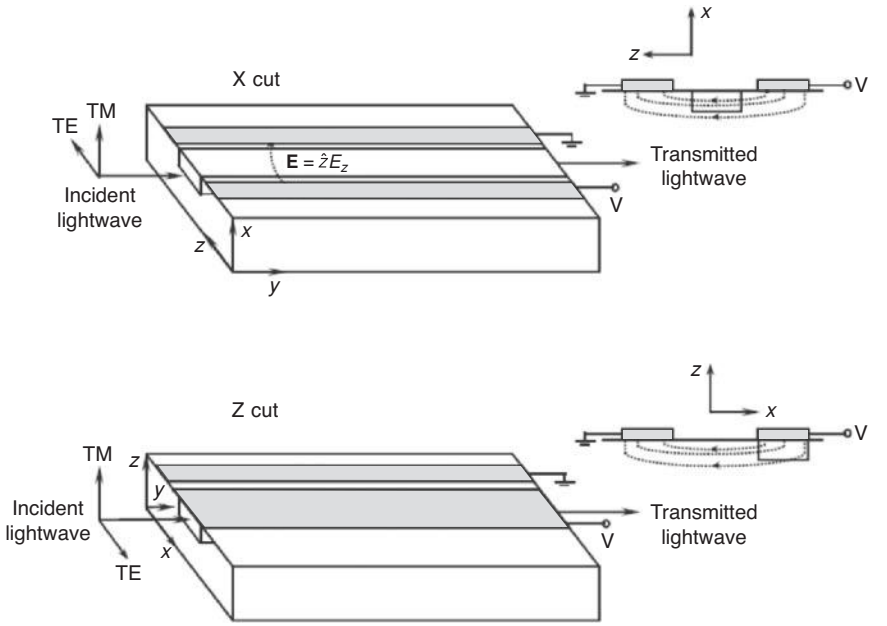


FIGURE 2.8 Structure of X and Z-cut lithium niobate phase modulators. To induce maximum refractive index change, an electric field in the z direction is preferred in lithium niobate modulators. For X-cut lithium niobate modulators, two electrodes are placed on each side of the optical waveguide. For Z-cut lithium niobate modulators, one electrode is placed on the top of the optical waveguide.

Since $r_{33} > r_{13}$, TE polarization is more efficient for phase modulation in the case of an X-cut lithium niobate crystal. For a Z-cut lithium niobate crystal, the electrodes are placed such that the waveguide is below one of the electrodes where the field is perpendicular to the Z-cut surface. In this case, the optical transmission for TE polarization will be

$$\mathbf{E}_{\text{TE}} = \hat{x} E_0 e^{-jk_0(n_o - n_o^3 r_{13} E_z/2)y}, \tag{2.122}$$

and for TM polarization,

$$\mathbf{E}_{\text{TM}} = \hat{z} E_0 e^{-jk_0(n_e - n_e^3 r_{33} E_z/2)y}. \tag{2.123}$$

In this case, TM polarization is preferred for most efficient phase modulation. The analysis above reveals that lithium niobate modulators have polarization-dependent response.

Lithium niobate modulator employs a Mach–Zehnder interferometer structure to convert the phase modulation (due to electrorefraction) to intensity modulation, as shown in Figure 2.9. Light input to the modulator is via a single-mode waveguide. A

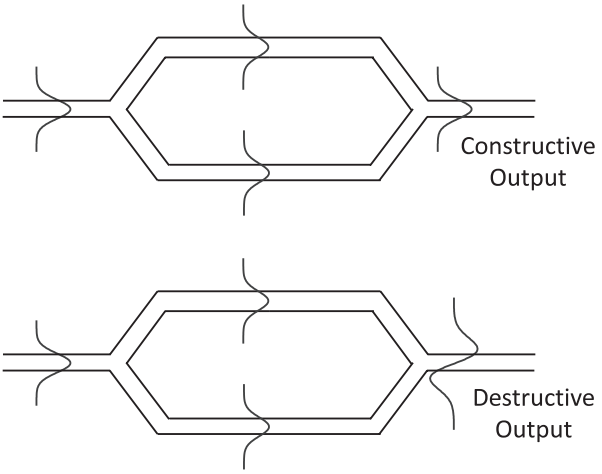


FIGURE 2.9 Intensity modulator with Mach–Zehnder interferometer structure. If the phase change in two arms of Mach–Zehnder interferometer is zero, constructive interference occurs at the output, and thus the output optical power is high. If the phase change in two arms of Mach–Zehnder interferometer is π, destructive interference occurs at the output, and thus the output optical power is low.

beamsplitter divides the light into two equal beams that travel through the two arms of a Mach–Zehnder interferometer. By applying a voltage to the electrodes, the effective path lengths can be varied. In an ideally designed modulator of this type, the path lengths L and guide characteristics are identical, so that with no applied voltage the split beams recombine in the output waveguide to produce the lowest-order mode once more. If an electric field is applied so as to produce a phase change of π radians between the two arms, the recombination results in an optical field that is zero at the center of the output waveguide, corresponding to the first-order mode. If the output waveguide is a single-mode guide, identical to the input guide, the first-order mode is cut off and dissipates rapidly over a short length by substrate radiation. Thus, the modulator can be switched from a transmitting to a nontransmitting state by application of a voltage. With proper choice of the polarization of the incident wave and the electrode design, the power transmitted, P_{out}, is given by

$$P_{\text{out}} = \frac{1}{4}\left|e^{j\beta_1 L} + e^{j\beta_1 L}\right| P_{\text{in}} = \cos^2(\Delta\beta L/2)\, P_{\text{in}}, \tag{2.124}$$

where P_{in} is the optical input power, $\Delta\beta = \beta_1 - \beta_2$, and β_1 and β_2 are the propagation constants in the two branches of the Mach–Zehnder interferometer, respectively. The difference between the propagation constants in two branches is given by

$$\Delta\beta = \frac{2\pi\,\Delta n}{\lambda_0}, \tag{2.125}$$

where λ_0 is the optical wavelength and Δn is the refractive index change resulting from electrorefraction.

Modern lithium niobate modulators can operate over a wide wavelength region and achieve an extinction ratio of more than 20 dB and/or a modulation bandwidth of about 100 Gb/s. However, lithium niobate modulators are bulky and require large modulation voltages. The frequency chirp produced by lithium niobate modulators can be very small, and if a symmetric push-pull modulation scheme (both branches of the modulator are applied with modulation voltages, one with positive voltage and the other with negative voltage) is used, zero-frequency chirp can be achieved.

2.2.2.2 Electroabsorption Modulators The electroabsorption effect is the effect with which the absorption coefficient changes with applied electric field. There are two main mechanisms for electroabsorption in semiconductors: the Franz–Keldysh effect and the quantum-confined Stark effect. Both of these electroabsorption effects are seen near the bandgap of semiconductors when electric fields are applied. The *Franz–Keldysh effect* is exhibited in conventional bulk semiconductors, whereas the quantum-confined Stark effect happens in quantum well structures. In fact, the quantum-confined Stark effect can be shown to be the quantized version of the Franz–Keldysh effect.

The concept of the Franz–Keldysh effect is shown in Figure 2.10. With an applied electrical field, we have a potential that varies linearly with distance. The

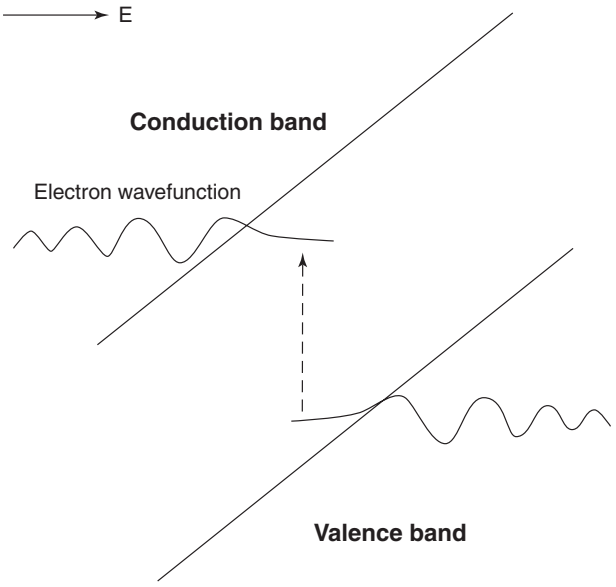


FIGURE 2.10 Franz–Keldysh effect. The sum of the band edge energy and the penetration of an electron wavefunction into the bandgap lead to the absorption of photons with less energy than the bandgap energy of the semiconductor.

wavefunctions for the electrons or the holes become Airy functions and "tunnel" into the bandgap region. Thus, overlap for electron and hole wavefunctions is possible even for photon energies lower than the bandgap energy, hence allowing optical absorption below the bandgap energy. However, in the spectral region just below the bandgap, the effects of excitons are important and possibly dominant in the optical absorption spectrum. In the presence of an electric field, the exciton can be field-ionized rapidly, leading to lifetime broadening of the exciton absorption peak.

In quantum wells, the electroabsorption for electric fields perpendicular to the quantum well layers is quite distinct from that in bulk semiconductors. While excitons are polarized by the electric field, quantum wells prevent excitons from field ionization. Therefore, instead of being broadened by the electric field, the exciton absorption peaks are strongly shifted by the field. The shifts can be tens of meV, leading to significant absorption below the bandgap energy. Since the shift of the energy levels with electric field in an atom is called the Stark effect, this shift of the exciton absorption peaks is called the *quantum-confined Stark effect*.

If a CW optical signal is incident at an electroabsorption modulator with length L, the output power of the modulator is given by

$$P_{\text{out}} = P_{\text{in}} \exp\left[-\alpha(V)L\right], \tag{2.126}$$

where α is the power absorption coefficient of the semiconductor material. As the absorption coefficient α can be varied by an applied electric field, the output optical power can be modulated by an external voltage. However, the modulation characteristics of electroabsorption modulators depend on the operating wavelength, and a small amount of frequency chirp always accompanies the intensity modulation. Commercially available electroabsorption modulators can produce a 20-dB extinction ratio with a small modulation voltage (<2 V). Bandwidths larger than 50 GHz have been realized with traveling-wave electroabsorption modulators. In addition, an electroabsorption modulator can be integrated with a semiconductor laser, leading to a compact device called an *electroabsorption-modulated laser*.

2.2.3 Transmitter Design

As a key component in optical communication systems, a transmitter must be designed and engineered with a number of specific requirements. Overall system performance, reliability, complexity, cost, size, and power consumption are just a few factors that come into play in transmitter design. More often than not, a designer has to balance all these requirements; in other words, trade-off among these factors is necessary for good design. The performance of an optical transmitter directly determines the overall system performance, especially the data rate and transmission distance. Key parameters of an optical transmitter include average transmitter power, modulation speed (bit rate), extinction ratio, optical output spectrum, intensity noise, and jitter.

Typically, for a low-cost system, direct modulation is used in optical transmitters. However, for a high-performance system, external modulation with optical modulators is utilized. As a general rule, telecommunication systems operating at 10 Gb/s

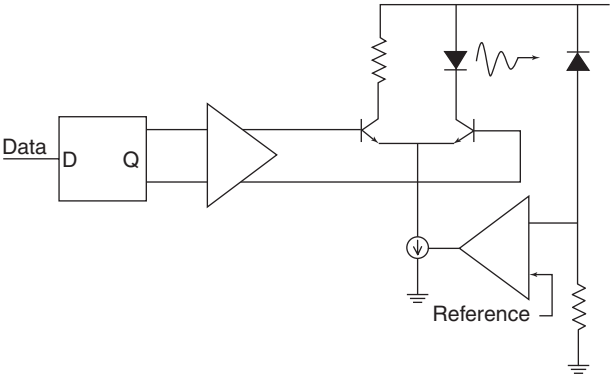


FIGURE 2.11 Block diagram of an optical transmitter. In an optical transmitter, data are first retimed by a flip-flop driven by the system clock. Then a laser/modulator driver, consisting of a predriver and a current steering output stage, modulates the laser or optical modulator. The output power of the laser is detected by a photodiode, and a feedback circuit is used to control the laser output power.

or higher use external modulation; 2.5-Gb/s systems use either direct modulation or external modulation, depending on the transmission distance; systems with data rates lower than 2.5 Gb/s use direct modulation. Short-reach optical links operating in the 1.3-μm region can also use direct modulation at 10 Gb/s, as the fiber dispersion is small.

Figure 2.11 shows typical structures of transmitters using direction modulation or external modulation. In direction modulation, a laser driver turns the semiconductor laser on and off, depending on the data input. The laser driver must be able to deliver tens of milliamperes of current with a very small rise and fall time. Designing a good laser driver at a high bit rate often presents an engineering challenge. For external modulation, the semiconductor laser is operated in CW mode, and a modulator driver turns the optical modulator on and off. The modulator drivers need to provide enough voltage swing to achieve a good extinction ratio. In addition, automatic bias control is often indispensable for modulator drivers. To keep the transmitted power constant (as required by most optical communication systems), a photodiode located at the back facet of the semiconductor laser monitors the optical power, and the photocurrent detected is used as a feedback control loop to stabilize the laser output power.

2.3 OPTICAL RECEIVERS

An optical receiver converts the transmitted optical signal back into an electrical signal and recovers the data carried by the optical signal. A key component in optical receivers is the photodetector, an optical–electrical converter that converts optical energy to electrical current. The electrical current from the photodetector is then amplified, and a clock and data recovery circuit recovers the data carried through the

optical communication system. In this section we introduce the operation principle of photodetectors and discuss the design of clock and data recovery circuits.

2.3.1 Photodetectors

Photodetectors absorb photons and generate electrical current that is proportional to optical power. As discussed in Section 2.2.1, when a photon has energy greater than the bandgap of a semiconductor material, the photon energy will be absorbed and an electron–hole pair will be generated in the semiconductor. With applied voltage, the electron–hole pair will carry electric current in an optical receiver. In optical communication systems, photodiodes (photodetectors with p-n and p-i-n junctions) are commonly used for photodetection.

2.3.1.1 PIN Photodiodes A PIN photodiode consists of an intrinsic semiconductor layer (in reality, a lightly doped semiconductor layer) sandwiched in the center of a p-n junction, as shown in Figure 2.12 Under reverse bias, a depletion layer across the p-n junction is created on both sides of the p-n junction and across the entire intrinsic layer. The depletion layer has a very low carrier density and high impedance, and hence a large built-in electric field is established in this region. When photons (with energy greater than the semiconductor bandgap energy) impinge on the PIN photodiode, electron–hole pairs are generated in the depletion layer. Because of the large electric field inside the depletion region, electron–hole pairs are separated and drift in opposite directions toward the p- or n- side, leading to electric current flow across the p-i-n junction. The photocurrent (generated by the PIN photodiode under light incidence) is proportional to the incident optical power, given by

$$I_p = R_d P_{\text{in}}, \tag{2.127}$$

where I_p is the photocurrent, P_{in} is the incident optical power, and R_d is known as the responsivity of the photodiode. The responsivity is related to the quantum efficiency by

$$R_d = \frac{q}{h\nu}\eta, \tag{2.128}$$

where the quantum efficienty η is defined as the ratio between the number of the photogenerated electrons and the number of incident photons. The responsivity can be improved by increasing the width of the intrinsic layer. Typically, the intrinsic layer

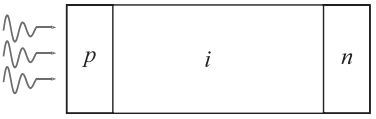


FIGURE 2.12 Structure of a PIN photodiode.

has a width of 20 to 50 μm to absorb all the incident photons, and the responsivity can be as high as 0.9 A/W. However, the speed of photodiode response is limited by the carrier transit time across the p-n junction, which is proportional to the width of the intrinsic region. The bandwidth of a photodiode is given approximately by

$$\mathrm{BW} = \frac{1}{2\pi\,(\tau_{\mathrm{tr}} + \tau_{\mathrm{RC}})}, \tag{2.129}$$

where τ_{tr} is the transition time and τ_{RC} is the RC time constant induced by electrical parasitics. To improve the bandwidth of a photodiode, it is desirable to have a thin depletion layer. In practice, pin photodiodes with bandwidth larger than 100 GHz have been made by using a very thin absorption layer. However, a thinner absorption layer reduces the responsivity of the photodiode. Therefore, there exists a trade-off between the photodiode bandwidth and responsivity. To improve photodiode performance, a Fabry–Perot cavity can be designed around the p-i-n junction, resulting in enhanced photon absorption with reduced intrinsic layer width. Alternatively, a waveguide photodiode can be implemented with an edge-coupled optical signal. While the optical signal is traveling along the waveguide, the carrier moves in a direction perpendicular to the waveguide. Therefore, photodiode bandwidth and responsivity can be optimized separately with little compromise. Modern photodiode design can achieve a bandwidth over 300 GHz with traveling-wave photodiodes, where the electrode structure is designed to support traveling RF waves and thus reducing the impact of the parasitic effect of electrodes on photodiode bandwidth.

2.3.1.2 Avalanche Photodiodes The receiver optical power is typically very low for long-haul optical communication systems. Therefore, photodiodes with a high responsivity value are often preferred to generate large photocurrent. However, the responsivity of a PIN photodiode is limited by $R_d = q/h\nu$ for the maximum quantum efficiency of $\eta = 1$. An avalanche photodiode (APD) can provide internal current gain and thus can achieve a higher responsivity value.

The internal current gain in avalanche photodiodes is provided through impact ionization under a large electric field. Photogenerated carriers in a photodiode are accelerated by the electric field in the depletion region, and thus these carriers acquire kinetic energy from the electric field. When these carriers collide with the crystal lattice, they will lose some energy to the crystal. If the kinetic energy of a carrier is greater than the bandgap energy, the collision can free a bound electron. The free electron and hole thus created can themselves acquire enough kinetic energy to cause further impact ionization. The result is an avalanche, with the numbers of free carriers growing exponentially as the process continues. This cumulative impact ionization produces a large number of free carriers in the depletion region, which is greater than the number produced by photoionization. Therefore, the total current is greater than the primary photocurrent, and current amplification is achieved inside the photodiode.

The ratio of the total current to the primary photocurrent is the current amplification of the APD,

$$M = \frac{I_{\text{APD}}}{I_{\text{ph}}}, \tag{2.130}$$

where I_{APD} is the total current generated by the APD and I_{ph} is the current generated by photon absorption. The responsivity of the APD is thus given by

$$R_{\text{APD}} = \frac{\eta q}{h\nu} M. \tag{2.131}$$

The structure of an avalanche photodiode is similar to that of the pin photodiode with an avalanche region added on one end of the intrinsic region, as shown in Figure 2.13. The avalanche region must have a high internal electric field for the avalanche process to occur. Let α and β represent the ionization coefficients for electrons and holes, respectively. Both α and β will be functions of the material, the electric field strength, and the temperature. For injected electrons, the current multiplication factors are

$$M_e = \frac{(1-k)\exp[\alpha w_a(1-k)]}{1-k\exp[\alpha w_a(1-k)]}, \tag{2.132}$$

and for injected holes,

$$M_h = \frac{(1-1/k)\exp[\beta w_a(1-1/k)]}{(1-1/k)\exp[\alpha w_a(1-1/k)]}, \tag{2.133}$$

where $k = \beta/\alpha$ and w_a is the width of the avalanche amplification region.

The avalanche process can extend the duration of the impulse response of avalanche photodiode well beyond that attributable to the one-pass transit times of electrons and holes. For the case of electron injection, the total duration is the sum of three times: the time required for the most distant primary electron to reach the avalanche region, the duration of the avalanche process, and the time required for all holes produced during avalanche to move back across the avalanche and intrinsic

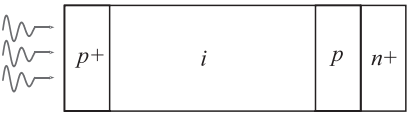


FIGURE 2.13 Structure of an APD. Optical absorption occurs in the intrinsic layer, and the large electric field in the p layer creates an avalanche effect.

regions. The multiplication factor of the avalanche photodiode shows a frequency dependence as

$$M = \frac{M_0}{1 + jM_0\tau_1\omega}. \tag{2.134}$$

Therefore, the 3-dB bandwidth of a avalanche photodiode is given by

$$\text{BW} = \frac{1}{2\pi M_0\tau_1}, \tag{2.135}$$

where τ_1 is the effective transit time of the avalanche region. The bandwidth is inversely proportional to the low-frequency gain M_0. This relation shows the trade-off between the APD gain and the bandwidth.

The noise associated with the avalanche current-multiplication process consists of the sum of the amplified shot noise of the primary photocurrent and the excess noise produced by the multiplication process. The excess noise is represented by an excess noise factor F, which is defined as the ratio of the total noise associated with I_{APD} to the noise that would exist in I_{APD} if the multiplication process produced no excess noise (i.e., F is the total noise divided by the multiplied shot noise). The mean-squared total noise is F times the multiplied shot noise, given by

$$\langle i_n^2 \rangle = 2qIM^2FB, \tag{2.136}$$

where I is the sum of the primary photocurrent and the dark current. When the avalanche process is initiated by electrons, the excess noise factor F is given by

$$F_e = M_e\left[1 - (1-k)\frac{(M_e - 1)^2}{M_e^2}\right]. \tag{2.137}$$

Thus, a small value of $k = \beta/\alpha$ is preferred for a smaller excess noise factor. However, when the avalanche process is initiated by holes,

$$F_h = M_h\left[1 - (1 - 1/k)\frac{(M_h - 1)^2}{M_h^2}\right]. \tag{2.138}$$

Thus, for hole injection, a smaller $1/k = \alpha/\beta$ is preferred; that is, a larger impact ionization coefficient for holes leads to a small access noise factor F_h.

2.3.2 Optical Receiver Design

An optical receiver consists of a photodiode, a baseband preamplifier, a limiting amplifier, and a clock and data recovery circuit, as shown in Figure 2.14. Typical requirements of optical receivers include large bandwidth, large dynamic range, and high receiver sensitivity. The noise, gain, and bandwidth of the preamplifier and

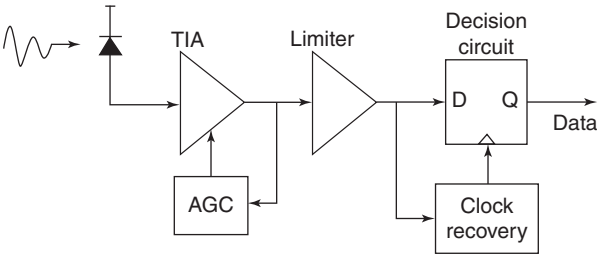


FIGURE 2.14 Block diagram of an optical receiver. TIA, transimpedance amplifier; limiter, limiting amplifier.

the limiting amplifier directly affect the receiver sensitivity and data rate of the overall communication system. The clock and data recovery circuit must provide fast response, low jitter, and long run (consecutive identical digits) tolerance.

The photocurrent generated by the photodiode is usually very low, so the signal is prone to noise contamination. The preamplifier amplifies the photocurrent for further processing in later stages. To increase the receiver sensitivity, a large load resistor can be used to increase the input voltage to the preamplifier. In addition, a large impedance can reduce the thermal noise and improve the receiver sensitivity. However, the high load resistor will reduce the bandwidth of the optical receiver. Therefore, there exists a trade-off between the bandwidth and the receiver sensitivity for optical receivers. Occasionally, an equalizer is designed to improve the bandwidth of high-impedance optical receivers. However, the equalizer will introduce more noise, especially at high frequencies. In high-performance optical receivers, a transimpedance amplifier (as shown in Figure 2.14) is generally used to provide both large bandwidth and high receiver sensitivity. Because of the negative feedback, the input impedance of the preamplifier is reduced significantly, thus improving the bandwidth of the optical receiver. However, to improve the receiver sensitivity, the bandwidth of the transimpedance amplifier is designed to be 0.7 times the bit rate, which is a reasonable compromise between the noise and the intersymbol interference resulting from limited bandwidth.

The voltage level produced by the preamplifier is usually inadequate to drive the clock and data recovery circuit. Limiting amplifiers following the preamplifier are used to boost the signal level further. Therefore, limiting amplifiers are required to provide enough gain with negligible intersymbol interference. The limiting amplifiers are designed with a cascade of differential amplifiers with enough bandwidth and relative linear phase response.

Clock and data recovery circuits present another challenge in designing optical receivers. Figure 2.15 shows the architecture of clock and data recovery circuits. A phase-locked loop is commonly utilized to recover the clock. Then the output of the phase-locked loop samples and retimes the data, reducing the jitter and intersymbol interference. High data rate, stringent jitter, and loop bandwidth specifications require significant design efforts in clock and data recovery circuits.

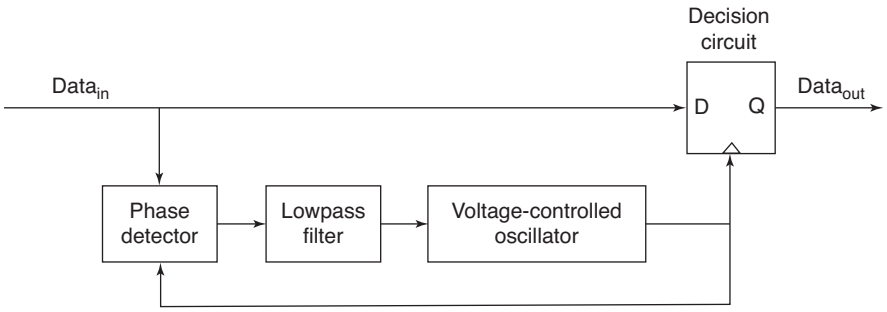


FIGURE 2.15 Block diagram of clock and data recovery circuits.

2.4 OPTICAL AMPLIFIERS

As optical signals propagate along optical fibers, the signal power is attenuated due to fiber loss. For reliable communication, optical signal power needs to be kept at a level above the receiver sensitivity. Traditionally, when an optical signal is weak, it is converted back to electrical form and the signal is amplified electronically. All the noise and distortion can be removed in the case of digital communications. The amplified waveform is then used to modulate a CW light wave to generate a clean optical signal with higher power for further transmission. This type of O-E-O (optical–electrical–optical) repeater was commonly used before optical amplifiers, especially erbium-doped fiber amplifiers, became dominant in long-haul optical communications. An optical amplifier provides optical gain through stimulated emission when the gain medium is pumped optically or electrically to achieve population inversion. Whereas O-E-O repeaters work only for single channels, optical amplifiers can amplify multiple incident optical signals simultaneously, and they are transparent to data format (e.g., bit rate and modulation format). Therefore, optical amplifiers provide a simpler and cost-efficient means to boost optical signal power for long-distance transmission. Optical amplifiers commonly used in optical communication systems are erbium-doped fiber amplifiers (EDFAs), semiconductor optical amplifiers (SOAs), and Raman amplifiers.

When a signal with low optical power propagates through an optical amplifier, each section of the amplifier adds optical power to the original signal through stimulated emission. Amplification of the optical signal is described by

$$\frac{dP(z)}{dz} = g_0 P(z), \tag{2.139}$$

where $P(z)$ is the optical power at a distance z from the input end and g_0 is the small-signal gain coefficient of the amplifier. Neglecting the gain saturation, the optical signal grows exponentially due to optical amplification,

$$P(z) = P_i \exp(g_0 z), \tag{2.140}$$

where P_i is the input optical power. At the output of the optical amplifier, the signal power is

$$P(L) = P_i \exp(g_0 L), \tag{2.141}$$

where L is the length of the optical amplifier. Thus, the power gain of the optical amplifier is given by

$$G = \exp(g_0 L). \tag{2.142}$$

This equation gives the amplifier gain under small-signal conditions. When the optical power is large (comparable to the saturation power P_s), the gain coefficient g of the optical amplifier is reduced as

$$g = \frac{g_0}{1 + P/P_s}. \tag{2.143}$$

Thus, the amplifier gain G decreases with an increase in the signal power. This phenomenon is called *gain saturation*. In this case, the amplification of optical signal is described by

$$\frac{dP(z)}{dz} = \frac{g_0 P(z)}{1 + P(z)/P_s}. \tag{2.144}$$

The optical power $P(z)$ can be determined by

$$\ln \frac{P(z)}{P(0)} + \frac{1}{P_s}[P(z) - P(0)] = g_0 z. \tag{2.145}$$

By using the initial condition, $P_i = P(0)$ and $P_o = P(L) = G P_i$, we have

$$\ln G + \frac{1}{P_s}\left(P_o - \frac{P_o}{G}\right) = g_0 L. \tag{2.146}$$

Hence, the amplifier gain under large-signal conditions is given by

$$G = 1 + \frac{P_s}{P_i} \ln \frac{G_0}{G}, \tag{2.147}$$

where $G_0 = \exp(g_0 L)$. This equation shows that the amplifier gain decreases from its unsaturated value G_0 when the optical power becomes comparable to the saturation power P_s. In practice, a parameter known as the *output saturation power* is used to characterize the gain saturation effect of optical amplifiers. The output saturation power is defined as the output power for which the amplifier gain G is reduced by a factor of 2 (3 dB) from its unsaturated value G_0. From eq. (2.147) we can find the

saturation output power as

$$P_{\text{out}}^{s} = \frac{G_0 \ln 2}{G_0 - 2} P_s. \tag{2.148}$$

While providing signal gain, optical amplifiers add amplified spontaneous emission (ASE) noise to its output, and thus degrade the signal-to-noise ratio (SNR) of the optical signal. The spectral density of spontaneous emission-induced noise is nearly constant (white noise) and can be written

$$S_{\text{sp}}(\nu) = (G - 1) n_{\text{sp}} h\nu, \tag{2.149}$$

where n_{sp} is called the *spontaneous emission factor* (or *population inversion factor*) of the amplifier. The spontaneous emission factor is given by

$$n_{\text{sp}} = \frac{N_2}{N_2 - N_1}, \tag{2.150}$$

where N_1 and N_2 are the densities of the ground and excited states, respectively. The effect of spontaneous emission is to add fluctuation to the amplifier optical power.

Similar to the electronic amplifier, the SNR degradation in an optical amplifier is quantified by amplifier noise figure, F_n, which is defined as the ratio of the input SNR to the output SNR. As the dominant noise source in an optical amplifier is the ASE noise, its noise figure can be found as

$$F_n = \frac{\text{SNR}_{\text{in}}}{\text{SNR}_{\text{out}}} = \frac{2n_{\text{sp}}(G - 1)}{G} \approx 2n_{\text{sp}}. \tag{2.151}$$

As the spontaneous emission factor n_{sp} is always larger than unity, the SNR of the amplified signal is degraded by a factor larger than 3 dB. Most optical amplifiers have a noise figure of 4 to 8 dB. The definition of noise figure above parallels that used for RF amplifiers; however, there are differences that require clarification. All of the noise on the output signal is assumed to be contributed by the amplifier, which implies that the input signal itself does not contribute to the output noise. The noise figure therefore relates to the spectral density of the added ASE.

An optical amplifier can be used in various locations on a fiber optical communication link. Depending on its location, it performs different functions (amplifying optical signal in general) and has different performance requirements. Typical applications of optical amplifiers include as in-line amplifiers, booster amplifiers, and preamplifiers. An in-line amplifier is used in the transmission path to amplify the weakened optical signal due to fiber loss. High gain, low noise, and high output power (high saturation power) are required. A booster amplifier is operated in the saturation region and used immediately after the transmitter to boost its power. For booster amplifiers, high-gain, high-output power (high saturation power) is desired. A preamplifier is used in front of the receiver to boost the optical signal power for

detection. High gain and low noise are preferred for preamplification, and an in-line narrow bandpass filter is often used to reduce the ASE noise.

Ideally, an optical amplifier should have high-gain, high-output power (no power saturation or high saturation power) and low noise. For WDM applications, a large bandwidth (wide gain spectrum) and flat gain spectrum are also required. The transient behavior of the optical amplifier is also an important consideration for system applications, and gain must be insensitive to variants in the input power of the signal.

2.4.1 Rare-Earth-Doped Fiber Amplifiers

Rare-earth elements such as erbium, neodymium, thallium, and ytterbium can be doped into optical fibers to make fiber amplifiers operate in different wavelengths from 0.5 to 3.5 μm. Amplifier characteristics such as the operating wavelength and the gain bandwidth are determined by the dopants rather than the silica fiber (which plays the role of a host medium). EDFA is, by far, the most widely used fiber amplifier, because its operating wavelength is coincident with the low-fiber-loss window at the 1.55-μm band. Because EDFAs can amplify multiple wavelengths (tens of wavelength or even over 100 wavelengths), their deployment in 1990 has led to commercial WDM optical communication systems with capacity exceeding 100 Gb/s. Commercial EDFAs can provide a gain of over 30 dB and 20 dBm of output power with about 35 nm of bandwidth and a very low noise figure (4 to 6 dB).

Typically, an EDFA consists of a length of optical fiber doped with Er^{3+} ions and a suitable optical pump. The gain characteristics of EDFAs depend on the pumping scheme as well as on other dopants, such as the germanium and aluminum present within the fiber core. The amorphous nature of silica broadens the energy level of erbium ions into bands. Many transitions can be used to pump the EDFA. Efficient pumping is achieved using semiconductor lasers operating near the 0.98- and 1.48-μm wavelengths. The pumping scheme can be either a forward pump or a backward pump, as shown in Figure 2.16. The performance is nearly the same in the two pumping configurations when the signal power is small enough for the amplifier to remain unsaturated. In the saturation regime, the power-conversion efficiency is generally better in the backward-pumping configuration, mainly because of the important role played by the amplified spontaneous emission. In the bidirectional pumping configuration, the amplifier is pumped in both directions simultaneously by using two semiconductor lasers located at the two fiber ends. This configuration requires two pump lasers but has the advantage that the population inversion, and hence the small-signal gain, is relatively uniform along the entire amplifier length. In some high-performance EDFAs, two-stage pumping is used: forward pumping in the first stage for high gain and low noise amplification, and backward pumping in the second for high output powers. In addition, an optical isolator is placed immediately after the first amplifying stage (critical for a low noise figure) to prevent degradation of the first-stage performance due to the backward-propagation ASE noise from the second stage. An optical filter can be placed after the first stage to prevent gain saturation caused by the ASE peak around 1530 nm. Gain flattened filters and/or dispersion compensation modules are usually inserted between these two stages.

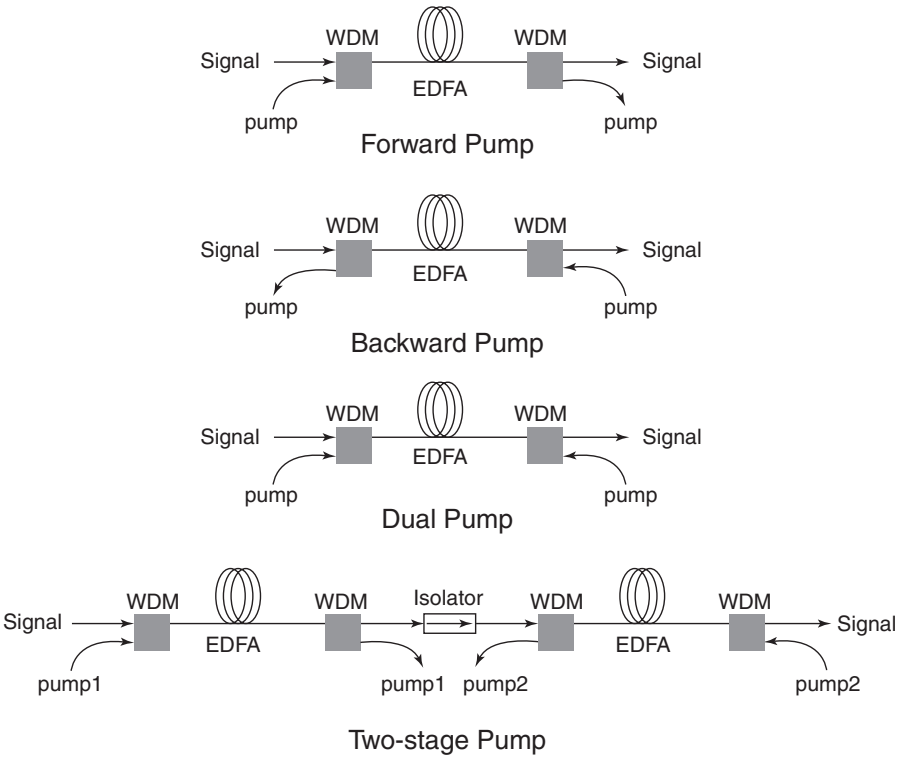


FIGURE 2.16 Various EDFA pump schemes.

Most EDFAs operate in the C-band (1528 to 1562 nm), which corresponds to the peak of the $^4I_{13/2}$–$^4I_{15/2}$ erbium energy-level transition. An erbium-doped fiber, however, has a relatively long tail to the gain shape, extending well beyond this range, to about 1605 nm, which corresponds to the tail of the $^4I_{13/2}$–$^4I_{15/2}$ energy-level transition. This has stimulated the development of EDFA for the L-band from 1565 to 1625 nm (note that part of the L-band in the 1610 to 1625 nm region is not covered by L-band EDFA). Because L-band EDFA exploits the tail of the erbium gain band, emission and absorption coefficients are three to four times smaller than in the C-band. In addition, L-band EDFA operates at low average inversion in order to minimize the intrinsic gain ripple. The comparatively flatter intrinsic gain spectrum simplifies the design and implementation of gain-flattening filters. These two operation conditions require erbium-doped fiber that is four to five times longer than that is required for the C-band (assuming typical erbium concentration levels) or high erbium doping concentration (as high as 1900 ppm, compared to 300 to 500 ppm for typical EDFAs), and the pump powers required for L-band EDFAs are much higher than that are required for their C-band counterparts. Due to the smaller absorption cross sections in the L-band, these amplifiers also have higher amplified spontaneous emission.

Whereas EDFAs can only be used for light-wave systems operating near 1.55 μm, fiber amplifiers operating at 1.3-μm wavelength can be realized by doping ZBLAN fiber (a fluoride fiber fabricated with ZrF_4–BaF_2–LaF_3–AlF_3–NaF) with praseodymium (Pr^{3+}) ions. Pr-doped fiber amplifiers can provide 30 dB gain and 20 nm bandwidth around 1.3-μm wavelength with 1-W pump power in the wavelength region near 1.01 m. The pumping efficiency of Pr-doped fiber amplifiers is very small (<0.2 dB/mW) compared with that of EDFA (as high as 10 dB/mW). The saturation power of such amplifiers is very high (200 mW). Relatively high cost and the practical difficulty of working with fluoride fiber hinder their wide application.

2.4.2 Semiconductor Optical Amplifiers

A semiconductor optical amplifier has a structure very similar to that of a semiconductor laser except that it has no resonant cavity. Figure 2.17 shows the structure of a semiconductor optical amplifier. In a semiconductor optical amplifier, an active layer (i.e., gain medium) with lower bandgap energy and higher refractive index is sandwiched between a p material and an n material (which form a p-n junction). Therefore, a waveguide is effectively formed by these three layers of semiconductor materials. Incident optical signal travels along the gain medium and undergoes amplification. Two end facets of the device are coated with an antireflection layer, so the device will not lase under a large injection current. This type of device is called a *traveling-wave semiconductor optical amplifier*.

For a semiconductor optical amplifier with active layer length L, the amplifier gain can be found from standard Fabry–Perot interferometers and is given by

$$G(\lambda) = \frac{(1 - R_1)(1 - R_2) G_s}{\left(1 - G_s\sqrt{R_1 R_2}\right)^2 + 4G_s\sqrt{R_1 R_2}\sin^2\phi(\lambda)}, \tag{2.152}$$

where $G_s = \exp(gL)$ is the single-pass amplifier gain and R_1 and R_2 are the facet reflectivities. The phase shift is given by

$$\phi = \frac{2\pi\bar{n}L}{\lambda} + \frac{\alpha_c g_0 L}{2}\frac{P}{P + P_s}, \tag{2.153}$$

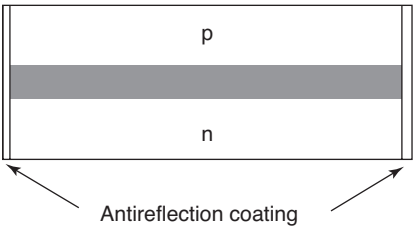


FIGURE 2.17 Structure of a semiconductor optical amplifier.

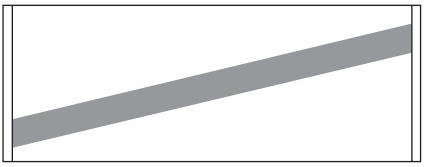
where $\bar{n}$ is the effective refractive index of the amplifier, α the linewidth enhancement factor, g_0 the unsaturated gain, and P and P_s the internal optical power and saturation power, respectively. For an ideal traveling-wave SOA, $R_1 = R_2 = 0$, and hence $G = G_s$. The gain spectrum of an ideal SOA depends only on the gain profile of the semiconductor material. However, in practice, the residual reflectivities can be made very small (<0.1%) but nonzero. In this case, the gain spectrum shows a small ripple where the maximum and minimum gain ratio is given by

$$GR = \frac{G_{\max}}{G_{\min}} = \frac{1 + G_s\sqrt{R_1 R_2}}{1 - G_s\sqrt{R_1 R_2}}. \tag{2.154}$$

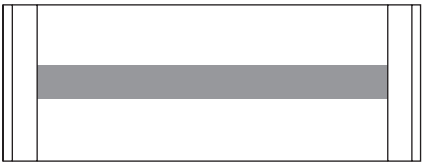
The gain ripple increases with increasing gain and increasing facet reflectivities. In addition to antireflection coating, angled- or buried-facet structure (as shown in Figure 2.18) is also commonly used to reduce the gain ripple. In an angled-facet structure, the reflection wave is separated from the incident wave, and thus the reflectivities are effectively reduced. As an alternative, buried-facet structure spreads the optical field before it is reflected at the facet, so little of the power of the reflected wave is coupled into the active layer. Commercially available traveling-wave SOAs can provide a peak gain of 20 to 30 dB with a gain ripple below 1 dB. The peak gain wavelength depends on the active layer material, and the gain bandwidth could be 30 to 80 nm. A typical output saturation power of SOAs is 5 to 15 mW.

In addition to optical amplification, SOAs also add ASE noise at its output. The noise figure of an SOA is given by

$$F_n = 2 \cdot \frac{N}{N - N_0} \cdot \frac{g}{g - \alpha_{\text{int}}}, \tag{2.155}$$



Angled facet



Buried facet

FIGURE 2.18 Structures to reduce reflection in SOAs.

where N is the carrier density, N_0 the transparency carrier density, and α_{int} the internal loss coefficient. Residual facet reflectivities can increase the noise figure by a factor of $1 + R_1 G$, where R_1 is the reflectivity of the input facet. For most traveling-wave SOAs, $R_1 G \ll 1$, and this additional factor is often neglected. Typically, $F_n = 5 - 8$ for SOAs.

Compared to EDFA, SOAs can be designed to operate at any wavelength, depending on the selection of gain medium. SOAs can also be integrated with a laser (as a booster amplifier) or photodiode (as a preamplifier). However, an SOA is inherently a polarization-sensitive device. The amplifier gain for TE or TM mode could differ as much as 6 dB. With proper design (e.g., square waveguide or polarization diverse structure), the polarization dependence could be reduced to about 1 dB. SOAs also show significant nonlinearity when operating with large signal powers. When used for multiwavelength amplification, SOAs suffer from interchannel crosstalk. Three prominent nonlinear effects are cross-gain modulation, cross-phase modulation, and four-wave mixing. Even though these nonlinear effects are detrimental for optical amplification applications, they are often employed for optical signal processing, such as wavelength conversion and all-optical regeneration.

2.4.3 Raman Amplifiers

Raman amplifier utilizes stimulated Raman scattering (as discussion in Section 2.1.5.1) to transfer the energy of optical pump to an optical signal at a longer wavelength. Under high pump power, the optical fiber itself can be used as a gain medium for optical amplification. The optical amplification gain of fiber Raman amplifiers depends on the properties of the transmission fiber, such as pump absorption, effective area, and Raman gain coefficient. The gain coefficient of a fiber Raman amplifier is proportional to the pump intensity and it can be written as

$$g(\omega) = g_r(\omega) \frac{P_p}{a_p}, \tag{2.156}$$

where $g_r(\omega)$ is the Raman gain coefficient and a_p is the cross-sectional area of the pump beam inside the fiber. The Raman spectrum has a gain peak at a wavelength shift of about 13 THz (about 100 nm for a 1.45-μm pump) and the gain bandwidth is about 6 THz (about 50 nm in a 1.55-μm region). The large bandwidth and low noise figure of fiber Raman amplifiers make them attractive for fiber optic communication applications. However, a relatively large pump power (>1 W) is required to realize a large amplification gain.

For small-signal amplification and co-propagation of pump and optical signal, the optical gain of a Raman amplifier is given by

$$G_r = \exp(g_0 L). \tag{2.157}$$

The small-signal gain g_0 is given by

$$g_0 = \frac{g_r P_{p0} L_{eff}}{\alpha_p L}, \tag{2.158}$$

where P_{p0} is the input pump power and the effective length of the amplifier is $L_{\text{eff}} = \left[1 - \exp(-\alpha_p L)\right]/\alpha_p$, where α_p is the fiber loss at the pump wavelength. For $\alpha L_{\text{eff}} \gg 1$, $L_{\text{eff}} \approx 1/\alpha_p$.

Raman amplifiers offer great flexibility in choosing the wavelength of the amplification region (provided that a suitable pumping laser is available). EDFAs provides gains in the C- and L-bands (1530 to 1630 nm). Thus, Raman amplifier can provide gain in the O-band (1260 to 1360 nm) and potentially open up other bands for WDM, such as the S-band (1460 to 1530 nm) and the E-band (1360 to 1460 nm). With the use of multiple pumps at different wavelengths and different powers simultaneously, we can tailor the Raman gain shape. Today the most popular use of Raman amplifiers is to complement EDFAs by providing additional gain in a distributed manner in an ultra-long-haul system.

Raman amplification is inherently low-noise. Spontaneous emission noise is relatively low. The major noise source is due to Rayleigh scattering of the pump signal in the fiber and spontaneous Raman scattering. However, Raman amplifiers exhibit significant cross-gain saturation. Raman scattering is inherently a very fast process, and the Raman gain responds to the pump power almost instantaneously. Fluctuations in pump power will cause the gain to vary and will appear as crosstalk to the desired signal. Therefore, for Raman amplifiers, it is important to keep the pump at a constant power level. Having the pump propagate in the direction opposite to the signal helps dramatically because fluctuations in pump power are then averaged over the propagation time over the fiber. Another major concern with Raman amplifiers is crosstalk between the WDM signals due to Raman amplification. A modulated signal at a particular wavelength depletes the pump power and affects the gain seen by the wavelengths. Again, having the pump propagate in the direction opposite to the signal reduces this effect dramatically.

2.5 PASSIVE OPTICAL COMPONENTS

Besides optical fiber, optical transmitters, optical receivers, and optical amplifiers, many passive optical components are used in optical communication systems. Commonly used passive optical components in passive optical networks include optical couplers, optical filters, and multiplexers. Optical splitters are used to split optical power into multiple fibers; optical filters are utilized to filter out optical signals with specific wavelengths; and optical multiplexers can combine or separate optical signals with different wavelengths. Passive optical components do not require any external power to perform their functionalities, and hence they are more reliable and require less maintenance effort.

2.5.1 Directional Couplers

A directional coupler consists basically of parallel optical waveguides sufficiently closely spaced that energy is transferred from one to the other by optical tunneling. This energy is transferred by a process of synchronous coherent coupling between the overlapping evanescent tail of the modes guided in each waveguide. Photons of the

driving modes, say in guide 0, transfer into the driven mode in guide 1, maintaining phase coherence as they do. This process occurs cumulatively over a significant length; hence, the light must propagate with the same phase velocity in each channel in order for this synchronous coupling to occur. The fraction of the power coupled per unit length is determined by the overlap of the modes in the separate channels.

Optical coupling between two waveguides can be described by coupled mode theory. Assume that $A_1(z)$ and $A_2(z)$ are optical fields propagating in two waveguides of a directional coupler. The coupled mode equations for these two optical fields can be written as

$$\begin{aligned} \frac{dA_1(z)}{dz} &= -j\beta A_1(z) - j\kappa A_2(z), \\ \frac{dA_2(z)}{dz} &= -j\beta A_2(z) - j\kappa A_1(z), \end{aligned} \tag{2.159}$$

where two waveguides are assumed to be identical and lossless, and β is the propagation constant in the waveguides. The coupling coefficient κ depends on the shape and separation between the waveguides. Solving the coupling wave equations, we have

$$A_1(z) = A\cos\kappa z \exp(j\beta z), \tag{2.160}$$

$$A_2(z) = -jA\sin\kappa z \exp(j\beta z). \tag{2.161}$$

Thus, the optical power flow in the waveguides is given by

$$P_1(z) = |A_1(z)|^2 = A^2\cos^2\kappa z, \tag{2.162}$$

$$P_2(z) = |A_2(z)|^2 = A^2\sin^2\kappa z. \tag{2.163}$$

and the power coupling coefficient is given by

$$C = \frac{P_2(z)}{P_1(0)} = \sin^2\kappa z. \tag{2.164}$$

The optical power transfers back and forth between the two waveguides as a function of length, as shown in Figure 2.19. In addition, there exists a distinct phase difference between the amplitude of the fields in the two waveguides. Initially, the optical field $A_1(z)$ in waveguide 2 lags 90° behind the optical field $A_1(z)$ in waveguide 1, and the power is transferred from waveguide 1 to waveguide 2. After $\kappa z = \pi/2$, all of the optical power has been transferred to waveguide 1. Then for $\pi/2 < \kappa z < \pi$, the optical field in waveguide 1 lags 90° behind that in waveguide 2, and so on. This phase relationship results from the basic mechanism that produces the coherent transfer of energy. The field in the driving guide causes a polarization in the dielectric material which is in phase with it and which extends into the region between guides because of the mode tail. This polarization then acts to generate energy in the mode of the driven guide. It is a basic principle of field theory that generation occurs when polarization leads the field, while dissipation occurs when polarization lags the field. Thus, the

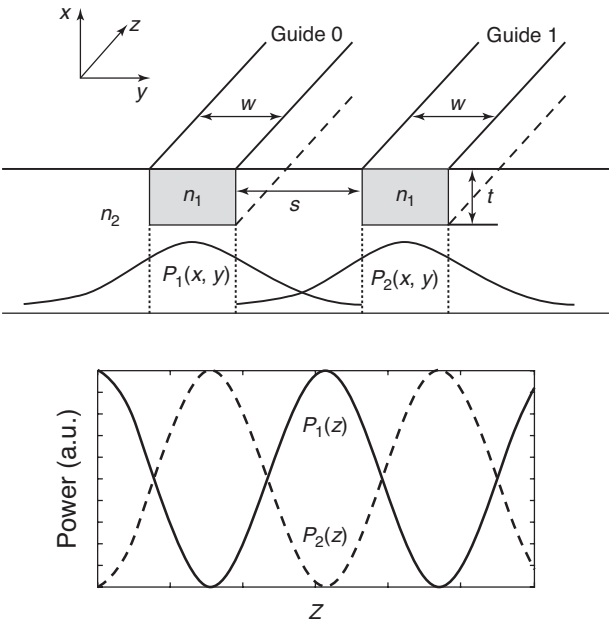


FIGURE 2.19 Optical power coupling between two parallel waveguides. $P_1(x, y)$ and $P_2(x, y)$ represent the mode distribution of two waveguides. Because of significant overlap between $P_1(x, y)$ and $P_2(x, y)$, optical power can be coupled back and forth between these two waveguides. The power variation along the z direction is shown by $P_1(z)$ and $P_2(z)$.

lagging field in the driven guide is to be expected. Because of this definite phase relationship, the dual-channel coupler is a directional coupler. No energy can be coupled into a backward wave traveling in the $-z$ direction in the driven waveguide.

Optical couplers (splitters) can be created by fused fibers. For example, two pieces of fiber are placed side by side in a flame, drawn, and fused together to form a 2×2 fiber coupler. In this coupler, a long tapered section is created in each fiber, where the core diameter is small. Thus, the normalized frequency V is reduced in the tapered section, and the field is more weakly confined and extends outside the cladding area so that efficient coupling occurs between the two fibers. The power coupling coefficient from one fiber to the other is

$$C = F^2 \sin^2 \frac{\xi z}{F}, \tag{2.165}$$

where

$$F^2 = \frac{1}{1 + (234a^3/\lambda^3)(\Delta a/a)^2}, \tag{2.166}$$

$$\xi = \frac{21\lambda^{5/2}}{a^{7/2}}. \tag{2.167}$$

The power coupled from one fiber to the other can be varied by changing z, the length of the coupling region over which the fields from the two fibers interact. Here a is the effective core radius in the coupling region, and Δa is the difference in effective core radii in the coupling region. If the core radius difference is zero ($\Delta a = 0$), $F = 1$, and the power coupling coefficient becomes

$$C = \sin^2 \xi z. \tag{2.168}$$

The coupling ratio can be made anywhere between 0 and 1 by varying the coupling length and by adjusting the two effective core diameters in the coupling region. Since the coupling coefficient depends on the wavelength, fused fiber couplers can also be designed to deliver all the optical power at one wavelength to one output port and all the optical power at another wavelength to the other output port. In this case we have a wavelength-division multiplexer/demultiplexer.

An optical power splitter can also be made with planar light-wave circuits. A Y-branch or multimode interference planar light-wave circuit can split the input optical power to two output waveguides. The coupling ratio of the Y-branches depends on the structure and dimension of the waveguides.

An optical coupler is reciprocal; in one direction it is a $1 \times N$ optical coupler; an optical signal with power P_{in} is split into N branches with equal power, $P_{\text{out}} = P_{\text{in}}/N$ (assume that the splitter is lossless). In the reversed direction, an optical signal with power P_i entering one of the N branches will also come out of the common port with power P_i/N.

2.5.2 Optical Filters

An optical filter is essentially a bandpass filter operating at optical frequencies which selectively passes one or more wavelengths and rejects others. Optical filters are used in WDM systems to separate optical signals with different wavelengths. Common optical filters include Fabry–Perot etalon, thin-film filters, Mach–Zehnder interferometers, fiber Bragg gratings, and arrayed waveguide gratings.

2.5.2.1 Fabry–Perot Etalon A Fabry–Perot interferometer (FP etalon) consists of a resonant cavity of length L formed by two parallel mirrors with power reflectivity R. An optical signal that enters the cavity will be reflected back and forth by two end mirrors of the FP etalon. Each time the optical signal reaches the output mirror, part of the optical signal is transmitted and the rest is reflected. The total output will be the sum of all these transmitted signals with interference among them. Adding up all the successive contributions to the output signal, the complex transfer function of the optical field is

$$H(f) = \frac{E_{\text{out}}(f)}{E_{\text{in}}(f)} = \frac{1 - R}{1 - R\ \exp(-j4\pi f \tau)} \exp(-j2\pi\tau), \tag{2.169}$$

where $E_{\text{out}}(f)$ and $E_{\text{in}}(f)$ are the input and output optical field, respectively, and τ is the one-way propagation time of the optical signal. The optical power transfer

function is given by

$$T(f) = |H(f)|^2 = \frac{(1-R)^2}{(1-R)^2 + 4R\sin^2(2\pi f\tau)}. \tag{2.170}$$

The optical power transfer function is a periodic function and its period is called a free spectral range, given by

$$FSR = \frac{1}{2\tau} = \frac{c_0}{2nL}, \tag{2.171}$$

where n is the refractive index of the cavity and c_0 is the speed of light in vacuum. The optical power transfer is maximum for optical signals with wavelength

$$\lambda = \frac{2nL}{m}. \tag{2.172}$$

In other words, constructive interference occurs when a round trip in the cavity gives a phase shift of a multiple of 2π. The half-power bandwidth (full width at half maximum) is given by

$$\text{BW} = \frac{c_0}{2nL}\frac{1-R}{\pi\sqrt{R}} = \frac{\text{FSR}}{F}, \tag{2.173}$$

where the finesse is defined as the ratio of FSR and bandwidth. To get high selectivity, it is necessary to have a high level of finesse, and this, in turn, requires a high mirror reflectivity R. It is possible to cascade several etalons, each having a modest finesse, to gain a higher overall effective finesse. The FP filter can be tuned with cavity length change through a piezoelectric effect or through the index changes of the cavity materials (such as liquid crystal). To tune the filter across one entire FSR, it is only necessary to change the mirror spacing by $\lambda/2$.

Thin-Film Filters A thin-film filter is essentially a Fabry–Perot etalon where the mirrors surrounding the cavity are realized by using multiple reflective dielectric thin-film layers (wavelength-dependent mirror). A thin-film resonant multicavity filter consists of two or more cavities separated by reflective dielectric thin-film layers. The effect of having multiple cavities on the response of the filter is that the top of the passband becomes flatter and the skirts become sharper, both very desirable filter features. The multiple dielectric slab structure is quite versatile, and a number of well-known transfer functions, such as the Butterworth and Chebyshev functions, may be synthesized using such a structure. However, the synthesis of these filters calls for a variety of dielectric materials with different refractive indices. This may be a difficult requirement to meet in practice. It turns out, nonetheless, that very useful filter transfer functions can be synthesized using just two different dielectric

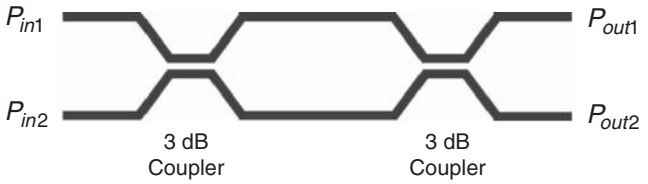


FIGURE 2.20 Structure of a Mach–Zehnder interferometer.

materials, a low- and a high-index dielectric. A particularly useful filter structure consists of a few half-wavelength cavities (a thickness of dielectric equivalent to a half wave-length, i.e., two layers of quarter-wavelength film) separated by several quarter-wavelength films (alternating layers of high- and low-index dielectrics with thickness equivalent to a quarter wavelength).

2.5.2.2 Mach–Zehnder Interferometer Figure 2.20 shows the structure of a Mach–Zehnder interferometer. When an optical signal is incident on one of the input ports of the Mach–Zehnder interferometer, the optical signal is split by a 3-dB coupler at the input. Each split light wave traverses different paths and merges in another 3-dB coupler at the output. The two paths differ by a delay of τ. The transfer function for an optical field can be found as

$$\begin{bmatrix} H_{11}(f) \\ H_{12}(f) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \exp(-j2\pi f\tau) - 1 \\ j\exp(-j2\pi f\tau) + j \end{bmatrix}, \tag{2.174}$$

and the power transfer function of a Mach–Zehnder interferometer is given by

$$\begin{bmatrix} |H_{11}(f)|^2 \\ |H_{12}(f)|^2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \sin^2(\pi f\tau) \\ \cos^2(\pi f\tau) \end{bmatrix}, \tag{2.175}$$

where H_{11} denotes the transfer function from input port 1 to output port 1, and H_{12} denotes the transfer function from input port 1 to output port 2. Both power transfer functions are raised sinusoids, 90° out of phase with respect to each other. A Mach–Zehnder interferometer can be tuned by varying the relative delay through electro-optic or thermo-optic effects. The finesse of a Mach–Zehnder interferometer is only 2. In practice, a single Mach–Zehnder interferometer cannot provide sufficient wavelength selectivity, so multiple stages of Mach–Zehnder interferometers are usually cascaded to achieve a finesse comparable to that of Fabry–Perot etalons.

2.5.2.3 Fiber Bragg Gratings Bragg gratings are periodic perturbations, usually periodic variations of the refractive index in an optical medium. If two waves with propagation constant β_1 and β_2 propagate through a Bragg grating, energy can

be coupled from one wave to the other if the Bragg phase-matching condition is satisfied,

$$|\beta_1 \pm \beta_2| = \frac{2\pi}{\Lambda}, \tag{2.176}$$

where λ is the period of the Bragg grating. For a reflective filter, optical signals can be scattered in the opposite direction at the same wavelength if

$$2\beta_0 = \frac{2\pi}{\Lambda}. \tag{2.177}$$

In other words, the wavelength of the reflective optical signal is given by

$$\lambda_0 = 2n_{\text{eff}}\Lambda, \tag{2.178}$$

where n_{eff} is the effective index of the Bragg grating.

A fiber grating can be formed in the fiber core by ultraviolet light imprinting with a two-beam interference method or phase mask method. When the germanium-doped fiber core is exposed to an intense ultraviolet pattern, structure defects are formed, thus changing the refractive index in the core. Fiber gratings are classified as short- and long-period gratings. Short-period gratings are Bragg gratings which are generally used as narrowband notch filters, where a forward-propagating mode with a Bragg wavelength is strongly coupled to a backward-propagating mode. In long-period fiber gratings, energy is coupled from the forward-propagating mode in the fiber core into other forward-propagating modes in the cladding. The phase-matching condition for long-period gratings is given as

$$\left|\beta_{\text{core}} - \beta_{\text{cladding}}\right| = \frac{2\pi}{\Lambda}. \tag{2.179}$$

In general, the difference in propagation constants between two forward-propagating modes (in core and cladding, respectively) is quite small, leading to a fairly large value of grating period. Since the cladding mode is highly lossy, long-period gratings are very efficient, and a band rejection filter can be tailored to provide almost exact equalization of an erbium gain spectrum. Hence, long-period fiber Bragg gratings are commonly used as gain equalizers in EDFAs.

2.5.2.4 Arrayed Waveguide Gratings The arrayed waveguide grating is a generalization of the Mach–Zehnder interferometer. Typically, an array waveguide grating consists of M input ports, N output ports, and two identical focusing planar star couplers connected by K uncoupled waveguides (arrayed waveguides), as shown in Figure 2.21. The length of adjacent waveguides differs by a constant value ΔL, so they form a Mach–Zehnder type of grating.

When an optical signal is incident on one of the input ports, it is split and coupled into those array waveguides. Since the waveguides have different lengths, when these

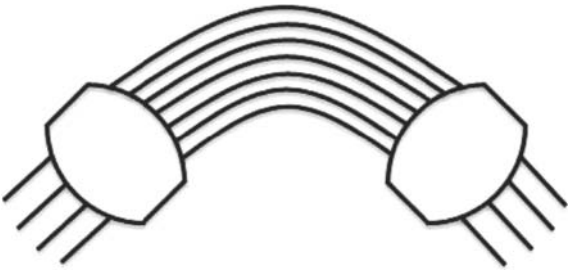


FIGURE 2.21 Structure of an arrayed waveguide grating.

signals arrive at the output port, a constructive or destructive interference pattern will be formed. Optical signals with different wavelengths will have constructive interference at different output ports. Therefore, when WDM optical signals are incident at one input port, they will separate into different outputs. Arrayed waveguide gratings are typically used as multiplexers or demultiplexers in WDM optical communication systems.

2.6 SYSTEM DESIGN AND ANALYSIS

The performance of an optical communication system depends on many factors; fiber loss, dispersion, nonlinearities, and receiver noise are the four most important limiting factors. Fiber loss attenuates the signal power; after transmission over a long length of optical fiber, the signal power could be very low. If the signal power after photodetection is lower than the noise level at the receiver, the signal cannot be recovered reliably. Fiber dispersion introduces pulse broadening and intersymbol interference and thus introduces bit errors at optical receivers. Fiber nonlinearities such as self-phase modulation, cross-phase modulation, and four-wave mixing may reduce signal power, lead to signal distortion, and introduce crosstalk, and hence degrade the overall system performance. In this section we analyze the system performance of an optical communication link and discuss the power budget and dispersion limit imposed by optical transmission channels.

2.6.1 Receiver Sensitivity

The primary factor in determining receiver sensitivity is the signal-to-noise ratio at the optical receiver. The signal power at the receiver depends on the optical signal strength and the photodetector. The major noise sources in optical receivers are thermal noise, shot noise, and ASE noise if an optical amplifier is used.

Noise Sources Thermal noise is due to the random motion of electrons in resistors or other electronic components. Electrons are always in motion because of the resistor's thermal energy. The velocity and directions of electron motion are random,

and the intensity of this motion is proportional to the temperature. The effect of this thermal motion of electrons is to produce a randomly varying voltage or current, called *thermal noise*. Thermal noise can be represented as a voltage source, v_{th}, in series with the resistor or a current source, i_{th}, in parallel with the resistor. The mean-squared values of the thermal noise are

$$\begin{aligned} \langle v_{\text{th}}^2 \rangle &= 4k_B T R_L \, \Delta f, \\ \langle i_{\text{th}}^2 \rangle &= 4k_B T \, \Delta f / R_L, \end{aligned} \tag{2.180}$$

where k_B is the Boltzmann constant, T the absolute temperature, R_L the load resistance, and Δf the effective bandwidth.

The randomness of the photogeneration process in a photodiode produces a current with average value I_{ph} (proportional to the optical power) but with fluctuations about this average value. The fluctuations can be explained on the basis of random emission and generation times and the discrete quantity of charge on the electron. The fluctuations are represented by a noise current called *shot noise*. The mean-squared value of the shot noise current is given by

$$\langle i_s^2 \rangle = 2q I_{\text{ph}} \, \Delta f, \tag{2.181}$$

where q is the elementary charge.

PIN Receivers For PIN receivers, the average photocurrent is

$$I_{\text{ph}} = R P_{\text{in}}, \tag{2.182}$$

where P_{in} is the average receiver optical power. The shot noise generated by the photodiode is

$$\langle i_s^2 \rangle = 2qIB = 2q \left(I_{\text{ph}} + I_d \right) \Delta f, \tag{2.183}$$

where I_d is the dark current (leakage current when no optical signal is incident) in the photodiode. Thermal noise due to the load resistor is

$$\langle i_{\text{th}}^2 \rangle = \frac{4k_B T \, \Delta f}{R_L}. \tag{2.184}$$

Therefore, the signal-to-noise ratio for PIN receivers is

$$\text{SNR} = \frac{(R P_{\text{in}})^2}{2q \left(R P_{\text{in}} + I_d \right) \Delta f + 4k_B T \, \Delta f / R_L}. \tag{2.185}$$

In most cases of practical applications, thermal noise dominates PIN receiver performance. Since the thermal noise current is inversely proportional to load resistor R_L,

SNR can be improved by increasing the load resistor. This is the reason that most optical receivers use a high impedance of transimpedance amplifier.

APD Receivers Optical receivers that employ an APD generally provide a higher SNR for the same incident optical power. The improvement is due to the internal gain, which increases the photocurrent by a multiplication factor M so that the photocurrent becomes

$$I_{\text{ph}} = R_{\text{APD}} P_{\text{in}} = MRP_{\text{in}}. \tag{2.186}$$

The shot noise generated by an APD is

$$\langle i_s^2 \rangle = 2qFM^2 IB = 2qFM^2 \left(I_{\text{ph}} + I_d\right) \Delta f, \tag{2.187}$$

where the excess noise factor is given by

$$F = kM + (1-k)(2 - 1/M). \tag{2.188}$$

Again, the thermal noise for the load resistor is

$$\langle i_{\text{th}}^2 \rangle = \frac{4k_B T\ \Delta f}{R_L}. \tag{2.189}$$

Therefore, the signal-to-noise ratio for APD receivers is

$$\text{SNR} = \frac{(MRP_{\text{in}})^2}{2qFM^2\,(MRP_{\text{in}} + I_d)\ \Delta f + 4k_B T\ \Delta f / R_L}. \tag{2.190}$$

For a given incident optical power P_{in}, there exist an optimum value of APD gain such that the signal-to-noise is maximized for an APD receiver. The optimum value of the multiplication factor can be determined from

$$kM_{\text{opt}}^3 + (1-k)M_{\text{opt}} = \frac{4k_B T}{qR_L\,(RP_{\text{in}} + I_d)}. \tag{2.191}$$

The optimum value M_{opt} depends on a large number of receiver parameters, such as the dark current I_d, the responsivity R, and the ionization coefficient ratio k. However, it is independent of receiver bandwidth. M_{opt} decreases with an increase in input optical power P_{in} and is very sensitive to the ionization coefficient ratio k. For $k = 0$, M_{opt} decreases inversely with P_{in} (In practice, the dark current is negligible compared with photocurrent in an APD.) By contrast, M_{opt} varies as $P_{\text{in}}^{-1/3}$ for $k = 1$.

PIN Receivers with Optical Preamplifiers When an optical preamplifier is placed in front of a PIN receiver, the optical signal power is improved but ASE noise is also added. Typically, ASE noise is greater than the thermal noise in a PIN receiver

and becomes the dominate noise source. The signal generated in the preamplified receivers is

$$I_{\text{ph}} = GRP_{\text{in}}, \tag{2.192}$$

where G is the optical amplifier gain. During photodetection, the ASE noise will beat with the optical signal and create noise in the receiver known as *signal-spontaneous beat noise*,

$$\langle i^2_{\text{sig-sp}} \rangle = 2h\nu R^2 G^2 F_n P_{in}\, \Delta f, \tag{2.193}$$

where F_n is the noise figure of the optical amplifier. In addition, the ASE noise can beat with itself, creating *spontaneous-spontaneous noise* in the receiver, given by

$$\langle i^2_{\text{sp-sp}} \rangle = (h\nu GRF_n)^2\ \Delta\nu\, \Delta f, \tag{2.194}$$

where $\Delta\nu$ is the bandwidth of the optical filter following the optical amplifier. This filter can reduce the ASE noise power. Also, there is shot noise given by

$$\langle i^2_s \rangle = 2qGRP_{\text{in}}\, \Delta f, \tag{2.195}$$

where the shot noise due to ASE noise current and dark current is neglected. The shot noise can beat with ASE noise as well, and the *spontaneous-shot beat noise* is given by

$$\langle i^2_{\text{s-sp}} \rangle = 2h\nu GRF_n\, \Delta\nu\, \Delta f. \tag{2.196}$$

The signal-spontaneous beat noise and the spontaneous-spontaneous beat noise dominate among all the noise sources. Therefore, the signal-to-noise ratio in preamplified receivers is given by

$$\text{SNR} = \frac{(GRP_{\text{in}})^2}{2h\nu R^2 G^2 F_n P_{\text{in}}\, \Delta f + (h\nu GRF_n)^2\, \Delta\nu\, \Delta f}. \tag{2.197}$$

Receiver Sensitivity In an optical receiver, the signals I received are the sum of the signal (I_1 for symbol 1 and I_0 for symbol 0) and the noise. The decision circuit compares the value sampled with a threshold value I_D. If $I > I_D$, the decision is binary 1 or binary 0 for $I < I_D$. Assume that the noise in the optical receiver is Gaussian with variance (noise power) σ_1^2 for symbol 1 received and σ_0^2 with symbol

0 received. The probability density function for the signal received is a Gaussian function which is given by

$$\begin{aligned} f_1(I) &= \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{(I-I_1)^2}{2\sigma_1^2}\right] \qquad \text{binary 1 sent,} \\ f_0(I) &= \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left[-\frac{(I-I_0)^2}{2\sigma_0^2}\right] \qquad \text{binary 0 sent.} \end{aligned} \tag{2.198}$$

The probability of error when a binary 1 is transmitted is given by

$$P(0|1) = \int_{-\infty}^{I_D} \frac{1}{\sqrt{2\pi}\,\sigma_1}\exp\left[-\frac{(I-I_1)^2}{2\sigma_1^2}\right] dI = \frac{1}{2}\text{erfc}\left(\frac{I_1 - I_D}{\sqrt{2}\,\sigma_1}\right), \tag{2.199}$$

and similarly, the probability of error when a binary 0 is sent is

$$P(1|0) = \int_{I_D}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_0}\exp\left[-\frac{(I-I_0)^2}{2\sigma_0^2}\right] dI = \frac{1}{2}\text{erfc}\left(\frac{I_D - I_0}{\sqrt{2}\,\sigma_0}\right), \tag{2.200}$$

where erfc$(\cdot)$ is the complementary error function, defined as

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} xe^{-u^2}\,du. \tag{2.201}$$

Assume that the probabilities of sending binary 1 and binary 0 are equal; that is, $P(0) = P(1) = 1/2$. Then the average probability of error (bit error rate) is

$$\text{BER} = P(1)P(0|1) + P(1|0) = \frac{1}{4}\left[\text{erfc}\left(\frac{I_D - I_0}{\sqrt{2}\,\sigma_0}\right) + \text{erfc}\left(\frac{I_1 - I_D}{\sqrt{2}\,\sigma_1}\right)\right]. \tag{2.202}$$

The optimum threshold is the one that minimizes the probability of error. The optimum threshold is determined by the equation

$$\frac{(I_1 - I_D)^2}{2\sigma_1^2} - \frac{(I_D - I_0)^2}{2\sigma_0^2} = \ln\frac{P(1)}{P(0)} - \ln\frac{\sigma_1}{\sigma_0}. \tag{2.203}$$

In practice, the threshold is chosen to make the probability of error $P(0|1)$ and $P(1|0)$ equal; that is,

$$\frac{I_1 - I_D}{\sigma_1} = \frac{I_D - I_1}{\sigma_0}. \tag{2.204}$$

In this case, the threshold is given by

$$I_D = \frac{\sigma_0 I_1 - \sigma_1 I_0}{\sigma_0 + \sigma_1}, \tag{2.205}$$

and the probability of error is

$$\text{BER} = \frac{1}{2}\text{erfc}\left(\frac{Q}{\sqrt{2}}\right), \tag{2.206}$$

where

$$Q = \frac{I_1 - I_0}{\sigma_1 - \sigma_0}. \tag{2.207}$$

The bit error rate decreases with increasing Q value. For $Q = 6$, the bit error rate is about 1.0×10^{-9}, and for $Q = 7$, the bit error rate is approximately 1.0×10^{-12}. To achieve a certain bit error rate (e.g., BER $= 1.0 \times 10^{-9}$), the optical P_r power received must be larger than a certain value P_s. This minimum received power P_s to reach a certain bit error rate is called *receiver sensitivity*.

2.6.2 Power Budget

In optical communication system design, a certain power budget is required to ensure that enough power will reach the receiver to maintain reliable performance during the entire system lifetime. The minimum average power required by the optical receiver is specified by the receiver sensitivity P_s, and the average launch power of a transmitter is generally specified by P_t. Then the power budget is expressed in the equation

$$P_t - P_s > C_L + M_s, \tag{2.208}$$

where C_L is the total channel loss, M_s is the system margin, and the optical power is expressed in units of dBm. The purpose of system margin M_s is to allocate a certain amount of power to additional sources of power penalty that may develop during a system lifetime because of component degradation or other unforeseen events. A system margin of a few (3 to 10) dB is generally allocated during the design process. The channel loss C_L includes fiber loss and all the other losses (such as connector loss, splice loss, and other passive component loss). In PONs, the channel loss consists of fiber and splitter losses (neglecting connector and splice losses), given by

$$C_L = \alpha_f L + C_{\text{splitter}}, \tag{2.209}$$

where α_f is the fiber loss in dB/km, L the fiber length, and C_{splitter} the splitter loss. Given a specific launch power of the transmitter and the receiver sensitivity, the

maximum transmission distance in PONs is given by

$$L \leq \frac{P_t - P_s - C_{\text{splitter}}}{\alpha_f}. \tag{2.210}$$

Or in the case of a 20-km transmission distance (as specified by GPON/EPON standards), the maximum splitter loss is

$$C_{\text{splitter}} \leq P_t - P_s - \alpha_f L. \tag{2.211}$$

Assume that the loss of a $1 \times N$ splitter is $10 \log N$; then the maximum number of users that can be supported by the PON system is given by

$$N \leq 10^{(P_t - P_s - 20\alpha_f)/10}. \tag{2.212}$$

Since PONs are mostly power-limited systems, the power budget is a very important design parameter in PONs.

2.6.3 Dispersion Limit

Fiber dispersion can lead to pulse distortion and intersymbol interference. In addition to fiber loss, fiber dispersion imposes another limit on the bit rate–distance product of optical communication systems. A common design criterion for the fiber dispersion limit is given by

$$B < \frac{1}{4\sigma_D}, \tag{2.213}$$

where B is the bit rate and σ is the rms (root-mean-square) dispersion induced by optical fibers. In PON systems, the upstream transmission uses a 1.3-μm window, where the fiber dispersion is minimum (for standard single-mode fiber). Hence, if narrow-linewidth DFB lasers are used at the transmitters, fiber dispersion is not a significant limiting factor for upstream transmission. However, if a Fabry–Perot laser is used, the rms fiber dispersion is given by

$$\sigma_D = \sigma_\lambda |D| L, \tag{2.214}$$

where σ_λ is the rms linewidth of the Fabry–Perot laser (usually, 2 nm), D is the fiber dispersion coefficient, and L is the fiber length. The dispersion parameter D depends on how close the operating wavelength is to the zero-dispersion wavelength of the standard single-mode fiber. The bit rate–distance (BL) product of upstream transmission is thus given by

$$BL < \frac{1}{4\sigma_\lambda |D|}. \tag{2.215}$$

At the 1.3-μm window, $|D|$ is less than 3 ps/nm·km. Therefore, the BL product is limited to about 50 Gb/s·km. As the maximum transmission distance in GPON/EPON is only 20 km, fiber dispersion will not cause significant system performance degradation at 2.5 Gb/s even if a Fabry–Perot laser is used.

For downstream transmission at the 1.49- and 1.55-μm windows, fiber dispersion could be a problem since the fiber dispersion coefficient is about 15 to 17 ps/nm·km. If a Fabry–Perot laser is used at the transmitter, the BL product will be limited to 8 Gb/s·km. Therefore, fiber dispersion will limit the transmission distance to 8 km for a bit rate of 1.0 Gb/s, less than the transmission distance required by the GPON/EPON standard. Therefore, a DFB laser operating in a single longitudinal mode is required for downstream transmission. For a narrow-source spectral width, the rms fiber dispersion is given by

$$\sigma_D = |\beta_2|\, L, \tag{2.216}$$

where the group velocity parameter β_2 is related to the dispersion coefficient by

$$\beta_2 = -\frac{\lambda^2 D}{2\pi c_0}. \tag{2.217}$$

Therefore, for transmitters with DFB lasers, the BL product is limited by

$$B^2 L < \frac{1}{16\beta_2}. \tag{2.218}$$

At the 1.55-μm window, $\beta_2 \approx 2$ s^2/km. Thus, the bit rate–distance product is limited to about 3000 (Gb/s)^2·km. However, eq. (2.218) works only for transmitters with external modulators. For direct modulation using DFB lasers, the frequeny chirp leads to a broadening of the signal spectrum and imposes a much more severe limitation on the BL product. If we assume an rms spectral width $\sigma_\lambda = 0.1$ nm for an optical signal from a direct modulated DFB laser, the dispersion limited is given by

$$BL < \frac{1}{4\sigma_\lambda\, |D|}. \tag{2.219}$$

Therefore, the BL product is limited to about 150 Gb/s·km. For a 20-km reach in PONs, direct modulation of DFB lasers works for the bit rate of current GPON/EPON standards (maximum 2.5 Gb/s). However, for next-generation PONs with a 10-Gb/s data rate, direct modulation could lead to a large power penalty in downstream transmission. In such cases, external modulation is required.

2.7 OPTICAL TRANSCEIVER DESIGN FOR TDM PONS

The optical transceiver (transmitter and receiver) used for optical-to-electrical conversion is a key component in optical communication systems. In a PON system, an optical transmitter and receiver at the optical line terminal (OLT) or optical network units (ONUs) are typically packaged together to form a bidirectional optical subassembly (BOSA). Figure 2.22 shows the architecture of optical transceivers for an OLT and ONUs. The transmitter optical subassembly (TOSA) consists of a semiconductor laser (Fabry–Perot laser or DFB laser) and a laser driver. The receiver optical subassembly (ROSA) includes a photodiode (PIN or APD), a transimpedance amplifier, a limiting amplifier, and a clock and data recovery circuit. In addition to TOSA and ROSA, a duplex or triplex (WDM filter) is used to separate upstream and downstream wavelengths. The duplex or triplex is usually a thin-film filter, but optical filters (e.g., Bragg grating or a Mach–Zehnder interferometer) based on planar light-wave circuits are becoming a preferred option, as planar light-wave circuits are more compact, more reliable, and easier to assemble with TOSA and ROSA. As TDM PONs are being deployed on a large scale, significant effort has been put into designing an optical transceiver with improved performance, lower cost, and better reliability. The key challenges in developing optical transceivers for FTTx applications are a higher level of integration, cost-effective packaging, and burst-mode optical transmission

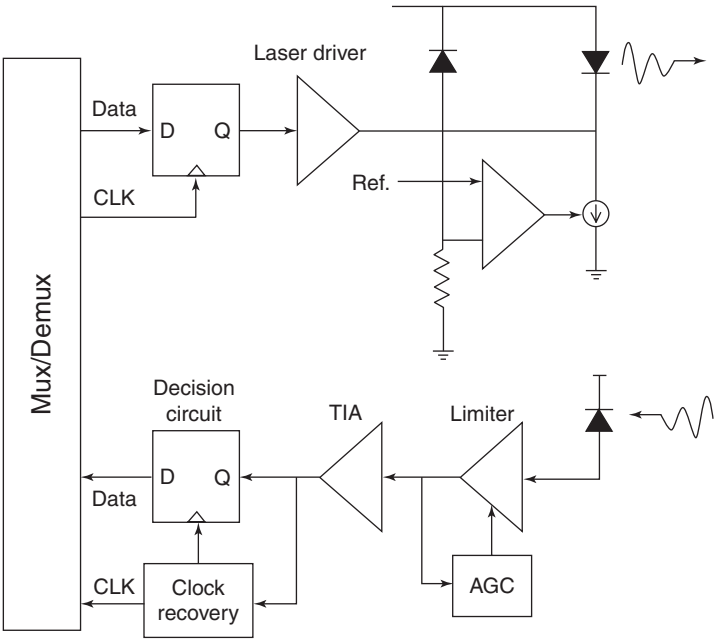


FIGURE 2.22 Block diagram of an optical transceiver.

technologies in the upstream link. For integration and packaging, optical transceivers are evolving toward planar light-wave circuits and monolithic photonic integrated circuits.

2.7.1 Burst-Mode Optical Transmission

In a TDM PON system, all the users share the same fiber infrastructure from an OLT to the distribution node. In the downstream direction, data packets are broadcast to all ONUs. OLT transmitters and ONU receivers operate in a continuous mode, where synchronization is maintained at all times. Even if there are no data to send to ONUs, the OLT transmitter has to transmit idle bit patterns so that the ONUs' receiver can retrieve the clock continuously from the downstream signal. However, in the upstream direction, all the users have to share the channel with the TDMA scheme. Packet collisions have to be avoided in the upstream transmission, so at any given time, only one packet (from one ONU) is allowed to reach the central office. The OLT coordinates the upstream transmission and schedules the transmission time for each ONU. When an ONU wants to send data to the OLT, it transmits a burst of data in the time assigned by the OLT and then turns off its transmitter completely to avoid interfering with other ONUs' upstream transmission. Bursts of data from different ONUs following each other to the receiver at the central office are separated by a certain guard time. This type of transmission is called *burst-mode transmission*. Figure 2.23 compares the data formats of continuous- and burst-mode transmission. Burst-mode optical transmission is a key technology for TDM PONs, in which a burst-mode transmitter at each ONU and a burst-mode receiver at the central office

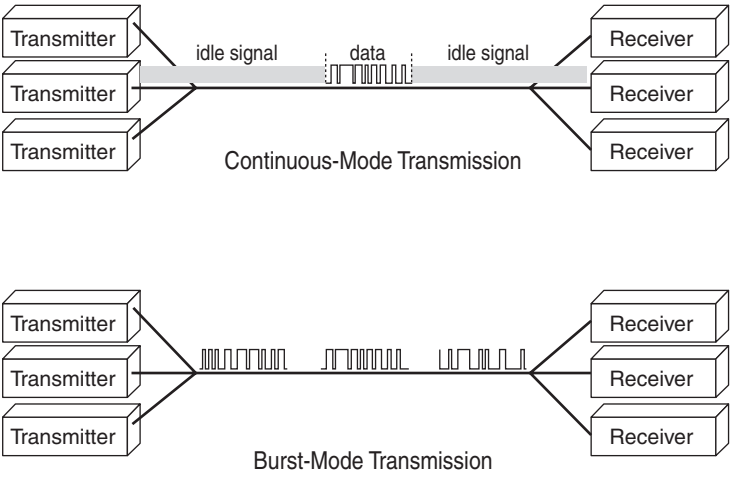


FIGURE 2.23 Comparison of continuous-mode transmission and burst-mode transmission. In continuous-mode transmission, continuous bit streams are sent from sources to destinations; if no data need to be sent, idle signals are sent. In burst-mode transmission, data are sent in bursts from sources to destinations. No idle signals are sent between bursts.

are indispensable. A burst-mode receiver at the central office is required to have various affordable input ranges and fast clock locking times. On the other hand, a burst-mode transmitter, located on the user side, has to show fast turn-on time as well as good power suppression during an idle state. Designing high-speed burst-mode optical transceivers is necessary and challenging when deploying passive optical networks.

Burst-Mode Laser Drivers A burst-mode transmitter has to show fast turn-on time as well as good power suppression. The design challenges for burst-mode laser/modulator drivers are rise and fall times and automatic power control. Conventional driver circuits are designed to maintain a constant bias current and voltage. However, good optical power suppression in the idle state requires that the bias be turned off quickly. Driver circuits have to be designed to have short turn-on/off performance. For automatic power control, conventional circuits often use slow monitor photodiode and/or analog filters to average the signal and an analog control loop to maintain the desired operating point. Burst-mode operation prevents the use of simple analog control loops. It is necessary to monitor optical output sampled at appropriate points in a burst waveform.

Burst-Mode Receivers Conventional optical receivers cannot be used for burst-mode detection because they are not able instantaneously to handle various packets arriving with large differences in optical power and phase alignment. It is therefore necessary to design receivers that can adapt to the variation in optical power and phase alignment on a packet-by-packet basis. The design challenges for burst-mode receivers are (1) dynamic sensitivity recovery, (2) level recovery, and (3) fast clock recovery.

When a weak burst follows a strong burst, it is difficult to detect the weak signal. Dynamic sensitivity recovery is necessary for detection of the weak signal. The recovery of the weak burst is limited by photodiode carrier transport effects, amplifier slew and charging rates, and unintentional automatic gain control effects.

For level recovery, a burst-mode receiver can be designed with a feedback or feedforward structure. For the feedback design, a differential input/output transimpedance amplifier and peak detection circuit form a feedback loop, whereas in the feedforward design, the signal from the preamplifier is feedforward into a peak detection circuitry. Both designs have been implemented in practice. A feedback structure enables the receiver to work more reliably, but a different input/output preamplifier is needed. A feedforward receiver has a faster response time and a conventional dc-coupled preamplifier can be used, but, the circuitry needs to be designed carefully to prevent oscillation in the receiver. Level recovery circuits with simple and robust design and good performance are still an open issue that requires further investigation.

For burst-mode clock recovery, the classical approach—phase-locked loops—can only recover the clock and phase after thousands of bits and hence are not suitable for burst-mode transmission. Novel clock recovery architectures such as gated VCO and oversampling have been proposed to fulfill the requirement of burst-mode clock

recovery. However, the jitter performance of these novel burst-mode recovery schemes is relatively poorer than that of the conventional phase-locked loop.

2.7.2 Colorless ONUs

Reliable and low-cost WDM light sources are indispensable for the successful deployment of future WDM PONs. In standard TDM PONs, conventional laser diodes, especially inexpensive FP lasers, are used routinely in both OLTs and ONUs. However, for WDM PONs, multiwavelength laser diodes can be very expensive, especially if wavelength locking is required in DWDM networks. Cost-effective implementation of WDM light sources at the ONU side is necessary for WDM PONs. Several approaches have been demonstrated for implementation of colorless ONUs in WDM PONs, such as spectral sliced broadband light sources, injection-locked FP lasers, and reflective SOAs. The wavelengths of these light sources are not determined by the gain media themselves but by external factors such as optical filters or injected signals, so the wavelengths are easier to manage without worrying about temperature or aging effects in ONUs. Table 2.1 compares the advantages and disadvantages of various colorless ONUs.

Spectral Sliced Broadband Light Sources A comb of optical signals, each with a unique wavelength, can be achieved by slicing the spectrum of a broadband light source. The broadband light source can be superluminescent light-emitting diodes or EDFA or FP lasers. For wavelength selection, an AWG, usually deployed at the remote node, will slice a narrow spectrum of the broadband optical signals, and different wavelengths will be selected for different ONUs. If a tunable filter is used, the wavelengths of the optical comb can be tuned. The advantage of using spectral slicing is simple implementation and low cost. The disadvantages are limited modulation speed, low power, and incoherent output, which will limit the transmission span.

Injection-Locked FP Lasers When an external narrowband optical signal is injected into a multiple longitudinal mode laser such as an FP laser, the lasing mode will be locked to a single mode. The external optical signal acts as a seed for laser oscillation in the FP cavity, and the mode that is nearest the peak wavelength of the injected optical signal will be locked to the injected light, and other modes will be suppressed. The injected optical signal can be a spectral sliced broadband light source (e.g., ASE) or a modulated downstream signal. The advantages of using injection-locked FP lasers are simplicity and low cost, and the modulation speed can be relatively high. The disadvantages are the limited locking range for wavelengths and power. Modulation index, laser bias current, and the power of external optical excitation must be chosen carefully to maximize the efficiency of injection-locked FP lasers. Since the optical signal from an ONU has the same wavelength as that of the injected optical signal, the backscattering of the injected signal could degrade the performance of upstream transmission.

TABLE 2.1 WDM-PON Approach Comparison Summary

Approach	Commmunication	Key Components	Advantages	Disadvantages
Tunable lasers	• Full-duplex • Gbit/s	ONU: tunable laser OLT: WDM demux	• Dynamic wavelength management • Can share fiber from ONU to RN if using tunable receivers for downstream	• Very high cost for access networks
BLS w/spectral slicing	• Full-duplex • Few Mbit/s	ONU: LED BLS RN: AWG OLT: WDM demux	• Inexpensive BLS	• High slicing power loss limits network reach • Incoherent output • BLS spectrum width limits number of users per RN • High crosstalk limitations • No fiber sharing from ONU to RN
Injection-locked FPLD	• Full-duplex • Few Gbit/s	ONU: FPLD + circulator RN: AWG OLT: ASE source	• Inexpensive FPLD	• Requires more than one fiber to the ONU due to backscattering • Limited locking range • No fiber sharing from ONU to RN
Centralized light sources	• Half-duplex or Full-duplex • 1-2 Gibt/s	ONU: SOA + circulator or RSOA RN: AWG OLT: WDM lasers, WDM demux	• No light source at ONU • SOA can simultaneously act as a detector	• ASE noise • Rayleigh backscattering and reflections need to be minimized. • No fiber sharing from ONU to RN
Shared resources (HPON, DWA)	• Half-duplex or Full-duplex	ONU: Fixed Lasers or RSOA OLT: Tunable Lasers OLT or RN: AWG	• Allow smooth transition from TDM to Hybrid TDM/WDM to full WDM	• Resource sharing require scheduling algorithms

Centralized Light Sources Another alternative for colorless ONUs is use of a centralized source at the OLT. In this approach, no light source is deployed at ONUs; the OLT provides optical signals to ONUs, where the optical signals are modulated by upstream data and sent back to the OLT. Since all the light sources are located at the OLT, the requirements for wavelength provisioning and management are relaxed in a controlled environment. Two types of modulators, external modulators and semiconductor optical amplifiers, have been used for upstream modulation. Compared to external modulators, semiconductor optical amplifiers can provide signal gain for the upstream transmission and thus achieve better performance. However, semiconductor optical amplifiers have limited modulation speed. To transmit efficiently both upstream and downstream in a full-duplex manner at the same wavelength, various data modulation schemes have been implemented: FSK-ASK, PSK-ASK, and ASK-SCM. The disadvantage of using a centralized light source is the backscattering due to the Rayleigh or Brillouin effect and the polarization sensitivity of the RSOAs or optical modulators. Also, the cost of RSOAs or external modulators can be an obstacle for practical deployment.

Self-Seeding RSOAs To eliminate the centralized source and reduce the associated cost, self-seeded reflective SOA has been demonstrated for colorless ONUs. In this scheme, the ASE noise emitted from RSOA is spectrally sliced by the AWG at the remote node, and a broadband filter reflectes the ASE noise back to RSOA. Thus, the RSOA is seeded by the spectrally sliced ASE noise, and no other light source is needed. The output from the self-seeding RSOA is incoherent and low RIN light, which can be directly modulated by upstream data. The self-seeding RSOAs have the advantages of simplicity and low cost. The cost of RSOA is relatively higher than that of FP lasers, but the scheme eliminates any other light sources and thus reduces the overall system cost. The disadvantage of self-seeding RSOA is that the incoherent output could limit the transmission distance.

2.8 SUMMARY

In this chapter we present the fundamentals of optical communication systems and discuss the key components and enabling technologies for optical transmission in passive optical networks.

Optical fiber is the transmission channel in optical communication systems. In the first section, characteristics of optical fibers, such as fiber mode, loss, dispersion and nonlinearities are examined in detail, and optical signal propagation in fibers is discussed. In the second section we discuss the operation and design of optical transmitters. The operating principles of semiconductor lasers and its major characteristics are presented, and two types of optical modulators, lithium niobate modulator and electroabsorption modulators, are discussed. The design of laser driver circuits is also presented, with a focus on laser drivers and power control circuits.

Next, we discuss two types of photodetectors commonly used in optical communications: pin and avalanche photodiodes. Optical receiver circuits are also presented.

Optical amplifiers are presented in the fourth section. Operating principles, the gain and noise figures of erbium-doped fiber amplifiers, semiconductor optical amplifiers, and Raman amplifiers are discussed in detail. Next we focus on passive optical components commonly used in passive optical networks. Operating principles of optical couplers and filters are also presented.

System performance of optical communication systems is analyzed in the sixth section. Various noise sources in optical receivers are presented and a system's bit error rate is analyzed. The power budget and dispersion limit of optical communication links are also discussed in this section.

Finally, optical transceiver design for FTTx applications is presented, burst-mode optical transmission in the upstream link of TDM PON systems is discussed, and colorless ONUs for future WDM PONs are described.

REFERENCES

1. G. P. Agrawal, *Lightwave Technology*, Vol. 1, *Components and Devices*, and Vol. 2, *Telecommunication Systems*, Wiley-Interscience, Hoboken, NJ, 2004.
2. I. P. Kaminow, T. Li, and A. E. Willner, Eds., *Optical Fiber Telecommunications*, 5th ed., Academic Press, San Diego, CA, 2008.
3. R. Ramaswami and K. Sivarajan, *Optical Networks: A Practical Perspective*, 2nd ed., Morgan Kaufmann, San Diego, CA, 2001.
4. J. A. Buck, *Fundamentals of Optical Fibers*, 2nd ed., Wiley-Interscience, Hoboken, NJ, 2004.
5. T. Suhara, *Semiconductor Laser Fundamentals*, CRC Press, Boca Raton, FL, 2004.
6. E. Kapon, *Semiconductor Laser*, Vol. I, *Fundamentals*, and Vol. II, *Materials and Structures*, Academic Press, San Diego, CA, 1998.
7. H. S. Nalwa, *Photodetectors and Fiber Optics*, Academic Press, San Diego, CA, 2001.
8. S. B. Alexander, *Optical Communication Receiver Design*, IEE, London, 1997.
9. E. Desurvire, *Erbium-Doped Fiber Amplifiers: Principles and Applications*, Wiley-Interscience, Hoboken, NJ, 2002.
10. N. K. Dutta and Q. Wang, *Semiconductor Optical Amplifiers*, World Scientific Publishing, 2006.
11. M. N. Islam, *Raman Amplifiers for Telecommunications*, Vol. 1, *Physical Principles*, and Vol. 2, *Sub-systems and Systems*, Springer-Verlag, New York, 2003.

CHAPTER 3

PASSIVE OPTICAL NETWORKS: ARCHITECTURES AND PROTOCOLS

In response to the steadily increasing demand for bandwidth and advanced networking services from residential as well as enterprise customers, passive optical networks (PONs) over the last decade have emerged as a matured access technology that offers flexibility, broad area coverage, and cost-effective sharing of expensive fiber links and networking equipment. Both ITU and IEEE have standardized solutions for PONs operating at gigabit per second line rates [1]. ITU-T extends the early work of the full service access network (FSAN) working group, which brings efficient fiber access to homes by creating ATM-based PON (APON). APON, the first standardized PON solution, supports the legacy of ATM protocols at 622 Mb/s (OC-12) of downstream bandwidth and 155 Mb/s (OC-3) of upstream bandwidth. Since the initialization of APON and an improved solution for ITU-T G.983.1 broadband PON (BPON), newer standards have been investigated to support gigabit rates and packet-based Internet applications. Gigabit PON (GPON) was rectified by ITU-T in the G.984.x Recommendations to support gigabit rates and a mix of TDM, ATM, and Ethernet services, and to enhance security. At approximately the same time, the Ethernet in the first-mile (EFM) 802.3ah study group has standardized Ethernet PON (EPON) to leverage the commercial success of Ethernet as a local area network (LAN) technology. The EFM effort creates a minimal set of extensions to the IEEE 802.3 medium access control (MAC) and control sublayers.

Despite their similarities in common transmission wavelength and optical transceiver budget, the GPON and EPON standards differ substantially in the

Broadband Optical Access Networks, First Edition. Leonid G. Kazovsky, Ning Cheng, Wei-Tao Shaw, David Gutierrez, and Shing-Wa Wong.

properties of their physical medium-dependent (PMD) layer, transmission convergence (TC) layer, OAM capabilities, and MAC layer. We will not discuss each PON standard in full detail. Instead, topics are presented to serve as introductory materials to gain insight into the various PON architectures and protocols. In the first part of the chapter we present the generic PON architecture and terminology, followed by specific packet encapsulation methods employed by IEEE and ITU standards. Then additional details are presented on user data and control planes. These details include bandwidth assignment, autodiscovery and ranging, security, and path protection. Finally, we discuss a number of new topics, such as 10-Gb line rate, WDM, reach extension, and energy efficiency, that have emerged in recent IEEE and ITU standard discussions. We include essential information on these recently proposed topics to bridge the discussion to Chapter 4, where we deal with next-generation broadband optical access networks.

3.1 PON ARCHITECTURES

PON employs a physical point-to-multipoint (P2MP) architecture. Figure 3.1 shows the architecture of a generic standard commercial TDM-PON. In this architecture, an optical line terminal (OLT) is connected and shared by multiple optical network units (ONUs) via a passive optical splitter. Downstream traffic is sent from OLTs using a 1490-nm wavelength, and upstream traffic from ONUs is carried on a 1310-nm wavelength. In some PONs, analog video signal is still carried on the 1550-nm wavelength, although most analog video signal today has been replaced by digital video signal.

The connection between an OLT and an ONU is called the optical distribution network (ODN). At the user side, the ONU offers one or more ports through its

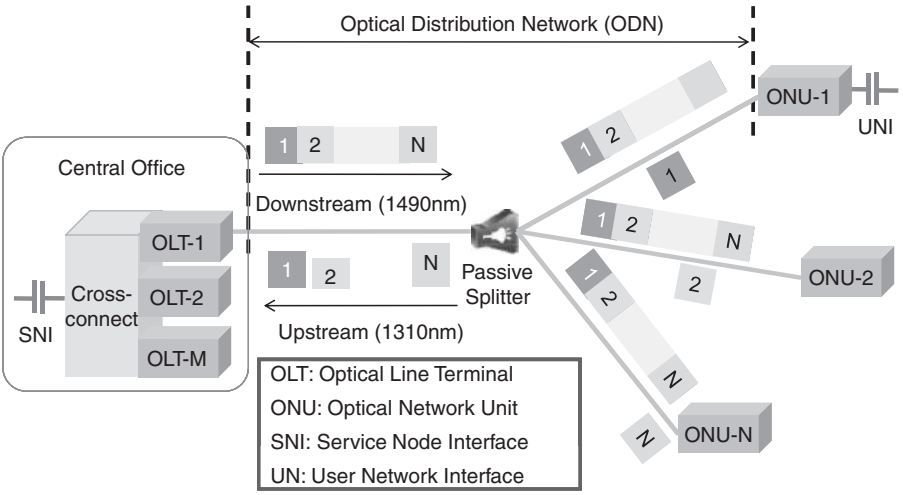


FIGURE 3.1 Standard TDM-PON architecture.

user network interface (UNI). Common UNI support, for example, includes voice connections such as T1/E1 for enterprise users and plain old telephone service (POTS) for residential users, and data connections such as 10/100 BASE-T Ethernet. Inside the central office, each OLT serves the PON ODN, and multiple OLTs are interconnected with a backbone switch or cross-connected to the backbone network. In a carrier environment, OLTs are usually constructed as line cards that are inserted into a chassis. The chassis then hosts the backbone switch or cross-connect to interconnect the OLTs through a high-speed back plane via a service network interface (SNI). The most common SNI used today are the gigabit Ethernet interface and the SONET interface.

In the ODN, signals are sent from and to different ONUs with a unique ONU identification in the frame header. These signals are encoded and multiplexed in different formats and schemes depending on the particular PON standard. Moreover, given the P2MP architecture of TDM-PON, the downstream link is a one-to-many broadcast connection and upstream link is a many-to-one connection. To avoid collisions by frames that arrive via an OLT from different ONUs, a multipoint control protocol (MPCP) allows only one ONU to transmit at any given time.

3.1.1 Network Dimensioning and Bandwidth

Limiting Factors to PON Dimensioning PON systems are limited in range and bandwidth. Current PONs typically support up to 32 to 64 splits and up to 20 km. The dimension of a PON system (i.e., the split-and-reach combination) is restricted in part by the power budget. In realistic deployment, however, this dimension is further restricted by the fiber dispersion and dynamic range of the upstream receiver. Subsequent subsections would describe in more detail the impact of the power budget, fiber dispersion, and dynamic range on PON dimensioning.

PON Bandwidths In terms of bandwidth, either EPON or GPON can support Gb/s transmission rates but differ in actual transmission rates. Although EPON supports a baud transmission line rate of 1.25 Gbaud/s, it actually supports a 1.0-Gb/s transmission rate (both upstream and downstream). This is because an EPON employs 8B/10B line codes to ensure that there are sufficient bit transitions to maintain DC balance and therefore making receiver design much easier. BPON and GPON, on the other hand, employ non-return-to-zero (NRZ) code to achieve 100% bandwidth efficiency, but at the cost of more expensive transceivers. As a result, GPON is able to support a 2.488-Gb/s (OC-48) rate downstream and a 1.244-Gb/s (OC-24) rate upstream.

3.1.2 Power Budget

As described in Section 2.6.2, the limiting split ratio N and fiber reach L(km) can be computed given the power budget $P_{\text{budget}} = P_t - P_s$(dB), fiber loss coefficient α_f(dB/km), and system margin M_s(dB). The power budget is determined by the type of transceiver employed by the systems since it is calculated by the difference

between transmitter launch power P_t and receiver sensitivity P_s. For the reasons given below, both EPON and GPON support very similar dimensioning since they rely on virtually the same underlying physical system.

Both EPON and GPON standards support similar transceivers in class A ($P_{\text{budget}} =$ 20 dB) and class B ($P_{\text{budget}} = 25$ dB) optical transceivers. GPON further supports class C (30 dB) optical transceivers. Typical class A transceivers could mean the use of a DFB laser/APD receiver at the OLT and PIN laser/FP receiver at the ONU. Class C transceivers would utilize a DFB laser/APD receiver at both the OLT and the ONU. In practice, both EPON and GPON have converged to adopt to the use of class B+ (28 dB)-type transceivers to cost-effectively enable sufficient power budget. It should be noted that most commercial DFB/APD transceivers can support more than a 28-dB budget. However, an additional 3 to 4 dB of margin is built in to account further for laser power fluctuation due to outdoor temperature variation and transceiver aging. Temperature variation is an important factor in PON because ONU lasers are being deployed without support from expensive temperature controllers.

EPON and GPON standards are also adopted to the same wavelength plan (i.e., 1490 nm downstream and 1310 nm upstream), for reasons described in Section 2.1.3. Specifically, the downstream transmission window is between 1480 and 1500 nm, and the upstream transmission window is between 1260 and 1360 nm. In addition, ITU G.983.3 specifies an enhancement band in the 1550 and 1560 nm window to reserve for analog video (ITU J.185/186) or additional digital service overlays. Among these wavelength windows, fiber losses in the upstream 1310-nm window is the limiting factor since signal experiences the greatest attenuation in the 1310-nm window.

To summarize, the following equation should be used when determining the maximum split and fiber reach of a PON system:

$$P_{\text{budget}}(\text{dB}) > \alpha_{1310\,\text{nm}} L + 10\ \log_{10}(N) + M_s(\text{dB}). \tag{3.1}$$

The typical fiber loss coefficient at 1310 nm is expected to be 0.38 dB/km, and thus a class B+ budget should be able to support a combination of 20 km and 64 users with 2.33 dB of system margin. The margin reserves a portion of budget to offset uncertainties, such as various points of connector losses, and transceiver aging. Reserving sufficient system margin is an important consideration for the operators because it could reduce the potential service outage time and thus save maintenance costs. In reality, most deployment guidelines are conservative and assume the fiber loss coefficient to be above 0.5 dB/km. A 1-dB connector loss is considered further for the worst-case scenario. As a result, a class B+ power budget would support only 12 km and 32 users with less than 3 dB of system margin.

Chromatic Dispersion As described in Section 2.6.3, fiber dispersion is one of the limiting factors for current PON systems. This is evident by eq. (2.219), where it is concluded that the bandwidth–physical reach product. The limit is 1000 Gb/s · km for a DM-DFB at only about 8 Gb/s · km for 1310 nm FP laser. In fact, the BPON and GPON standards further specify an aggressive 60-km-reach limitation based on the use of low-quality Fabry–Perot lasers with a large linewidth near 1 nm.

However, as pointed out in the discussion for power budget, the 60-km limit is overly optimistic without reserving system margin and considering worst-case fiber propagation effects. The situation, however, is expected to change when future better-quality photonic components are expected to support future PON systems. In these cases, the *BL*-related issues would be even more relevant as future PON systems are expected to drive for a combination solution of higher bandwidth and longer reach.

3.1.3 Burst-Mode Operation

In Section 2.7.1 we describe the burst-mode transmission observed in the upstream direction of PON traffic due to many ONUs sharing and taking turns to transmit over the same fiber link. Due to upstream burst-mode transmission, PONs need to account additionally for ONU laser ON and OFF times, OLT gain control adjustment, and clock recovery time. Figure 3.2 illustrates the effect of these parameters on the efficiency of the upstream transmission. The guard time includes the time for the previous ONU to turn the laser off and the new ONU to turn the laser on, a small tolerance for timing drift between successive transmitting ONUs, plus the time for the OLT to adjust the AGC gain level and recover the clock. In GPON, the "preamble" is used to allow both level recovery (AGC) and clock recovery. Delimiter (GPON) or code-group alignment (EPON) are used to find the start of the data.

In EPON, the standard body chose a relaxed set of parameters (i.e., laser ON time = laser OFF time = 512 ns and gain adjustment or recovery time <= 400 ns). GPON, on the other hand, allocates only 32 bits (25.7 ns) for laser ON and OFF and a very stringent 44 bits (35.4 ns) for gain control adjustment and clock recovery. As a trade-off, a fast AGC circuit would not be able to support as wide a dynamic range as would a slower AGC circuit. Thus, GPONs require further ONUs to suppress output power to reduce the dynamic range seen by OLTs.

To achieve fast gain adjustment, the dynamic range of the input signals must be limited. As a result, the differential distance between the nearest and farthest ONU must be limited to ensure that the AGC circuit can function within the specified gain adjustment time.

Overall, these parameters are selected to meet different design goals. IEEE EPON typically takes the most aggressive cost-cutting steps, whereas ITU-T BPON and GPON would select more aggressive specifications to maximize performance efficiency (i.e., 100% bandwidth utilization, shorter burst-mode overhead, packet assembly and fragmentation, and a higher power budget). Although the verdicts on

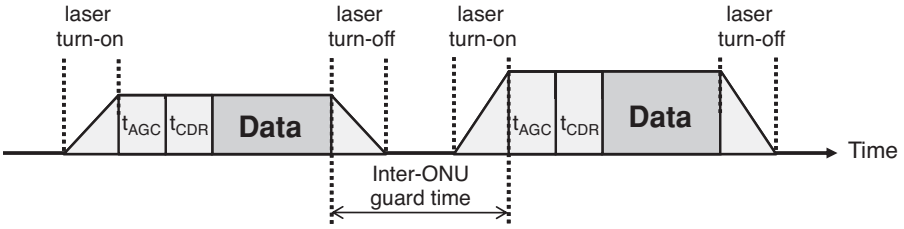


FIGURE 3.2 Timing diagram for upstream transmission.

the commercial success of EPON and GPON have yet to be determined, what these design choices do create are leaving some doors open for additional research and development. In particular, researchers and engineers have worked extensively on improving the remaining bandwidth efficiency of EPON, whereas GPON's stringent requirement on AGC and CDR time has inspired many innovative solutions to improve the performance and yield of the circuits.

3.1.4 PON Packet Format and Encapsulation

The main difference between EPONs and GPONs is in the type of packet or frame transmission they support. Today, the majority of Internet traffic is Ethernet, and so is the majority of the PON traffic. Ethernet PON is a direct extension of the IEEE 802.3 Ethernet LAN standard to fiber optic access networks. Figure 3.3 illustrates the MPCP physical data unit (PDU) frame format used by EPON. The MPCP-PDU frame is based on the IEEE 802.3z gigabit Ethernet (GbE) frame format and effectively hides the EPON header within the GbE preamble section. EPON also supports point-to-point emulation (P2PE) to make EPON links look virtually like a point-to-point system. EPON MPCP and P2PE extensions of Ethernet MAC are defined in clauses 64 and 65 of IEEE 802.3. Details of the MPCP and EPON control methods are described in Section 3.4.2.

There is a common misunderstanding that GPON cannot support Ethernet traffic efficiently. That is not true. What is true is that GPON inherits part of ATM networking and format from its predecessor APON and BPON. However, GPON further adopts the GPON encapsulation method (GEM), which uses a generic framing procedure (GFP) to support any packet type in addition to packet fragmentation and packet reassembly. From this perspective, GPON is more flexible than EPON because it can support a mix of ATM, Ethernet, and TDM service. Moreover, the ability to support

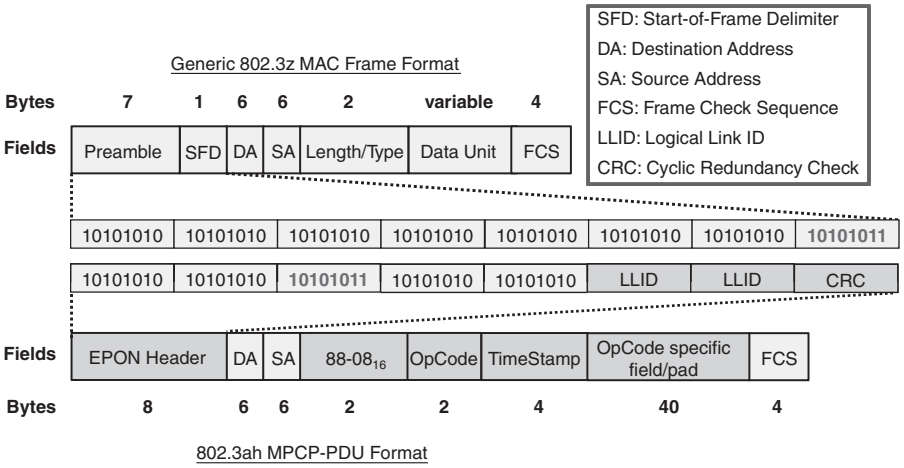


FIGURE 3.3 802.3z GbE frame format and 802.3ah MPCP–PDU frame format.

packet fragmentation and reassembly is not available in traditional Ethernet-based MAC. Details on the GPON GEM are given in Section 3.5.2.

3.1.5 Dynamic Bandwidth Allocation, Ranging, and Discovery

Dynamic Bandwidth Allocation To avoid packet collision in the upstream direction, a channel arbitration technique must be in place to avoid data collision and share the bandwidth resources fairly to the ONUs. Both EPON and GPON employ centralized dynamic bandwidth allocation (DBA) methods. Centralized DBA can guarantee service to individual ONUs while ensuring high bandwidth efficiency. Alternatively, a distributed arbitration method either cannot support guaranteed service, or else would require an ONU to communicate bandwidth arbitration messages. This would take a very long time because the fiber round-trip time (RTT) is approximately 200 μs for a 20-km PON. Because EPON and GPON use different frame formats, the DBA algorithm and operation look very different in each standard. However, in each case the DBA would allocate upstream time slots to different ONUs such that no collision is observed at the OLT. This would require accurate ranging and timing to measure the RTT between the OLT and ONU such that the OLT can appropriately offset the upstream time slots granted to ONUs to avoid upstream packet collision.

Ranging and Discovery Ranging is carried out in the autodiscovery process. In either EPON or GPON, the autodiscovery process allows newly powered-up (cold) ONUs or ONUs that have lost synchronization (warm) to join the network. Besides making logical registration of new ONUs (cold), the autodiscovery process additionally performs two functions. First, it allows an OLT to use the receiver burst to lock its clock to the receiver ONU clock and to adjust the gain level. Moreover, it performs a ranging process to measure the RTT between the OLT and the ONU.

3.1.6 Reliability and Security Concerns

Reliability Because standard TDM-PON employs a passive and point-to-multipoint architecture, the inherent reliability and security of the network are weak. In PON systems, service downtime could occur due to fiber cuts, splitter failures, and transceiver failures. The location of the failure could also be categorized at the truck fiber, at the distribution fiber, at the splitter, at the OLT, or at the ONU. Each type and location of failure could cause service downtime for every user (a trunk group) of PON or a subset of users (a branch group). Service downtime could be estimated by multiplying mean time to repair (MTTR) and the failures in time (FIT) factors. An unprotected tree PON architecture, as illustrated in Figure 3.1, is expected annually to experience approximately 40 minutes of service downtime. There are two ways to reduce the service downtime expected. The first approach enhances the optical diagnostic and monitoring techniques to ensure that the cause of the failure is identified correctly and efficiently when a failure is first observed. The second approach enhances the PON architecture by introducing redundant fiber path and automatic protection switching (APS) techniques. Although most of the TDM-PON deployment does adapt to

an unprotected tree architecture, ITU-T G.983.1 does suggest four types of protection switching architectures with different protection levels. Moreover, G.983.5 and G.983.2 standards specify the reference model for resiliency and redundancy as well as specifying the control and management interface to support the protection features. Details of these architectures and others are described later in the chapter. IEEE does not specify any particular protection switching or suggest any redundant switching architecture. EPON vendors, however, could elect to provide their own proprietary solutions, and Japanese operators are known for deploying ring-type fiber truck networks with APS capability.

Security In the downstream direction, TDM-PON signals are broadcast physically to all ONUs even if the logical traffic is only intended to be sent to one or a few ONUs. As a result, the downstream channel is highly susceptible to eavesdropping by malicious users. The biggest security weakness occurs during the ranging process, when the OLT broadcasts the identification serial number of the ONU ranged. Potentially, a malicious user could use this information to masquerade as a user victim or simply as a spoof. To improve the security, G.983.1 initially specifies a churning procedure to scramble downstream data with a key established between an ONU and an OLT. Although churning does provide some enhanced level of security, encryption should be employed in the application layer when security concerns become very important. ITU-T 984.3 further recommends the use of the 128-bit advanced encryption standard (AES) technique in GPON to enhance downstream security. The IEEE 802.ah EPON standard again does not specify any downstream encryption standard, but many ASIC vendors today do provide their own encryption mechanisms on commercial EPON chip sets.

3.2 PON STANDARDS HISTORY AND DEPLOYMENT

3.2.1 Brief Developmental History

The history of PON could be dated back to the early 1990s, when an FSAN consortium studied the feasibility of extending fiber optics access service to end users. Officially, APON was proposed in 1995 by FSAN to provide business customers with multiservice broadband offerings. ATM (multiservice) and PON (inexpensive) presented the best mix for achieving that target. However, ATM does not live up to the expectation of becoming the universal network protocol; instead, IP and Ethernet quickly evolved into that role. As a result, the ITU-T study group 15 (SG15) deliberately chose the name *broadband PON* (BPON) to avoid a naming association with the ATM technology. Nevertheless, BPON still relies on ATM networking by nature. BPON defines a line rate as a multiple of 8 kHz, the most basic frame repetition rate for SONET/SDH. An OLT distributes the 8-kHz clock from the OLT to an ONU to support TDM services such as voice applications more easily. The ITU-T G.983.1 BPON standard was published originally in 1998 and redefined in 2005 to add a higher line-rate option.

TABLE 3.1 Comparison of Standardized BPON, GPON, and EPON Parameters

Parameter	BPON	GPON	EPON
Maximum/diff. reach	20 km/20 km	60 km/20–40 km	20 km/20 km
Maximum split ratio	128	128	—
Line rate (up/down)	155.52–622.08/ 155.52–1244.16 Mb/s	1244.16/2488.32 Mb/s	1250/1250 Mb/s
Coding	NRZ + scrambling	NRZ + scrambling	8b10b
Data rate	Equals the line rate	Equals the line rate	1000 Mb/s
ODN loss tolerance	20/25/30 dB 28 dB (best practice)	20/25/30 dB 28 dB (best practice)	20/24 dB 29 dB (actual systems)
US overheads	Fixed 3 bytes 154 ns (155.52 Mb/s) 38.4 ns (622.08 Mb/s)	Guard: 25.7 ns Preamble: 35.4 ns Delimiter: 16.1 ns	Guard: 2 μs Laser on/off: 512 ns AGC/CDR: 400 ns

Due to its underlying ATM structure, BPON ONUs and OLTs are required to implement ATM switching capabilities. ITU-T G.983 actually specifies OLTs and ONUs as a whole to function as virtual path (VP) and virtual connection (VC) switches. Moreover, an additional translation between ATM and Ethernet needs to be implemented at the ONU user network interface (UNI) and OLT service network interface (SNI). These complications make BPON equipment unfavorably costly and hinder the growth of BPON systems in a fast-evolving broadband market. Gigabit-capable PON (GPON) is developed to be the next-generation PON technology developed by ITU-T after BPON. GPON is specified such that it can cope better with the changes toward Ethernet and IP communication technologies and meet the fast-growing bandwidth demands. The GPON standard was first rectified by the ITU-T G.984 standard in 2003.

At approximately the same time, the EPON standard was being developed independently by the IEEE standard body to extend basic Ethernet LAN MAC to support new fiber access physical layers. The work of EPON was begun in March 2001 by the IEEE 802.3ah FEM study group and finished in June 2004.

Table 3.1 summarizes the operational parameters for BPON, GPON, and EPON standards. As explained earlier, note that all B−, G−, and EPON have converged toward a class B+ type of optics. Class B+ (28 dB ODN loss tolerance) can usually support up to 20 km of reach and a 64 split ratio. The upstream burst overhead for GPON is calculated based on a 1.244-Gb/s bit rate.

3.2.2 FTTx Deployments

Optical access networks come in many flavors: point-to-point (P2P), active, or passive point-to-multipoint (P2MP). The term *FTTx* is used broadly and refers to "fiber to the x," where x could refer to one of several popular deployment scenarios. Figure 3.4

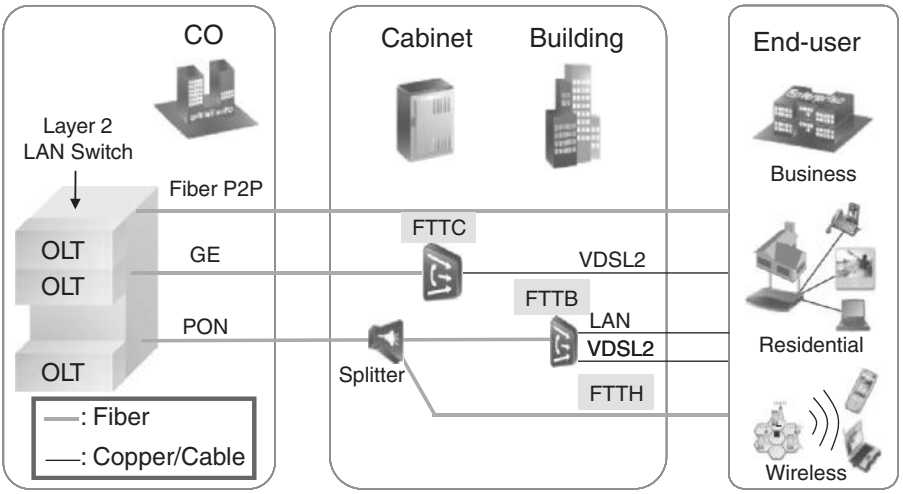


FIGURE 3.4 Architecture for an optical access network.

shows the most common reference FTTx deployment types. In the figure, fiber to the cabinet (FTTC), fiber to the building (FTTB), and fiber to the home (FTTH) represent different levels of fiber penetration. FTTC and FTTB refer to deployment where end users are connected through the hybrid use of coaxial cable (also called HFC), twisted pair, or radio. The differences between FTTC and FTTB can be seen by the level of fiber penetration, where FTTC generally connects to a remote active cabinet and FTTB places ONU inside a multidwelling unit (MDU). For the FTTH type of deployment, fiber to the business (FTTB) and fiber to the premises (FTTP) are used to further indicate connection to either business or residential users. Sometimes, in perhaps a confusing manner, the term fiber to the curb (FTTC) is used interchangeably with FTTB, and the term fiber to the node (FTTN) is also used interchangeably with FTTC.

Although each type of FTTx deployment could be understood with a different level of fiber penetration, most users of the term FTTx do not necessarily describe the particular physical configuration. For example, FTTC and FTTN are often used interchangeably, and FTTP is sometimes viewed as a generic term for either FTTB or FTTH. Users employ the term FTTx to indicate expectations regarding service-level agreements (SLAs), customer devices, service offerings, and investment payback periods.

3.3 BROADBAND PON

Since the initial rectification by ITU-T in 1998, broadband PON (BPON) has been deployed in North America by the regional Bell operating companies for both FTTN and FTTP types of deployment. ITU-T 983.1 specifies the reference architecture,

transceiver characteristics, transport fame structure, and ranging functions in BPON. In addition, the remaining 983.x standards make explicit recommendations in the following areas:

- G.983.2: ONT management and control interface
- G.983.3: third wavelength band for downstream video transmission
- G.983.4: dynamic bandwidth assignment
- G.983.5: enhanced survivability
- G.983.6: control and management interface to support protection feature

Note that user terminals are called optical networking terminals (ONTs) instead of ONUs in the ITU standards. In this chapter, the term *ONT* is used when referring to BPON or GPON, and the term *ONU* is used when referring to EPON systems. In the rest of this section we present details of the BPON standard to guide readers to an understanding of BPON operation as well as the implications during upgrade from BPON to future PON systems.

3.3.1 BPON Architecture

Permissible Data Rates The original ITU-T 983.1 standard supported 155.52 Mb/s (OC-3) and 622.08 Mb/s (OC-12) in either the downstream or upstream direction. Later, a new 1244.16 Mb/s (OC-24) transmission rate is added in response to increasing bandwidth demand. BPON leaves the choice of a specific combination of downstream and upstream rates to vendors, and Table 3.2 shows the permissible downstream/upstream data-rate pairs for BPON.

Wavelength Allocation The original downstream and upstream wavelength plan for BPON was set at 1260 to 1360 nm (1310 $\pm$ 50 nm) and 1480 to 1580 nm (1530 $\pm$ 50 nm), respectively. One hundred nanometers of bandwidth window was selected to give sufficient bandwidth after taking into account the fiber dispersive effect and the use of inexpensive FP laser components for the upstream band. In response to the pressing demand for broadcast video and data, however, BPON adds an additional downstream band at the original G.983.1 DS band.

TABLE 3.2 Permissible BPON Downstream/Upstream Data-Rate Pairs

Downstream (Mb/s)	Upstream (Mb/s)
155.52	155.52
622.08	155.52
622.08	622.08
1244.16	155.52
1244.16	622.08

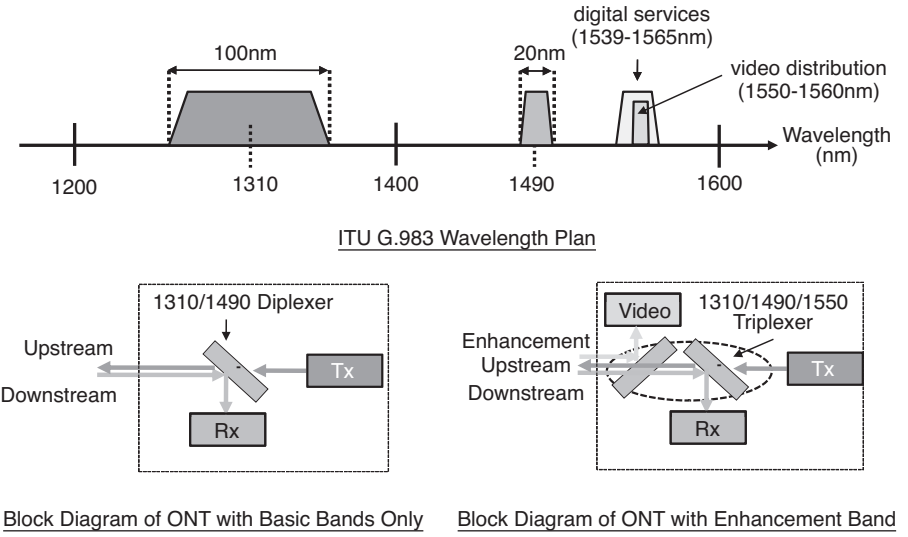


FIGURE 3.5 BPON wavelength plan and ONT options.

Figure 3.5 shows the wavelength allocation plan for BPON. This wavelength plan is also used for GPON systems. Two enhancement band options provide the ability to carry analog TV overlays (option 1: 1550 to 1560 nm) or additional digital services such as SONET/SDH links for high-bandwidth TDM services (option 2: 1539 to 1565 nm). Additional wavelength filters must be added to the ONT to support the enhancement band wavelengths. Additional filter losses and wavelength crosstalk must also be considered in enhancement-capable systems. These enhancement filter loss and isolation requirements are specified as part of the G.983 standard.

Video Delivery Architectures Enhancement band option 1 was used in the BPON system to overlay analog video transmission. Figure 3.6 depicts a common analog video overlay architecture. In this architecture, video is overlaid in the 1550-nm enhancement band using a subcarrier multiplexing (SCM) technique. The source of the video is fed from a remote headend that receives, for example, the traditional radio TV broadcast mentioned in Section 1.4.1. The RF-modulated signals are then subcarrier-multiplexed onto a 1550-nm laser source at the video server. The benefit of this architecture is that users are not required to use anything other than the traditional analog television. The number of broadcast channels that can be supported is limited, however. More channels can still be supported under this architecture, but it would require narrowcast transmission architecture (i.e., the ability to switch the receiving channels based on user selection). The dotted component in Figure 3.6 further illustrates narrowcast support in analog video overlay architecture. In this architecture, the user selects the channel using a set-top box with analog tuning ability, such as a cable set-top box, to communicate with the headend via Internet

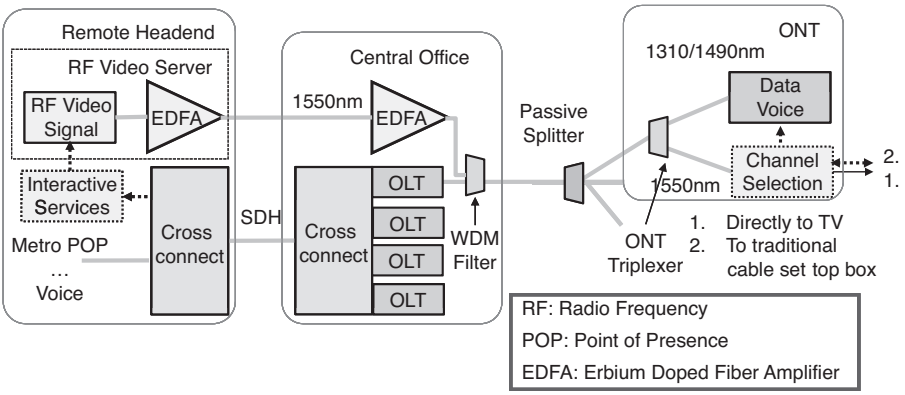


FIGURE 3.6 BPON analog video overlay architecture.

group management protocol (IGMP). The headend video server then determines the proper program and thus the RF spectrum to transmit over PON.

With the advent of high-definition (HD) TV, and as demands for in-home distribution of multimedia contents continue to rise, the analog video transmission model can no longer sufficiently support the bandwidth scalability and user responsiveness under the new demands. On the other hand, in-band IP-based video and the multimedia delivery model address these issues by supporting on-demand content delivery and enable more interactive capability between a home and a video service provider. An IP-based solution also has the benefit of converging video with voice and data transmission. Figure 3.7 shows the in-band IP video distribution architecture, in which the ONT no longer requires a triplex filter. For other than technical reasons, a user needs to obtain a new set-top box from the video service provider (VSP) to support the video distribution services via IGMP and/or the video on demand (VOD) protocol. To support fast channel selection, the VSP builds a hierarchy of VOD servers at the headend and access central office to cache popular content nearby the users. The in-band IP-based model is gaining momentum in recent deployments as access

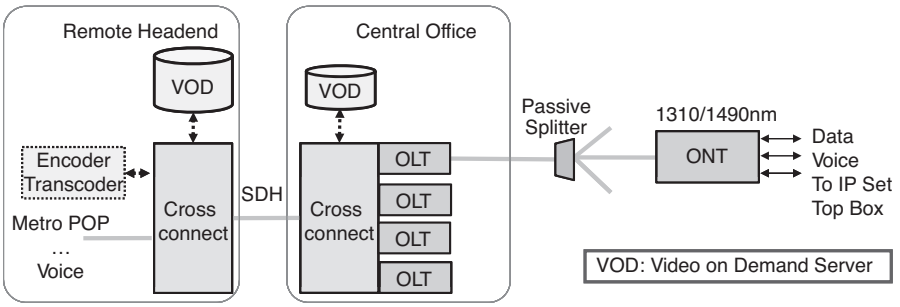


FIGURE 3.7 BPON in-band IP video delivery architecture.

TABLE 3.3 Comparison of Analog, IP, and Hybrid Video Delivery Architectures

	Pros	Cons
Analog	Mature technology	Analog video quality
	No new set-top box required	Requires ONT triplexer
	Scalable to a large take rate (broadcast)	Not a converged network
IP	Supports HDTV and better video quality	Requires IP set-top box
	More interactive	Data bandwidth is shared
	No ONT tripler required	Channel selection requires IGMP
	A converged network	Does not scale well
Hybrid	Supports HDTV, interactive, and more	Not a converged network
	Large combined bandwidth	Requires triplexer, set-top box, and IGMP
	Scales well	

operators aggressively migrate into providing HDTV to compensate for continuing revenue erosion in traditional voice and data services.

Where analog video overlay infrastructure is already in place, there is an additional benefit to employing a hybrid IP/analog video delivery architecture. The hybrid infrastructure allows the operator to support a combined large bandwidth in the PON data and RF overlay spectrum. Since most subscribers already have an RF-based solution, IP service could be used additionally to support a narrowcast type of service. Table 3.3 summarizes the pros and cons of employing each of the three most common video delivery architectures. In short, IP service is expected to provide better-quality pictures and interactive features, but at a greater price. In reality, the choice for the specific deployment depends heavily on the availability of legacy video transmission infrastructure and the market in the particular community (i.e., the take rate for IP service expected if they were to be deployed).

Interoperability BPON stresses vendor interoperability in system implementations. It is important because interoperability allows operators to exercise choices in selecting either OLT or ONU vendors. In particular, an OLT card is expected to be able to support ONTs implemented from other vendors, and vice versa. In addition, the OLTs should interact with operators' heterogeneous mix of service network interfaces. As a result of the interoperability requirement, however, the cost of BPON equipment suffers because it has to implement additional SNI ports that it may not be used in certain deployment scenarios.

3.3.2 BPON Protocol and Service

BPON transport and signaling protocol is specified based on ATM since telecommunication network operators had had extensive ATM switching infrastructure at the time that ITU G.983 recommendations were being developed. Besides being a mature technology, ATM also supports scalable traffic management and robust quality-of-service (QoS) controls. As a result, BPON protocol is very closely related to ATM networking.

3.3.2.1 ATM Cell and Services ATM is a circuit-switching technique based on fixed-size 53-byte (53B) packets called *cells*. The impetus for a short fixed-size cell is the ability to switch or multiplex quickly between traffic flows. This is an important feature during the development of ATM because it is designed primarily to support voice traffic. Small data cells enable an ATM network to reduce jitter by multiplexing data streams. In other words, urgent voice cells can be multiplexed within 53B intervals, avoiding the large jitter that could otherwise be incurred by large data cells.

An ATM cell consists of a 48B payload and a 5B header. The 5B header has two types of format: a user–network interface (UNI) and a network–network interface (NNI). An ATM header supports a virtual path identifier (VPI), virtual circuit identifier (VCI), payload-type indicator (PTI), cell loss priority (CLP), and header error check (HEC) (Figure 3.8). The UNI format also supports a 4-bit (4b) generic flow control (GFC) field. In BPON, the GFC field is not used and is always set to zero when a UNI format is used.

The details of the fields are described below.

- *VPI*. A virtual path is a logical bandwidth pipe that can consist of multiple virtual connections. Depending on whether the ATM cell is sent via a UNI or NNI formatted header, the VPI header supports either 8b (256) or 12b (4096) virtual path connections. The G.983 recommendation limits the number of virtual paths to less than 4096.
- *VCI*. Virtual connection is the lowest level of connection defined in an ATM network. VCI is 16b long and supports up to 65,536 different virtual channels for each virtual path. The combination of VPI and VCI identifies a unique virtual circuit for the ATM cell in a ATM port, or in each PON. A virtual circuit is allocated and defined through specified QoS characteristics.
- *PTI*. PTI is 3b long and the first bit indicates whether the cell is user data (0) or OAM data (1). The second bit indicates if the cell has experienced congestion (yes = 1; no = 0) during its transmission. If the current switch experiences

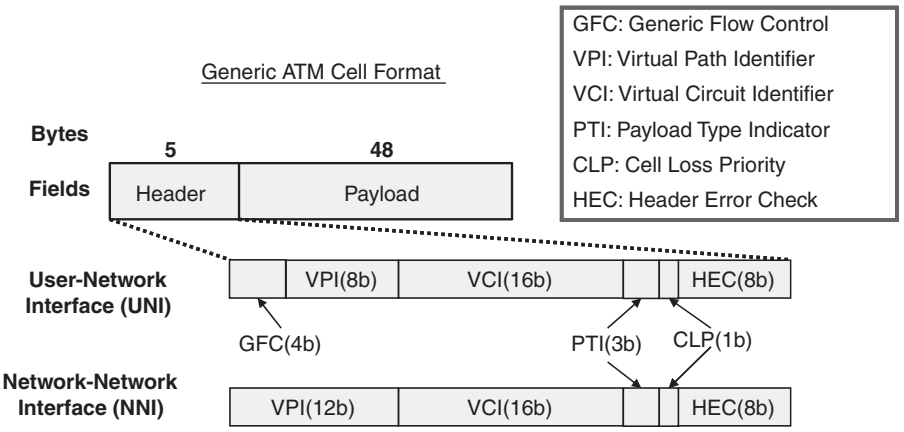


FIGURE 3.8 ATM cell format.

congestion switching the cell, the bit is turned to 1. If the cell is OAM data, the third bit is used for OAM functions. Otherwise, the third bit is used to indicate if the cell is the last in a series of ATM adaptation layer 5 (AAL5) cells. ATM defines several AAL layers to support different levels of QoS. AAL5 is used for packet data communication that could be segmented at the transmitting end and reassembled at the receiving end in a connectionless service.

- *CLP*. CLP is a 1b field to indicate the priority of the cell. CLP = 0 indicates a high priority and the interim switch must forward the cell with its every effort, even with congestion. CLP = 1 indicates that the interim switch could drop considerably in congestion situations. An example of a cell labeled as CLP = 1 is when the cells represent additional traffic above the committed information rate (CIR) in a negotiated connection.
- *HEC*. HEC is a 8b checksum to protect the ATM header.

An important characteristic of ATM is its ability to support different levels of QoS. In ATM networks, five classes of QoS are specified, as summarized in Table 3.4. Each QoS service class is further characterized by the parameters described below.

According to the service-level agreement (SLA), the following user-rate parameters are defined in terms of cells per second:

- *Peak cell rate* (PCR): maximum rate at which a user can be allowed to transmit
- *Sustainable cell rate* (SCR): average cell rate measured over a period
- *Minimum cell rate* (MCR): guaranteed transmission rate of a user by a service provider according to the SLA

In addition to SLA-defined rate parameters, users also negotiate network-related traffic attributes according to different QoS service classes. These attributes include:

- *Cell transfer delay* (CTD): end-to-end cell delays
- *Cell delay variation* (CDV): measured difference between minimum and maximum CTD

TABLE 3.4 ATM QoS Service Classes

Service Class	Description	AAL Level	Example
Constant bit rate	Guaranteed PCR, supports real-time applications	1	Voice and VOD
Variable bit rate real time (VBR-rt)	Specified PCR and SCR, guaranteed CDV	2	Video, packetized voice
Non-real time	Similar to VBR-rt without guaranteed CDV	—	Bank transactions, reservation systems
Available bit rate	Guaranteed MCR, PCR limited	—	Credit card validation, e-mail, fax, etc.
Unspecified bit rate	PCR limited, analogous to best-effort traffic	5	General TCP/IP applications

BPON Layers		Functions
Path Layer		ATM cell/ BPON frame conversion Cell transmission
Transmission Convergence (TC) Layer	Adaptation Sub-layer	Network/user frame conversion
	PON Transmission Sub-layer	Ranging Cell slot allocation Bandwidth allocation Privacy and security Frame alignment Burst synchronization Bit/byte synchronization
Physical Medium Dependent (PMD) Layer		O/E/O conversion Fiber connection

FIGURE 3.9 BPON protocol layer structure.

3.3.2.2 BPON Protocol Layers and Functional Blocks Figure 3.9 shows the BPON protocol layer structure as defined in the G.983.1 recommendations. The protocol layer structure consists of a physical media dependent layer (PMD), a path layer, and a transmission convergence layer (TC). The PMD layer defines the bit rate, line code (NRZ), operating wavelengths, and transceiver requirements.

In BPON, the path layer corresponds to the ATM virtual path (VP) layer as defined by ITU-T recommendation I-732. The OLT and ONT in BPON as a whole can be viewed as a VP switch. G.983.1-specified functions of the OLT and ONT are presented to illustrate the concept. Figure 3.10 shows the OLT and ONT functional blocks.

Details of the function for these OLT and ONT functional blocks are provided below.

OLT Functional Blocks

- *Service port*. OLT service ports interface with an SNI to connect to the core network. The service port handles the insertion of ATM cells into the upstream SONET/SDH payloads and retrieval of ATM cells from the downstream payloads. An additional feature such as protection switching could also be implemented at the service port level.
- *Transmission multiplexer* (MUX). An OLT MUX provides VP connections between the service port and the ODN interface. Different VPs are assigned to different services at each PON. Different information types, such as data, signaling, and OAM flows, are classified further by VCs of the VP.
- *ODN interface*. The line terminal handles the optoelectronic conversion process. Moreover, the ODN handles the insertion of ATM cells into the downstream transmission frames and extraction of ATM cells from the upstream frames.

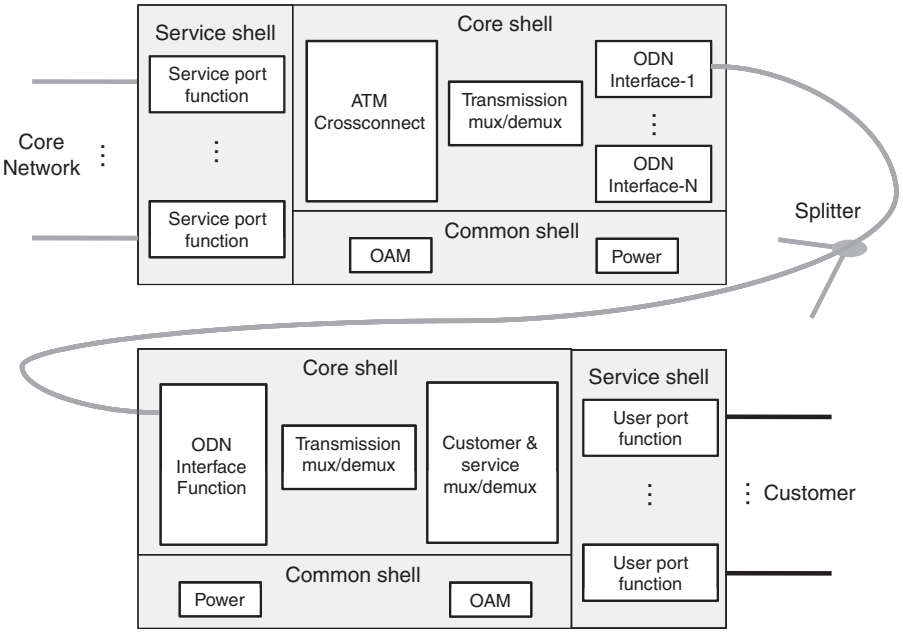


FIGURE 3.10 BPON OLT and ONT functional blocks.

ONT Functional Blocks

- *User port*. ONT user ports interface over UNI to connect to customer terminals.
- *Transmission multiplexer* (MUX). ONT MUX multiplexes service interfaces to the ODN interface. Only valid ATM cells would pass through the ONT MUX.
- *ODN interface*. This is similar to the ODN interface at OLT. The ODN interface at ONT handles the optoelectronic conversion process and the conversion of ATM cells to and from BPON transmission frames. It is worthwhile to note that the interface must acquire synchronization on the downstream frame bit before transmitting upstream frames.

The Recommendation G.983 series does not specify the functions of VC. However, its use is left open to vendor implementations. In addition to providing VP switching and optoelectronic conversion, the OLT and ONT combine to provide OAM functionalities. Details of the OAM, as well as other topics regarding the TC layer, are described next.

3.3.3 BPON Transmission Convergence Layer

In BPON, information is transmitted in the form of either an ATM cell or a physical-layer operation, administration, and maintenance (PLOAM) cell. Both cells have a fixed 53-byte (53B) format. ATM cells are used to carry data, and PLOAM cells are

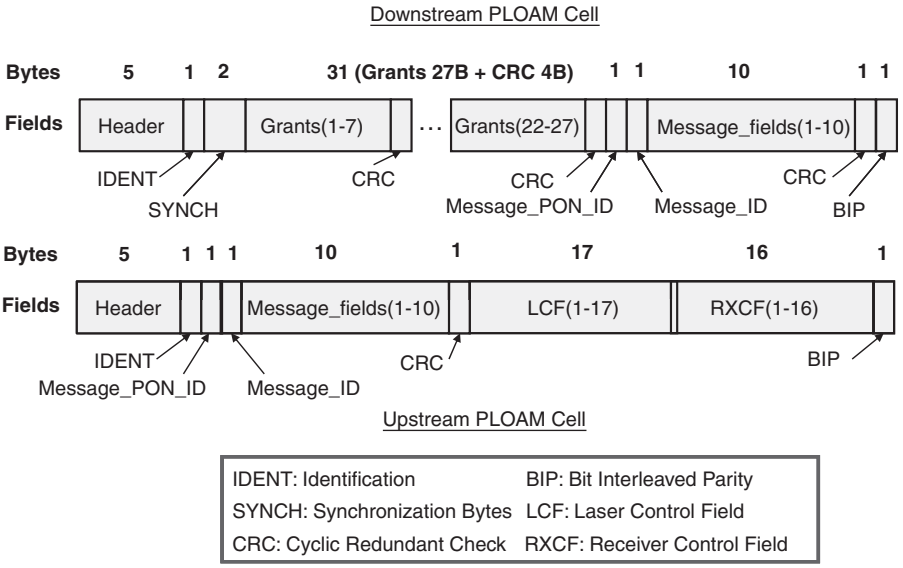


FIGURE 3.11 BPON PLOAM cell format.

used to carry physical-layer management information. ATM and PLOAM cells are transmitted in a fixed frame structure.

3.3.3.1 PLOAM Cell Format Since the ATM cell format has been introduced, the PLOAM cell format is described here before introducing BPON transmission frame structures. Figure 3.11 shows the downstream and upstream PLOAM cell formats.

The PLOAM cell uses the same 5B header as the ATM cell. The rest of the PLOAM fields are described below.

Downstream PLOAM

- *Identification* (IDENT). The last bit of the IDENT byte is set to "1" if it is the first PLOAM cell of a downstream frame. The PLOAM identification field is used further by ONT to synchronize with the downstream frame by correctly detecting the PLOAM header and receiving the "1" IDENT frame bit.
- *Synchronization bytes* (SYNCH). This 2B SYNCH field is used to transport a 1-kHz reference clock signal from the OLT to the ONT. At the OLT, a counter increments every 51.4 ns ($t_{\text{interval}} = 8$ bits/155.52 Mb/s) or every 1B when operating at a 155.52-Mb/s bit rate. The OLT writes the least significant 15b of its counter in the least significant 15b of the SYNCH field. The most significant 1 bit of the SYNCH field is filled with the most significant bit of the OLT counter. ONT synchronizes its local counter upon the reception of the SYNCH field. The counter is reset every 1 ms at the OLT to convey the 1-kHz reference signal.

- *Grants*. Each PLOAM cell has 27 grants. Each 1B grant could contain either a data grant, a PLOAM grant, a divided slot grant, a ranging grant (FD_16), an unassigned grant (FE_16), or an idle grant (FF_16). Each grant corresponds to an upstream time slot in the upstream frame. The value of the data grant corresponds to 1B ONT ID (0–63) for indicating an upstream time slot grant. The ONT can response a data grant with either an upstream data cell or an idle cell. On the other hand, an ONT always responds to a PLOAM grant with an upstream PLOAM cell. A divided slot grant is used to assign a minislot in the DBA process, which would be described in later sections. Unassigned grants are used to indicate an unused upstream slot. When an idle grant is transmitted, the ONT simply ignores the idle grant. An idle grant is used when the number of upstream slots available is less than the number of grants available in the PLOAM cells, which is the case when the downstream operating bit rate is above 155.52 Mb/s. The type of grant is indicated by the message field.
- *Downstream PLOAM messages*. A downstream message is defined by Recommendation G.983 to convey information such as ranging, churning key assignment, VP or VC configuration, and so on. In the PLOAM cell, the message is further divided into three fields. The PON ID is a unique ONU number assigned during the ranging process. This PON ID could be from 0 to 63. The format for the message type and message contents is available in Section 8.3.8 of Recommendation G.983 and is outside the scope of this book.
- *Cyclic redundancy check* (CRC). The CRC fields protect groups of seven grants using generating polynomials $g(x) = x^8 + x^2 + x + 1$. It is able to detect up to 3 bits of error. Error correction is not performed and the entire block of grants is ignored when an error is detected. For the last group of six grants (22–27), a dummy seventh grant with value 00_16 is assumed. For the message group, the CRC field is the reminder of the division between $g(x)$ and the polynomial created with the content of the 12B message field.
- *Bit interleaved parity* (BIP). The BIP field is used to monitor the bit error rate (BER) in the downstream link. The BIP is a good estimate of the BER when the BER is smaller than 10^{-4}.

Upstream PLOAM

- *IDENT*. The IDENT field is not used in upstream PLOAM cells.
- *Laser control field* (LCF) and *receiver control field* (RXCF). In concept, LCF and RXCF are used by the ONT to maintain the nominal optical output power and by the OLT to recover the appropriate gain threshold level. The implementation of LCF is vendor dependent and depends on the type of laser driver employed by the ONT.
- The rest of the upstream fields (i.e., message fields, CRC, and BIP) have properties similar to those of the corresponding downstream PLOAM fields.

3.3.3.2 Transmission Frames The BPON downstream bit rate operates at 155.52, 622.08, or 1244.16 Mb/s. The downstream transmission frame has a fixed

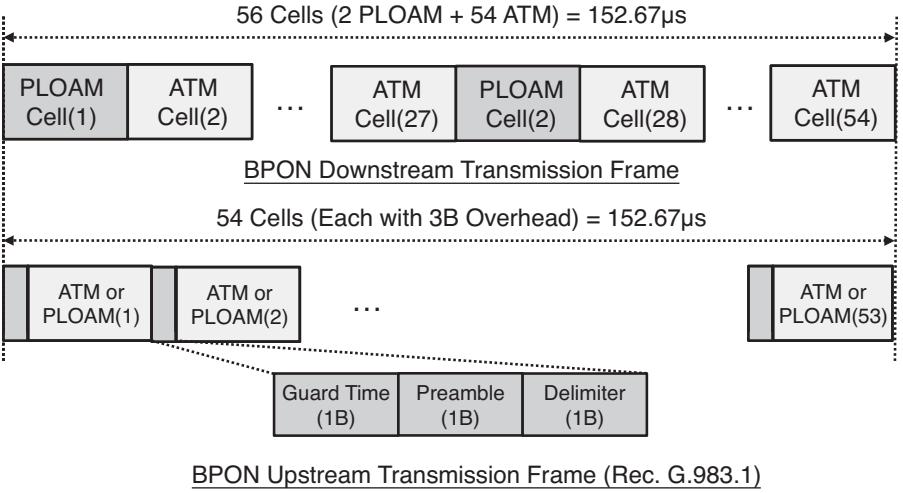


FIGURE 3.12 BPON transmission frame (at 155.52 Mb/s).

size of 152.67 μs. Figure 3.12 illustrates a downstream transmission frame at a rate of 155.52 Mb/s. The frame contains two PLOAM cells and 54 ATM cells, where the PLOAM cell is inserted in between every 27 ATM cells. The downstream frame size is multiplied by 4 to 224 cells for a 622.08-Mb/s bit rate and multiplied by 8 to 448 cells for a 1244.16-Mb/s bit rate. Downstream frames are broadcast in a TDM fashion, and each ONT will only pass on the ATM data that are destined to its subscriber based on the VPI/VCI combination.

The upstream frame contains 53 time slots, each of which is 56B long. Each upstream time slot contains a 3B overhead in addition to the 53B ATM/PLOAM cell. The content of the overhead field is programmed by the OLT and is defined by the Upstream_overhead message in the downstream PLOAM cells. The purpose of these 3B overheads is noted below.

- *Guard time*. The guard time is reserved for laser on/off time. It provides enough distance between two consecutive cells to avoid a collision. The minimum guard time length is 4b, which corresponds to 25.6 ns for a 155.52-Mb/s bit rate. The guard time could be expanded to 1B long to 12.8 ns for 622.08-Mb/s bit rate.
- *Preamble*. The preamble defines the clock and data recovery time in BPON. Bit-level synchronization is achieved here. The preamble is only 1B long and the BPON has a stringent clock and data recovery time requirement of 51.2 ns for 155.52 Mb/s and 12.8 ns for a 622.08-Mb/s bit rate. In BPON, the guard time plus the preamble is equivalent to the inter-ONU guard time in Figure 3.3.
- *Delimiter*. The delimiter is a unique pattern used to indicate the start of an ATM or PLOAM cell. Byte-level synchronization is achieved using the delimiter field.

An ONT sends an upstream PLOAM when it receives the PLOAM grant from OLT. The PLOAM cell rate is defined by OLT, and for each ONT the minimum PLOAM cell rate is at least one PLOAM cell for every 100 ms.

3.3.3.3 Upstream TDMA BPON upstream employs time-division multiple access (TDMA), where each ONT takes a turn in transmitting an upstream time slot. The assignment of ONT transmission to a time slot is allocated through the corresponding data grant in the downstream PLOAM cell. The PLOAM cell is available to all ONTs. An ONT always checks the PON ID in the data grant and compares it to its own assigned serial number. If there is a match, the ONT can then send its information or send an idle data cell when it does not have information to send.

In the case of a symmetrical 155.52-Mb/s transmission bit rate, each upstream frame contains 53 time slots and the downstream frame needs to provide the corresponding grants, which are mapped to the two PLOAM cells. As a result, the last grant of the second PLOAM cell is an idle grant (FF_16). For a symmetrical 622.08-Mb/s transmission bit rate, the last grant of every even-numbered PLOAM cell is an idle grant. For an asymmetrical transmission bit rate (e.g., 622.08-Mb/s downstream and 155.52-Mb/s upstream), PLOAM cells 3 through 8 are filled with idle grants.

In the early BPON implementations, the bandwidth was guaranteed at a fixed value to the specific subscriber. The service sold to the subscriber is fixed and an early scheme makes fixed bandwidth allocations to the subscribers. However, this method is not efficient, as a large amount of bandwidth is either under- or overutilized, due to the bursty nature of Internet traffic. As a result, a more efficient method of allocating bandwidth to each user dynamically is amended by Recommended G.983.4.

3.3.3.4 Ranging Since the ONT sends information to the OLT using time-division multiple access (TDMA), each ONT must be precisely synchronized with all other ONTs. To achieve synchronization across all ONTs, the OLT first uses the ranging process to determine the distance of each ONT from the OLT. In BPON, the maximum logical differential distance is 20 km. This corresponds to 200 μs of round-trip delay (RTT) or 69 cells (56B) + 192 bits. After ONT distances and their relative single trip-delay differences are known, the OLT uses the grant assignment to ensure that upstream slots from any two ONTs would not collide.

In addition to measuring the logical distance between the ONT and the OLT, the ranging procedure is also used to connect to new ONT (cold) or in-service ONT (warm) that has detected lost byte-level synchronization. In the case of cold ONT, the OLT assigns the PON ID after the ranging process. Figure 3.13 illustrates the BPON ranging procedure. In sequence, the OLT first sends an unassigned grant (FE_16) to open the ranging window. Then the OLT sends ranging grants in successive downstream PLOAM cells (downstream data and upstream data grants are queued during the ranging procedure). The ONT immediately responses to the ranging grant with the ranging cell. Upon receiving the ranging cell, the OLT measures the arrival phase of the arriving cell and notifies the ONT of the equalization delay. The ONT then adjusts the transmission phase to the notified value to achieve bit-level synchronization.

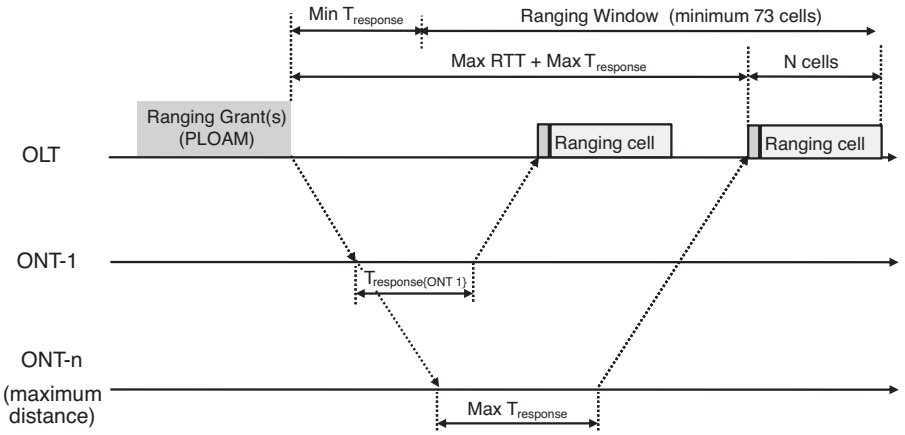


FIGURE 3.13 BPON ranging procedure.

BPON ONTs are required to process PLOAM grant information before PLOAM message fields. The standard further mandates a response time between 3136b and 4032b (at 155.52 Mb/s), which translates to seven to nine cells. The response time, T_{response} is specified to allow sufficient signal processing time at the ONT. As a result, the ranging window is considered open after at least minimum T_{response} since the first ranging grant is sent. In addition, the ranging window is reserved for at least 73 cells (minimum T_{window}) to receive at least one ONT ranging cell. This is because minimum T_{response} + minimum T_{window} − maximum T_{RTT} − maximum $T_{\text{response}} = 2T_{\text{cell}} - T_{192\text{bits}} > 1T_{\text{cell}}$. The ranging window could be expanded if the OLT is expecting more than one ranging cell.

During the ranging process, multiple ONTs could respond to the broadcast ranging request at the same time and cause a collision. To resolve the contention, a binary tree mechanism is used to range the ONTs individually. The binary tree mechanism sends a Serial_number_mask message before a ranging grant to all ONTs in the standby state. The Serial_number_mask message is increased with one bit at a time until only one ONT responds with the ranging cell. In a binary tree mechanism, ONT responses to the ranging grant only when its unique serial number matches the Serial_number_mask message. After the contention is resolved, the remaining ranging would be performed using normal ranging grants.

3.3.4 BPON Dynamic Bandwidth Allocation

Dynamic bandwidth allocation (DBA) is a methodology to allow OLT to reallocate bandwidth quickly across the entire PON according to current traffic conditions. In the original Recommendation G.983.1, DBA is not specified. To improve upstream bandwidth use, Recommendation G.983.4 adds DBA to BPON. G.983.4 also introduces traffic container (T-CONT) buffers, and all bandwidth grants are allocated in the T-CONT level.

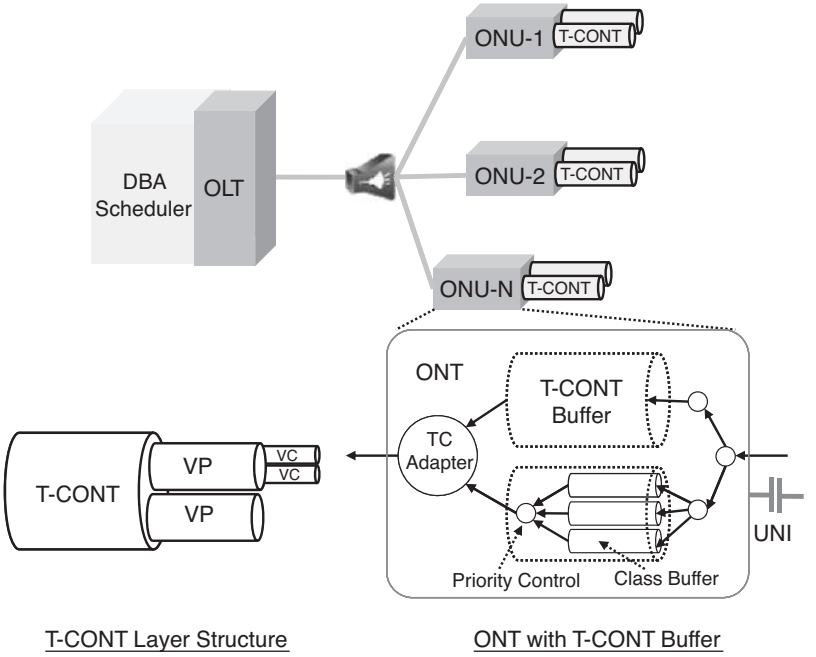


FIGURE 3.14 BPON T-CONT architecture.

3.3.4.1 T-CONT and Divided Slot In a BPON DBA process, grant allocation is associated directly with T-CONT-type classification. This classification enables PON to better handle traffic flows with similar characteristics. An ONT can have multiple T-CONTs, and the T-CONTs in an ONT operate independent of one another. Figure 3.14 illustrates the BPON T-CONT architecture.

A T-CONT can multiplex multiple streams of traffic, and each ONT could have multiple T-CONTs. Recommendation G.983.4 further classifies T-CONT into five different types. OLTs could create or delete T-CONTs for an ONT depending on the current traffic and operational parameters specified in the SLA. Examples of the parameters include service priority and QoS. The priority of the data grants include fixed bandwidth, assured bandwidth, nonassured bandwidth, and best-effort bandwidth. Both fixed and assured bandwidth are guaranteed, but bandwidth is reserved only during actual communications. Nonassured and best-effort bandwidth are assigned dynamically under control of the DBA algorithm. Each one of the T-CONT type is specified to support one or more of these priority classes. Table 3.5 summarizes the classification of the various T-CONT types.

3.3.4.2 DBA Protocol To dynamically determine the number of data grants assigned to a T-CONT, the OLT needs to know the traffic status of the T-CONT. Two possible schemes are specified by Recommendation G.983.4. In an ONT status reporting (SR) scheme, the ONT reports T-CONT buffer depths to the OLT. In an

TABLE 3.5 BPON T-CONT Types

Traffic Container	Assignable Bandwidth	Examples
T-CONT1	Fixed	Real-time services (e.g., voice)
T-CONT2	Assured	Non-real-time services (e.g., CBR video)
T-CONT3	Assured + nonassured	Non-real-time services (e.g., VBR video)
T-CONT4	Best-effort	TCP/IP services (e.g., file transfer and e-mail)
T-CONT5	All types	Support all service types

OLT idle cell monitoring scheme, the ONT is not assumed to have the status-reporting capability. As specified in G.983.1, an ONT sends an idle cell in the upstream direction if it does not have any information to send. OLT then uses the reception of idle cells to indicate that the grant for that T-CONT could be assigned to other ONTs. Here, the SR DBA scheme is presented further.

In upstream queue-length reporting, each T-CONT length is encoded with a 1B coded value to indicate the range of the queue length. To facilitate an efficient upstream status-reporting mechanism, G.983.4 further introduces an upstream divided_slot, as illustrated by Figure 3.15. An upstream divided_slot (56B) can hold multiple minislots, where each minislot is assigned one ONT. Each minislot embeds multiple T-CONT-length reports between 3B overhead and 1B CRC fields. The OLT allocates an upstream divided_slot to a set of ONTs using a divided_slot grant in the downstream PLOAM cell. The set of ONTs is specified by the Divided_slot_grant_configuration message. In addition, OLT uses a Divided_slot_grant_configuration message to specify minislot_offset, which determines the location of an ONT minislot in the divided_slot. The OLT determines the minislot_offset based on knowledge of the number of ONT

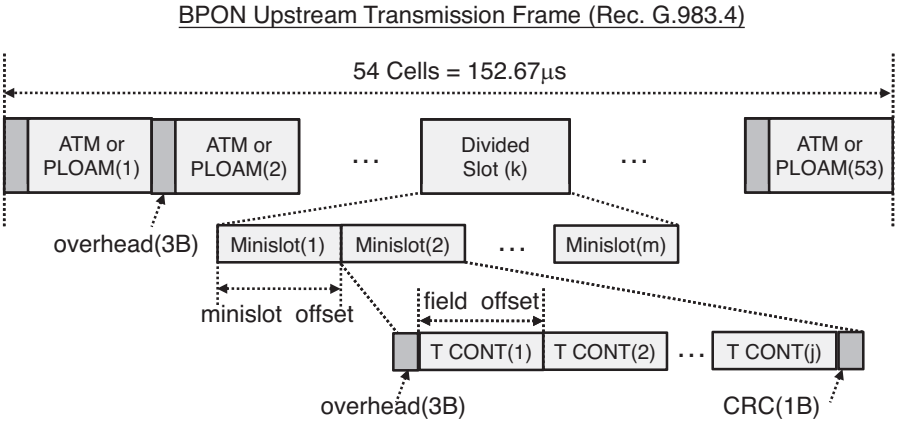


FIGURE 3.15 BPON T-CONT architecture.

T-CONTs, which is also created or deleted using a Divided_slot_grant_configuration message. A field_offset value is further specified if the ONT has more than one T-CONT.

3.3.5 Other ITU-T G.983.x Recommendations

Since the initial ratification of G.983.1 in 1998, the ITU-T body has developed a series of G.983 recommendations to enhance the original system. In this chapter, G.983.1, 983.3, and 983.4 standards are covered. In Section 3.3.1-3 we cover important aspects of G.983.1 concerning the PMD and TC layer of BPON. In Section 3.3.1 we present an enhanced wavelength range and associated filter requirement to support additional service signals, such as video, as specified by G.983.3. In Section 3.3.4 we describe G.983.4 protocol and requirements to incorporate DBA into BPON systems.

Recommendation G.983.2 Management and Control Interface Recommendation G.983.2 specifies ONT management and control interface specification (OMIC) for BPON. The G.982.2 standard specifies the protocol-independent management information base (MIB). The MIB describes the information exchange procedure between the OLT and the ONT. The OLT manages the ONT in a master–slave interface using an ATM connection called an ONT management and control channel (OMCC). An OMCC connection is established by setting up a unique VPI/VCI combination for each ONT using PLOAM messages. G.983.2 also specifies various OAM functions and parameters in the standard.

Protection Switching Recommendation G.983.5 specifies functions that extend the original G.983.1 standard to enable enhanced survivability and network protection mechanisms. In particular, the standard "describes BPON survivable architecture, protection performance criteria, protection switching, and protection switching protocols." It is noteworthy to point out that incorporation of a G.983.4 DBA function affects the protection switching protocol. As a result, part of the G.983.4 standard specifies an updated protection switching protocol to support DBA. More details on protection switching are presented separately in a later section.

3.4 GIGABIT-CAPABLE PON

ITU-T G.983 standards were developed based on ATM technology with the expectation that ATM would become the universal network protocol. Instead, however, Ethernet evolved successfully into that role. To support non-ATM-based SNI and UNI, BPON needs to implement expensive translation functions between ATM and the Ethernet interface. To adapt to the change in the telecommunication world, ITU-T created a G.984 series standard for gigabit-capable PON (GPON). GPON is not only an enhancement of BPON, but is specified with higher supported bit rates, better security with advanced encryption standard (AES), and other features. GPON further

TABLE 3.6 GPON and BPON Architecture Comparison

Characteristic	GPON	BPON
DS bit rates (Mb/s)	155.52, 622.08, 1244.16, 2488.32	155.52, 622.08, 1244.16
US bit rates (Mb/s)	155.52, 622.08, 1244.16, 2488.32	155.52, 622.08
Transmission frames	GTC frame with GEM	Fixed frame with ATM cells
Downstream security	AES	Key churning AES (later)
services	Full services (i.e., additionally include 10/100BASE-T Ethernet)	Telephony, SONET/SDH, TDM, ATM

supports Ethernet protocol and increases the bandwidth efficiency significantly by incorporating the GPON encapsulation method (GEM).

GPON Architecture GPON inherits its main characteristics from BPON in the PMD and TC layer. However, drastic changes are made in terms of a faster bit rate, use of GEM, and enhanced security. Table 3.6 summarizes the key differences between GPON and BPON architecture.

In GPON, the aggregate bit rate in either direction is improved to support up to 2.488 Gb/s in addition to the original 155.52 Mb/s, 622.08 Mb/s, and 1.244 Gb/s (downstream only) bit rate. Most of the GPON deployment selects the 2.488-Gb/s downstream and 1.244-Gb/s upstream bit-rate pair to support high-bit-rate access services and avoid OLT burst-mode receiver implementation issues. A GPON transmission convergence (GTC) layer, which specifies the GTC frame and GEM method, is described in full detail next.

3.4.1 GPON Physical Medium–Dependent Layer

The GPON PMD layer has most of the same requirements as that in a BPON. For example, both PON standards specify the use of class A, B, and C optics with the same logical reach and split ratio. Besides improving bit rates, a few notable additions are added to the G.984.2 GPON PMD standard to enhance GPON performance. These key additions are described in this subsection.

3.4.1.1 Upstream Burst Overhead and Timing

As explained before, successive upstream bursts must reserve enough guard time to allow laser on/off time, timing drift tolerance, level recovery, clock recovery, and the start of the burst delimitation. Whereas the G.983.1 BPON standard specifies a fixed 3B overhead to support these functions, the G.984.2 GPON standard specifies variable-length upstream physical layer overhead (PLO_u) and allows more time to adapt to upstream burst operations.

In summary, G.984.2 specifies 4B, 8B, 12B, 24B (PLO_u) based on 155.52-, 622.08-, 1244.16-, and 2488.32-Mb/s upstream bit rates, respectively. Each (PLO_u) field could be further divided into a subfield for guard time, preamble, and delimiter. The length and division of (PLO_u) is recommended based on a combination of

physical and implementation constraints. The guard time (T_g) is set to be larger than the expected laser on/off time (T_{on} and T_{off}), and the total timing uncertainty arises in propagation time due to fiber and component variations that could be caused by temperature drifts and other environmental effects. The delimiter time (T_d) is selected to allow the OLT to perform byte-level synchronization with a very low probability of severed error burst (P_{seb}) under severed upstream BER conditions.

The delimiter is constructed such that its shifted version over the preamble bit has a large Hamming distance d. It can be verified numerically that the $d = \text{int}(N/2) - 1$ for a delimiter of size between 8 and 20 bits (with a repeating "1010" pattern). The relationship between the number of bits in the delimiter N and the number of bits of error the receiver can tolerate, E, or the error-correcting capability, could then be approximated as $E = \text{int}(N/4) - 1$. Given a certain BER rate, P_{seb} could be computed by

$$P_{seb} = \frac{N}{E+1} \times \text{BER}^{E+1} = \frac{N}{\text{int}(N/4)} \times \text{BER}^{\text{int}(N/4)} \tag{3.2}$$

It can then be verified that N has to be at least larger than 16 bits to suppress P_{seb} to below 1.8×10^{-13} given BER at the 1×10^{-4} level.

After considering the minimum required (T_g) and (T_d), the rest of the time in (PLO_u) is dedicated to the preamble time (T_p) to complete burst-mode clock and data recovery (CDR) operation. Although GPON allows a longer T_p than that allowed by BPON to relax the CDR requirement, meeting the GPON CDR timing requirement is still a challenging task for circuit developers. As a result, the cost of GPON CDR is typically higher than for EPON CDR. Table 3.7 summarizes the recommended allocation of upstream burst overhead (PLO_u).

3.4.1.2 Laser Power Control In GPON, the OLT burst-mode receiver (BMRx) circuit requirement affects the use of a highly sensitive APD receiver at a 1244.16-Mb/s bit rate. The OLT receiver (APD + BMRx) must provide both high sensitivity and large dynamic range for burst-mode reception above a 1-Gb/s bit rate. A large dynamic range imposes a compromise to the multiplication factor M of the APD receiver, and a high incoming burst could potentially overload the OLT receiver.

To reduce the dynamic range experienced by the OLT receiver, the transmitting power of an ONT laser that located near the OLT could be regulated to avoid overload

TABLE 3.7 Allocation of GPON Upstream Overhead

Upstream Bit Rate (Mb/s)	(T_g) (bits)	(T_p) (bits)	(T_d) (bits)	Total Time (bits/ns)
155.52	6	10	16	32/205.8
622.08	16	28	20	64/102.9
1244.16	32	44	20	96/77.2
2488.32	64	108	20	192/77.2

of the OLT receiver. G.984.2 specifies three ONT output power modes. Each mode specifies a range for the mean launched power for coarse ONT power leveling control. The OLT receiver measures the incoming power level and compares it to power threshold(s) to determine if it needs to issue a power control message to the ONT. OLT power measurement is usually implemented by monitoring a small part of the receiver current I_{APD}. Four decibels of uncertainty is allowed to ensure a stable power-leveling mechanism.

In practice, the receiver monitoring circuitry is not very accurate, due to electronic effects. As a result, the OLT could combine the use of monitored BER values to trigger ONT power control. ONT power control could be either OLT- or ONT-activated, although in practice the former is used (if at all) because ONT implementations typically lack power monitoring tools. To control ONT power, OLT sends a downstream PLOAM Change_Power_Level (CPL) message and moves the ONT to one of the three power-level modes.

3.4.1.3 GPON Forward Error Correction G.983.2 further recommends optional use of forward error correction (FEC) coding to either lowering the cost of the transceiver component or further extending the system's physical reach. GPON investigates the use of FEC for two reasons. First, a higher bit rate reduces the receiver sensitivity. This is because a higher bit rate introduces more noise to the receiver because it has a wider bandwidth. As mentioned in Section 2.3.2, optical receiver sensitivity and bandwidth have a reciprocal relationship and thus fundamental improvement of receiver quality is required to improve the bandwidth without reducing the sensitivity. In addition, a high bit rate induces a larger chromatic dispersion. As mentioned in Section 2.6.2, the chromatic dispersion induced by a DFB type of laser is small even in a GPON system. However, the use of a multilongitudinal mode (MLM) laser such as an FP laser would induce more dispersion due to a wider laser linewidth. Moreover, an MLM laser would introduce mode partition noise (MPN) that causes a power penalty, to further degrade the power budget.

G.983.3 recommends the use of a cyclic code called Reed–Solomon (RS) coding to provide *optical coding gain*, defined as the difference in optical power at the receiver power, with and without BER, for BER $= 10^{-10}$. The RS FEC scheme transmits redundant information along with the original information. If some of the original data are lost or received in error, the redundant information is used to reconstruct the data. The design and error correction capability of the cyclic code is defined by $(n, n-m)$, where n is the size of the codeword (original bits + redundant bits) and m is the number of cyclic check bits (redundant information). G.983.3 defines an optional RS(255,239) FEC to encode the entire GTC frame (with the exception of the existing CRC in the GTC frame). RS(255,239) is first defined in the ITU-T G.709 optical transport network (OTN) standard and can provide 3 to 4 dB of optical coding gain in OTNs. The FEC, however, adds about another 6.2% overhead to the payload.

In current practice, GPON FEC option is rarely turned on because of the popularity of class B+ optics that can satisfy a 2.488/1.244-Gb/s transmission bit rate. The use of FEC in future 10-Gb/s PON systems, however, seems inevitable, due not only to

the implementation limitation of the transceiver but also to the physical constraint due to chromatic dispersion.

3.4.2 GPON Transmission Convergence Layer

The G.984.3 GPON TC (GTC) layer specification defines the GPON frame format, bandwidth allocation, ranging, OAM, and encryption method. In this subsection we focus on presenting details of the GTC frame format and GPON encapsulation method (GEM). The advanced encryption standard (AES) technique is also mentioned briefly, with more details presented in a later section.

3.4.2.1 GPON Transmission Convergence Downstream Frame GPON downstream transmission consists of fixed 125-μs GTC frames, as illustrated by Figure 3.16. Each GTC downstream frame consists of a downstream physical control block (PCB) and payload. The payload could carry either a fixed-size ATM cell (only in GPON equiment built on an earlier G.984-2004 version of the standard) or variable-size GEM frames. Details of the PCB are provided below.

Downstream Physical Control Block

- *Frame synchronization* (Psynch). The 4B Psynch uses a 0xB6AB31E0_16 pattern that allows the ONT to find the beginning of the frame. During initialization, an ONT could move from the Hunt state to the Synch state by identifying two consecutive Psynch fields. An ONT falls back to the Hunt state from the Synch State if it misses five consecutive Psynch fields. The Psynch field is also the only field in the GTC frame that is not scrambled. The rest of the GTC frame, including the rest of the PCB and all of the payload, is scrambled by a scrambling polynomial of $x^7 + x^6 + 1$.

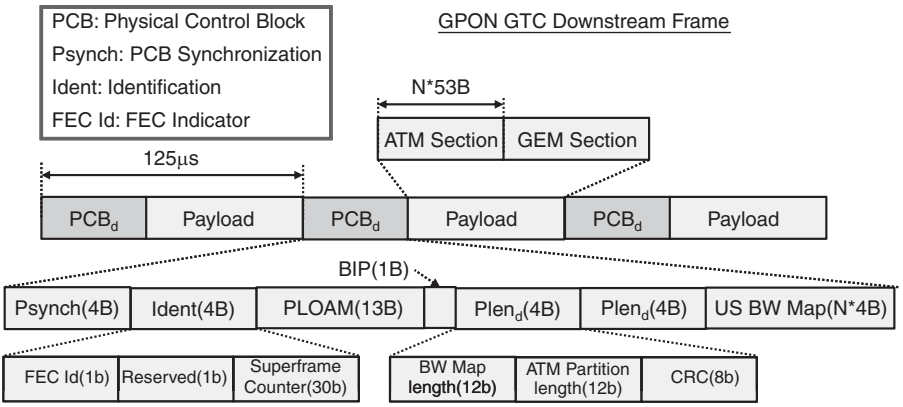


FIGURE 3.16 GPON GTC downstream frame format.

- *Identification* (Ident). The Ident field includes a 1b FEC option on/off indicator and 30b superframe counter. The superframe counter is used by the encryption system. The superfarme counter is incremented at the reception of every correctly received Ident field. Thus, it may also be used as a low-rate synchronization reference clock, similar to Psynch.
- *Downstream PLOAM field* (PLOAM_d). PLOAM_d has the same 13B field format and similar message format as the BPON PLOAM message. Some additional messages, such as threshold-crossing alerts and power-leveling control, are specified further.
- *BIP*. Similar to a BPON PLOAM BIP, this BIP can be used to detect errors that occurred in PCB and may also be used to estimate the link BER.
- *Downstream payload length indicator* (Plen_d). The 4B Plen_d gives the length of the upstream bandwidth allocation map in the PCB part (Blen) and the length of ATM section in the payload part (Alen). The Plen_d field is protected by a CRC-8b field and is sent twice to provide additional error robustness.

Upstream Physical Control Block

- *Upstream bandwidth allocation map* (US BW map). The US BW map is a variable-size field that allocates upstream transmission slots. Figure 3.17 illustrates the US BW map format. Each US BW map is 8B long and contains an allocation-ID (12b), flags (12b), a start slot time (2B), a stop slot time (2B), and a CRC (1B). The lowest 254 allocation-IDs are used to address the ONT directly, and the other 4 bits are used to be associated with a particular T-CONT or OMCC within that ONT. Note that allocation-IDs 254 and 255 are reserved for unknown (during ranging) and unassigned ONTs (no upstream allocation), respectively. The flags of the ONTs are used to indicate a certain control function that requests the ONT to send PLS_u, PLOAM_u, a dynamic bandwidth report (DBR_u), or to turn on upstream FEC encoding. In addition, the start and end time fields are used to locate the exact timing for that allocated upstream time slot. Figure 3.17 also shows an example of upstream bandwidth allocation with three ONTs upstream with a time slot allocated. Note that the start and stop time counter is reset to zero at the beginning of the GTC frame.

3.4.2.2 GPON Transmission Convergence Upstream Frame

A GTC upstream frame consists of multiple GTC bursts from ONTs based on an upstream transmission time slot allocated by the OLT. Figure 3.18 illustrates the frame structure of a GTC upstream frame. Each upstream frame matches the size of the downstream transmission frame to 125 μs. Each GTC burst has the following fields.

- *Upstream physical layer overhead* (PLO_u). The PLO_u first contains a 8B burst overhead (at 1244.16 Mb/s) to account for the preamble, and a delimiter of the burst overhead. Note that the guard time part of the burst overhead is accounted for by the OLT during upstream time slot allocation. As mentioned earlier, this part of PLO_u is required only between successive upstream bursts

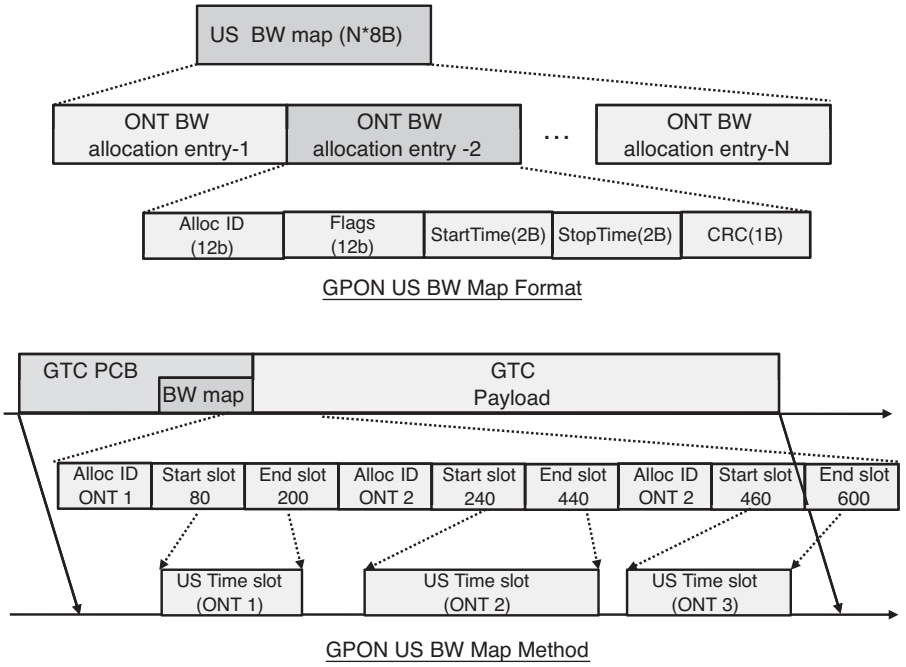


FIGURE 3.17 GPON GTC upstream bandwidth map format and allocation method.

sent by different ONTs. In addition to the burst overhead, the PLO_u includes another 3B of burst header. The burst header includes a BIP (1B), ONT-ID field (1B), and indication (Ind) field (1B). The BIP field records all bit error detected since the last BIP was transmitted. The Ind field reports the status related to PLOAM_u waiting, FEC status, and a remote defect indicator (RDI) status to indicate a high downstream BER rate. Burst-synchronous bit scrambling is applied to an entire upstream burst content, with the exception of the PLO_u.

- *Upstream PLOAM field* (PLOAM_u). The 13B PLOAM_u field is defined by G.983.1 and, similarly, to GPON PLOAM_d; a few new additional message types are added to enhance OAM functions and a power-leveling mechanism.

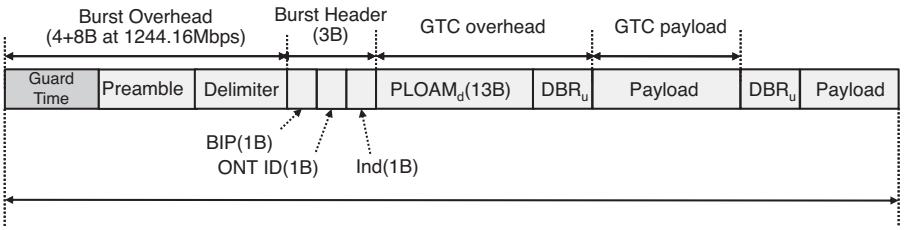


FIGURE 3.18 GPON GTC upstream frame format (G.984.3-2008).

- *Upstream dynamic bandwidth report* (DBR_u). Each DBR_u field reports the traffic status of the transmitting T-CONT (through the alloc_ID field). The length of the DBA_u reporting field could be coded into either 1B or 2B, depending on the DBA allocation format requested. OLT informs ONT of the type of DBA report to be used during the ONT activation process using the OMCI channel. Because the DBA_u field reports traffic status when an upstream T-CONT allocation is made to the ONT, it is also considered as a piggyback reporting mechanism. The DBR_u field is appended with a 1B CRC field to protect the DBR_u.

It is noteworthy to point out that the latest ratification of the G.984.3 standard (March 2008) made a number of significant changes in the GTC layer from the earlier version (February 2004 and its subsequent amendments). Most notably, the ATM partition of the GTC payload is altogether deprecated. As a result, the Alen within the Plen field is always set to 0, and no more ATM cells are transmitted in either the downstream or upstream GTC payload.

In terms of DBA, G.984.3-2004 relies on three mechanisms to report upstream queue lengths in the status-reporting (SR) DBA mode. In addition to the DBR type of piggyback reporting mechanism, G.984.3-2004 further specifies the status indication (using the idle Iden field of the PLO_u header) and entire upstream DBA payload-reporting mechanisms. Moreover, the DBR_u field originally also employs a third reporting mode, which is 4B in size, but it is also dropped in the new standard. In fact, the entire G.984.3-2008 standard builds the DBA service performance model based on the recommendation of using only a 1B DBA_u report field. Despite these changes in G.984.3-2008, however, specific implementation of the upstream scheduler as well as the traffic-monitoring (TM; known as nonstatus reporting in G.984.3-2004) DBA method is still open for vendor implementation.

Other notable changes in G.984.3-2008 include deprecation of the ONT power-leveling sequence (PLS) mechanism due to the lack of an ONT transmitter power-monitoring tool. Thus, the upstream PLS field is also dropped from the upstream frame. To accommodate these changes, G.984.3-2008 further specifies that a set of compliant rules be outlined further to ensure interoperability with GPON devices built on the earlier 2004 version.

3.4.2.3 GPON Encapsulation Method The GPON encapsulation method (GEM) allows the GPON protocol to support variable-size frames such as Ethernet. GEM also provides (1) multiplexing of GEM ports and (2) payload fragmentation and reassembly. In the G.984.3-2008, GEM frames further become the only traffic being carried by the GTC protocol. Figure 3.19 illustrates the structure of the GEM frame.

The following list summarizes the GEM header fields.

- *GEM payload length indicator* (PLI). The PLT is a 12-bit field used to indicate the length of the GEM payload section in bytes. As a result, GEM permits fragments up to a size of 4095 bytes.

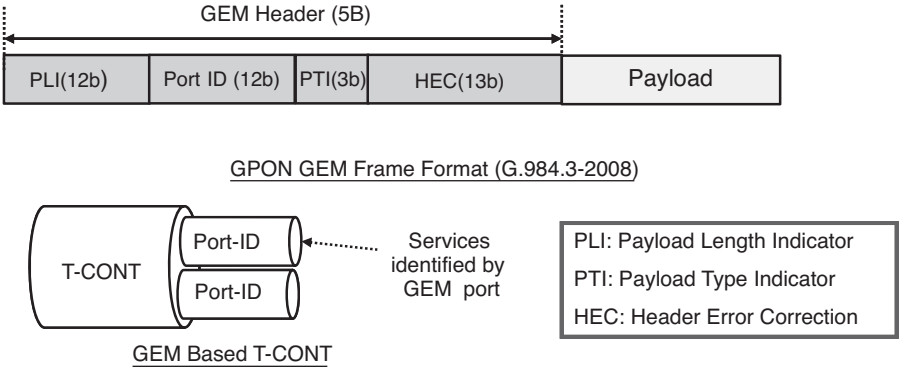
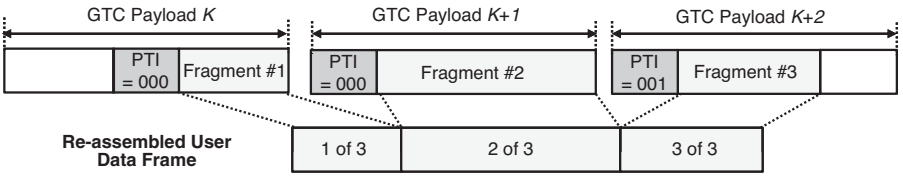


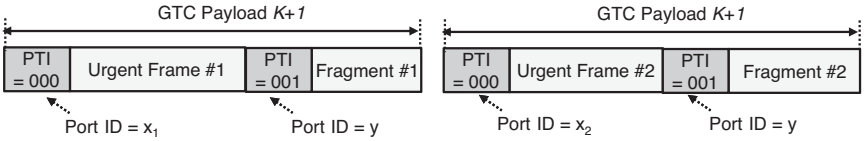
FIGURE 3.19 GPON encapsulation method (GEM) frame format.

- *GEM port identification* (Port-ID). The Port-ID is a unique traffic identifier on the PON to provide traffic multiplexing. Each Port-ID contains a user transport flow (port services), and an Alloc-ID/T-CONT could have one or more active port services.
- *GEM payload-type indicator* (PTI). The PTI field is used to indicate whether the current GEM frame is the end of a data fragment. The LSB is set to 1 if the GEM frame is either an unfragmented frame or the end of a data fragment. The second bit is used similarly to indicate if the current GEM frame is the end of a GEM OAM.
- *GEM header error correction* (HEC). GEM uses a combination of cyclic BCH(39,12,2) code and a single parity check bit.

ONTs receive downstream data based on a matched Port-ID for local service. In the upstream burst frame, each T-CONT subburst frame could contain data from multiple service ports, as illustrated by the GEM-based T-CONT layer structure in Figure 3.19. In addition to service multiplexing, the GEM frame also enables fragmentation. Figure 3.20 illustrates two cases of GEM fragmentation. In the first case, GEM is fragmented across three different GTC payloads. However, GEM fragmentation cannot straddle a multiple GTC frame boundary. As a result, the fragmentation must be aware of the number of bytes remaining in the frame and fragment the user data accordingly. In addition, if the remaining time in the GTC payload section is less than the minimum size of the GEM frame (5B), an idle (and incomplete) GEM header is sent to fill the void. In the second case, user data are fragmented to allow the insertion of a time-sensitive data frame. Each ONT is required to have at least two GEM reassembly buffers to support the use of time-urgent fragmentation. OLT should not interleave more than two users' data to the ONT unless it can determine that the ONT has additional reassembly capability. Similarly, OLT is required to have at least two GEM reassembly buffers for each Alloc-ID.



General GEM Fragmentation and Re-assembly Process



GPON Pre-emption Process Using GEM Fragmentation

FIGURE 3.20 GEM fragmentation and reassembly process. The figure shows two examples: a generic fragmentation process and a GPON preemption process using GEM fragmentation.

3.4.2.4 GPON Downstream Encryption Method Downstream data in GPON, as in any PON system, is broadcast to all ONTs via a shared feeder fiber. GPON standards strongly enhance the downstream data integrity over BPON standards, where downstream data are protected using only a key churning technique. This technique is, however, a low-cost hardware solution that has severe flaws. A BPON key churning technique employs a 16-bit key and applies XOR directly on the data with the key. Data protected by key churning can be broken as long as the text of the key is known. A 16b key, for example, could be derived in as few as 512 attempts. An attempt to strengthen the key using periodic updates does not improve the situation significantly.

To ensure better privacy and data integrity, GPON specifies a much stronger encryption process to secure downstream transmissions. Specifically, the OLT launches the key exchange process by sending a PLOAM message to the ONT. The ONT then generates a key and feeds it back to the OLT. The upstream communication is assumed to be secured, and thus the key is known only to honest ONTs and OLTs. After that, the OLT employs the advanced encryption standard (AES) mechanism to encipher the downstream data on a 16-byte block basis using that key. Among all ONTs, only the one that supplied the key can decipher the encrypted downstream data correctly.

3.4.3 Recent G.984 Series Standards, Revisions, and Amendments

Besides making notable changes to the TC layer in G.984.3-2008, the ITU-T body recently introduced a series of new standards, standard revisions, and amendments.

These include the G.984.5, and G.984.6, G.984.4-2008 revisions, and G.984.2 amendment 2. Some of the more significant changes are noted below.

3.4.3.1 G.984.2 Amendment 2: Transceiver Performance Monitoring

G.984.2 amendment 2 adds physical-layer transceiver performance monitoring to provide optical layer supervision (OLS) capability. The specification defines the performance requirement for OLT/ONT temperature and transceiver power monitoring. In the amendment, both OLTs and ONTs are expected to support $\pm 3^{\circ}$C. OLTs are expected to provide ± 2 dB and ONTs ± 3 dB of transceiver power measurement. The measured values are expected to serve as an alarm and be sent using an element management system (EMS) over OMCI channels. Details of the OMCI and the use of EMS are outlined in the G.984.4-2008 revision.

3.4.3.2 G.984.5: Enhancement Band

G.984.5 specifies the wavelength allocation and ONT/OLT filter requirement to reserve an enhancement wavelength band for future wavelength-division multiplexing (WDM) types of upgrades. The standard considers a coexistence situation where new data services could be overlaid in WDM wavelengths over existing GPON signals. The standard provides several wavelength options, depending on the existence of the legacy enhancement band signal (1550–1560 analog video service) and the cost consideration of the ONT filter types.

In the most ambitious scenario, the standard specifies an enhancement band between 1524 and 1625 nm. The standard requires that current ONTs implement a wavelength-blocking filter to avoid service interruption at the time of upgrade. OLT also requires a reciprocal triband filter to separate legacy signals from newly inserted wavelength signals at the time of upgrade. The particular WDM system is, however, outside the scope of this standard.

3.4.3.3 G.984.2 Amendment 2 and G.984.6: Single-Sided and Midspan Reach Extension

G.984.2 amendment 2 also indicates the use of class C+ optics to support 32 dB of power budget and up to 60 km of maximum fiber length. The amendment expects to turn on the FEC option in both the upstream and downstream directions to enhance receiver sensitivity. On the OLT receiver side, optical preamplification techniques such as SOA preamplification could be used to further enhance the upstream receiver (1290 to 1330 nm) sensitivity.

G.984.6 describes reference midspan reach extender architectures that would provide an excess of 27.5 dB of the power budget in both the optical trunk line (OTL) section and the ODN. The OTL is defined as the extended single feeder fiber that connects the OLT to the reach extender box. The reference architectures investigate the use of an all-optical amplifier option, an O/E/O conversion option, and a hybrid option. The particular type of optical amplifier and the specific architecture are outside the scope of this standard.

3.5 ETHERNET PON

As explained earlier, when the full service access network (FSAN) initiative was begun in 1995, the impetus of selecting ATM networking technology is that it would become the universal networking standard. FSAN submitted the BPON standard proposal in 1997, and BPON was formally ratified by ITU-T in 1998. However, Ethernet has quickly become the universal standard that ATM envisioned. Widespread use of Ethernet in both a local area network and a metro area network makes it an attractive alternative to support an access network. In January 2001, a new IEEE study group called Ethernet in the First Mile (EFM) was created to extend existing Ethernet technology into subscriber access areas. Ethernet technology over point-to-multipoint (P2MP) fiber, also known as *Ethernet PON* (EPON), quickly gained momentum for its ability to support full optical access with more relaxed timing requirements and the capability to encapsulate variable-size frames. IEEE EFM formally ratified IEEE standard 802.3ah in June 2004 to support the physical and data link layers of the EPON network. Since the introduction of EPON, it has quickly captured the interest of industry because EPON significantly simplifies the interoperability with Ethernet MAN and WAN equipment compared to the use of BPON equipment. Unlike ITU-T B/GPON standards, IEEE 802.3ah specifies only a small portion of an EPON system and thus also creates significant interests from the research community to address interesting challenges that are omitted from the standard.

3.5.1 EPON Architecture

The scope of IEEE 802.3ah, like the rest of the IEEE 802.3 standards, focuses exclusively on the physical and data link layers of the Open Systems Interconnection (OSI) reference model. IEEE 802.3ah specifies the EPON architecture in terms of its physical medium–dependent sublayer, P2MP protocol specification, and extensions for reconciliation, physical coding, and physical medium attachment sublayers.

3.5.1.1 EPON Physical Medium–Dependent Sublayer EPON PMD sublayer is defined by clause 60 of the IEEE 802.3ah standard. The PMD sublayer defines 1000BASE-PX10-U/D and 1000BASE-PX20-U/D for 10- and 20-km P2MP transceiver specifications. Table 3.8 summarizes the EPON PMD options. The operating wavelength range is the same as the one for BPON (i.e., 1260 to 1340 nm upstream and 1480 to 1500 nm downstream).

Details of the OLT and ONU transceiver specifications could be found in clause 60 of the IEEE 802.3ah standard. The difference between 1000BASE-PX10 and -PX20 specifications reside primarily in the OLT transceiver, where 1000BASE-PX20 uses better transmitter and APD receiver specification's than those of PIN-type receivers. The specifications of ONU transceivers were intentionally made very similar so that the same type of ONU could be made to support both -PX10 and -PX20 systems. This helps facilitate lower-cost EPON component development by encouraging economies of scale.

TABLE 3.8 EPON PMD Options

Parameter	1000BASE-PX10-U/D	1000BASE-PX20-U/D
Distance (km)	10	20
Data rate (Gb/s)	1	1
Minimum channel insertion loss (dB)	5.0/5.0	10.0/10.0
Maximum channel insertion loss (dB)	20.0/19.5	24.0/23.5
Maximum split ratio	16 (without FEC) 32 (with FEC)	16 (without FEC) 32 (with FEC)

One of the key distinctions between EPON and B/GPON is the relaxed timing requirement. BPON first specifies a stringent set of laser-on/off time (guard time) and preamble time (CDR). In EPON, the timing requirement is relaxed significantly to make the component as inexpensive as possible. As a result, the standard specifies the following parameters: laser-on/off times = 512 ns, CDR time $\leq$ 400 ns. This set of requirements is defined to facilitate high yield and to minimize the digital control interface. However, high physical burst overhead does translate into low efficiency. As a result, the CDR time is negotiable and the standard permits the use of better performing PMD components to enhance efficiency.

In a traditional Ethernet system, there is no global synchronization. All transmitters run on their local clocks to maintain low cost. A receiver derives the clock from the data received, and mismatches between clock sources are accounted for by adjusting the interframe gap (IFG) between Ethernet frames. In an EPON system, however, upstream ONU bursts are multiplexed using a TDMA method, and precise timing across all ONUs must be maintained. To maintain a common reference clock with an OLT, EPON ONUs are mandated to acquire loop timing from the OLT. This means that the ONU clock (MAC layer MPCP clock) should track the OLT MPCP clock from the received downstream signal. In this way, the OLT could still operate an inexpensive clock crystal with $\pm$100 parts per million from the nominal frequency. ONUs are able to remain synchronized at all times because OLT should constantly transmit either data or idle frames.

3.5.1.2 Extensions to Other Sublayers

Because of its P2MP architecture, EPON requires a certain extension to conform to existing IEEE 802.3 clauses.

Reconciliation Sublayer (RS) In Ethernet, layer 2 connection is achieved using an IEEE802.1D-based bridge. An Ethernet switch forwards an incoming packet from the input port to one or more of its output port(s). Because a traditional Ethernet system has P2P Ethernet connectivity, packets are only directed forward without being sent back to the same input port. In effect, a layer 2 Ethernet switch examines the SA and DA of each frame received and drops those from the same domain.

Due to the directionality of the passive splitter, ONUs cannot see each other's upstream traffic. Thus, OLT is required to help forward any inter-ONU communications. Without a modification to the standard, an IEEE 802.1D switch connecting to the OLT would simply discarded any such inter-ONU transmission frames because they would appear to come from the same domain. To resolve this issue, subclause 65.1 of IEEE 802.3ah defines a point-to-point-emulation (P2PE) function in the RS sublayer. The P2MP function tags each Ethernet frame with a unique logical link ID (LLID) in the preamble for each upstream ONU frame. The MSB of the LLID field indicates the mode bit, where "0" indicates P2PE operation and "1" indicates single-copy broadcast (SCB) that broadcast the traffic back to all ONUs.

Physical Coding Sublayer (PCS) PCS defines the 8b/10b line coding and an optional FEC. EPON adapts the 8b/10b line coding used by the IEEE 802.3z gigabit Ethernet standard. An 8b/10b line code produces a sufficient number of zero–one and one–zero transitions and dc balanced output to ensure easy clock recovery. However, this increases the baud transmission rate to 1.25 Gbaud/s. After 8b/10b line coding, the PCS further defines a data detection function to control the laser-on/off function. In short, the data detection function could be seen as a delay line that turns the laser on or off at the appropriate time after detecting waiting or completion of data from MAC. In EPON a 10-bit 8b/10b encoded data signal is called a *code group*. EPON PCS also defines the same RS(255,239) cyclic code as in GPON FEC. The use of FEC in EPON is also optional.

Physical Medium Attachment Sublayer (PMA) The PMA sublayer specifies the time interval required by the OLT to acquire bit-level synchronization using the CDR process and autocorrelation process. As mentioned earlier, T_CDR is required to be within 400 ns, and the PMA specifies an additional 32 ns for code group alignment.

3.5.1.3 EPON Framing EPON transmits variable-size Ethernet frames such as the one shown in Figure 3.3. Both a generic Ethernet frame and an EPON frame contain the destination and source MAC address (DA and SA, 6B each), payload length/type (PLT 2B), a variable-size payload section, and CRC (4B). When PLT is above 1500 (the maximum payload size), it is used to indicate a specific type of Ethernet frame. The actual payload length ranges from 46 to 1500 bytes.

The main difference is in the EPON header. The EPON header modifies the preamble and start of the frame delimiter (SFD) part of the Ethernet MAC frame to include LLID so as to enable the P2PE function. The SFD byte is moved to the third byte of the preamble, which is renamed to the start of the LLID delimiter (SLD). EPON data frames could further incorporate an optional VLAN control, as illustrated in Figure 3.21. The optional 4B VLAN tag control is added to identify a virtual network associating a set of communication entities. The tag control field carries a 3b priority value that selects one of the eight priority queues, which for upstream transmissions will be the OLT.

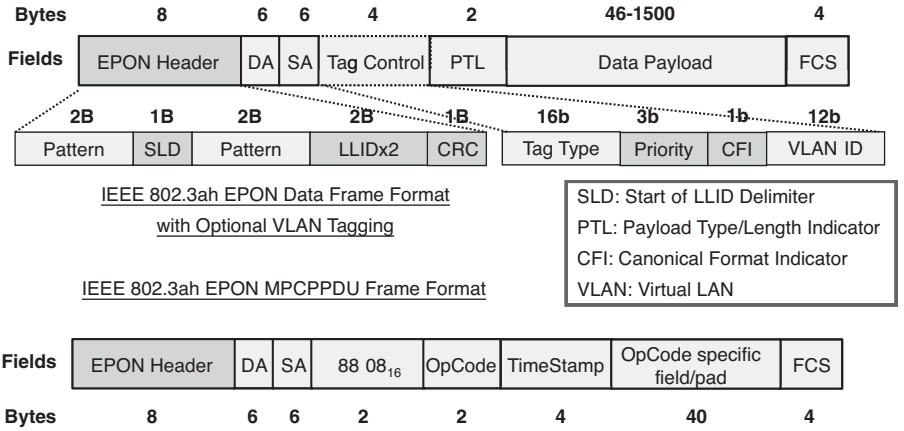


FIGURE 3.21 EPON Data frame format with optional VLAN tags and GEM MPCP frame format.

3.5.2 EPON Point-to-Multipoint MAC Control

EPON point-to-multipoint MAC control (MPMC) is defined by clause 64 of the IEEE 802.3ah standard. It defines primarily the multipoint control protocol (MPCP) that performs ranging, bandwidth arbitration, and discovery functions.

3.5.2.1 Multiple Point Control Protocol The MPCP controls rely on the use of the MPCP data unit (MPCPDU). As illustrated in Figure 3.21, the MPCPDU has a fixed 64B frame size (without counting the 8B EPON overhead). Six types of MPCP control frame are defined in the standard. The MPCP control frame is identified by 88-08$_{16}$ in the PTL, and the type of MPCP control frame is identified by the 2B opcode. The corresponding opcode and the type of MPCP message are listed below.

- 0001$_{16}$: PAUSE. The PAUSE message was used originally for flow control purpose. It was developed for a P2P Ethernet network for an overloaded receiver to stop receiving from its peer.
- 0002$_{16}$: GATE. There are two types of GATE messages: discovery GATE and normal GATE. Discovery GATE is used to advertise a discovery slot for all uninitialized ONU. Normal GATE is used to grant an upstream transmission opportunity to a single ONU.
- 0003$_{16}$: REPORT. ONU uses a REPORT message to report its local queue status to an OLT. An ONU can report the status of a variable number of queues using an REPORT message.
- 0004$_{16}$: REGISTER_REQ. An initialized ONU uses a REGISTER_REQ message to respond to a discovery GATE message.
- 0005$_{16}$: REGISTER. The OLT sends a REGISTER message to a newly discovered ONU and assigns a unique LLID.

- 0006_{16}: REGISTER_ACK. REGISTER_ACK is the final registration acknowledgment sent by an ONU in the discovery process.

The format of these control frames is presented in more detail during our discussion of ranging, autodiscovery, and DBA processes. Other details of the MPCP control frames may be found in clause 64 of the IEEE 802.3ah standard.

Ranging Process In EPON, all ONUs are synchronized to the OLT clock based on a loop-timing mechanism. Thus, an ranging process, which measures the RTT between an OLT and an ONU, can be carried out using normal GATE and REPORT control messages. Figure 3.22 illustrates how RTT can be determined by the GATE and REPORT processes. The OLT sends a normal GATE message with the time stamp T_0. The ONU would respond to the GATE message with a REPORT message after some delays, T_R. The time stamp on the REPORT message T_1 represents exactly $T_0 + T_R$ because the ONU has precise loop timing from the OLT. Upon receiving the REPORT message at T_2, OLT could then use the information $T_2 - T_1$ to determine the RTT.

Autodiscovery Process Similar to the GPON discovery process, an EPON autodiscovery process allows an ONU to register and join or rejoin the system. The term *auto-* is used because an EPON OLT broadcasts periodically to open up a discovery window for all uninitialized ONUs to respond. Figure 3.23 shows the autodiscovery process. The corresponding formats for the MPCP control messages used in autodiscovery are shown in Figure 3.24.

The autodiscovery process consists of the following steps.

- An OLT periodically opens a discovery window (discovery GATE). The discovery GATE message advertises a discovery time slot to every uninitialized ONU.

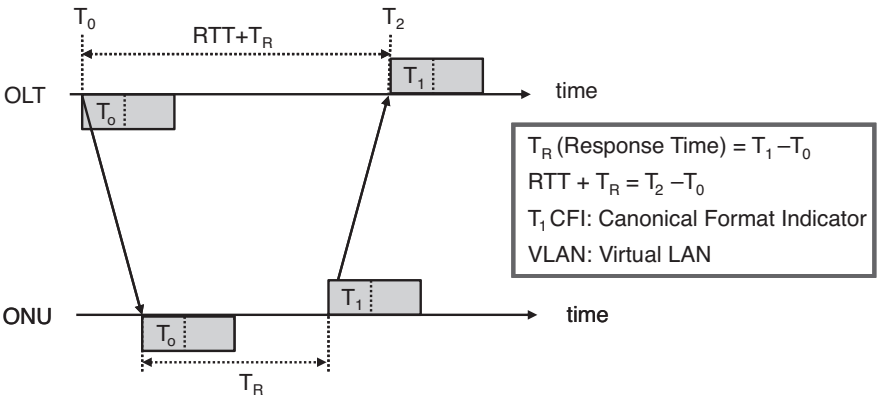


FIGURE 3.22 EPON ranging process.

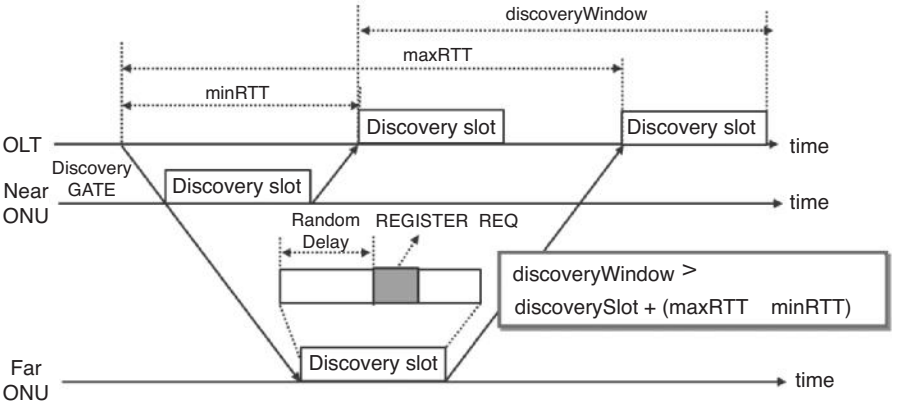


FIGURE 3.23 EPON autodiscovery process.

The length of the discovery slot is selected to be longer than the size of the ONU REGISTER_REQ. The minimum size of the discovery window is related to the discovery time slot, as shown in Figure 3.23. The reason that the discovery slot and discovery window should be larger than the minimum size requirement is to avoid upstream collisions during the autodiscovery process. In EPON, the minimum RTT time is assumed to be zero. Upstream collision is possible during autodiscovery because there could be more than one uninitialized ONU in the system, and their discovery slots could be overlapped.

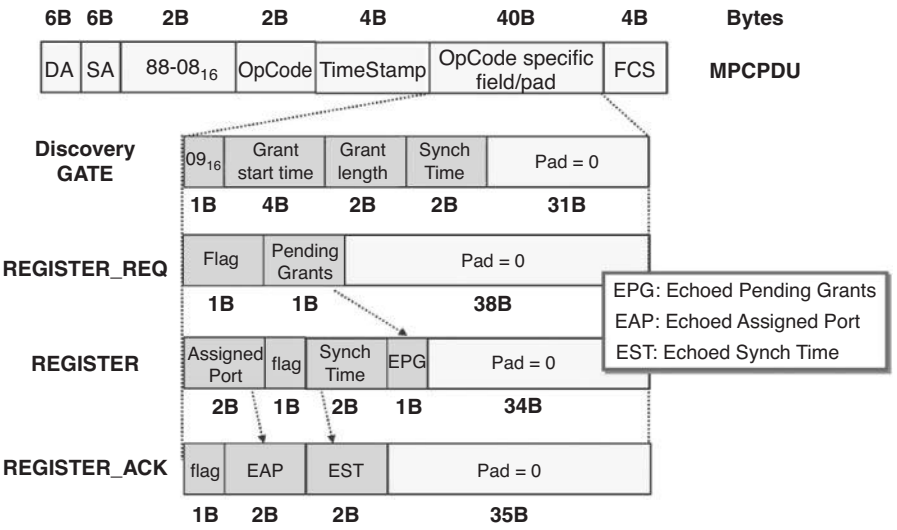


FIGURE 3.24 EPON discovery GATE, REGISTER_REQ, REGISTER, and REGISTER_ACK message formats.

- Uninitialized ONU responds to a discovery GATE message (REGISTER_REQ). An ONU sends the REGISTER_REQ message during the discovery time slot after a random delay. The random delay is added to avoid persistent collision from more than one uninitialized ONU that share an overlapped discovery slot.
- OLT registers ONU (REGISTER). Upon receiving a valid REGISTER_REQ message from an uninitialized ONU, the OLT assigns a unique LLID to the single ONU. Because the message is sent without establishing a logical link between the OLT and the ONU (i.e., the ONU has not learned its LLID), the message has a broadcast LLID field but uses the ONU MAC address in the DA. For all other MPCPDU control messages, the DA MAC address is set to a default $0180C2000001_{16}$ multicast address.
- ONU acknowledges (REGISTER_ACK). An ONU replies to the OLT with a simple REIGSTER_ACK message to indicate either Nack or Ack.

Figure 3.24 and Table 3.9 show further details of the autodiscovery control message formats.

- Discovery GATE. The discovery GATE message includes only a single grant and, additionally, advertises a synch field to an ONU. Because of different burst-mode receiver implementation at the OLT, the AGC and CDR intervals may be different. The synch field communicates the OLT's burst-mode synchronization time to the ONU. Since both T_{AGC} and T_{CDR} cannot be longer than 400 ns in EPON, the maximum allowed synchronization time is 832 ns (including an additional 32 ns for code group alignment), or 52 time quanta (EPON defines a 16-ns TQ, and the MPCP clock counter advances based on a TQ interval). Additionally, the discovery GATE message specifies the MPCP clock start time and discovery time slot length.
- REGISTER_REQ. The REGISTER_REQ message could indicate either a registration (flag = 1) or deregistration (flag = 0) request. When ONU sends a REGISTER_REQ message to the OLT, it also sends a pending grant field to indicate its ability to store future grants. This is because an ONU is required to store grant parameters until its local clock reaches the grant start time value.
- REGISTER. The REGISTER message assigns LLID to an ONU using the assigned port field. It also uses the ONU MAC address in the DA since the

TABLE 3.9 Contents of DA and OpCode for Autodiscovery Messages

MPCPDU Type	Destination Address	OpCode
Discovery gate	$0180C2000001_{16}$	0002_{16}
REGISTER_REQ	$0180C2000001_{16}$	0004_{16}
REGISTER	ONU MAC address	0005_{16}
REGISTER_ACK	$0180C2000001_{16}$	0006_{16}

logical connection has not yet been established. The flag field is used to indicate the registration type to ONU; these types could register, deregister, nack, and ack. IEEE 802.3ah does not specify that the synch time in the REGISTER be the same as that in the discovery GATE message. Thus, it is possible that a vendor could use information learned about the ONU in the REGISTER_REQ message to reduce the synchronization time. Finally, the OLT echoed the pending grants reported by ONU to indicate that it is aware of the ONU's buffer limitation.

- REGISTER_ACK. The REGISTER_ACK message completes the autodiscovery process and includes a single flag to indicate ack or nack. In addition, it echoed the assigned port and synchronization time values received.

As mentioned earlier, a random delay is added by the ONU after the beginning of each discovery time slot to avoid persistent collision with another ONU's upstream control message. In the case of a collision, the OLT would not receive a valid REGISTER_REQ or REGISTER_ACK message from the ONU. The OLT would periodically open a discovery window to allow these ONUs to join the network and resolve contention in the next discovery window. In these cases, the size of the discovery time slot and the periodicity of the discovery window would affect the performance of collision probability. The IEEE 802.3ah standard does not specify the details of these parameters and leaves implementation to the vendor.

Bandwidth Grant and Report Process One of the main functions of EPON MPCP is to enable bandwidth arbitration. It is important to note that IEEE 802.3ah does not specify the implementation details of DBA, and the standard leaves the implementation of DBA open to vendor innovation. Instead, the standard enables a bandwidth grant and report mechanism using normal GATE and REPORT control messages.

The following list provides details of the normal GATE and REPORT control message formats, which are illustrated in Figure 3.25.

- *Normal GATE message*. A normal GATE message includes a mandatory 1-byte field to indicate the number of grants (3b), the type of GATE message (1b), and the force upstream queue report bitmap (4b). In a normal GATE, each grant requires a 6-byte field to indicate the start of grant time (4B) and the length of grant time (2B). The number of grants of a normal GATE message is between 0 and 4, and as a result of zero padding could be 39, 27, 21, or 15 bytes long. A normal GATE is indicated by a "0" in the type of the GATE field, where the discovery GATE sets the bit to "1." The force upstream queue report bitmap mandates the ONU to report the upstream queue status in the corresponding OLT for possible upstream grant time slots. When an OLT mandates the ONU to report upstream queue status, it would possibly grant a larger upstream time slot, to include room for the upstream REPORT message in addition to data messages requested.

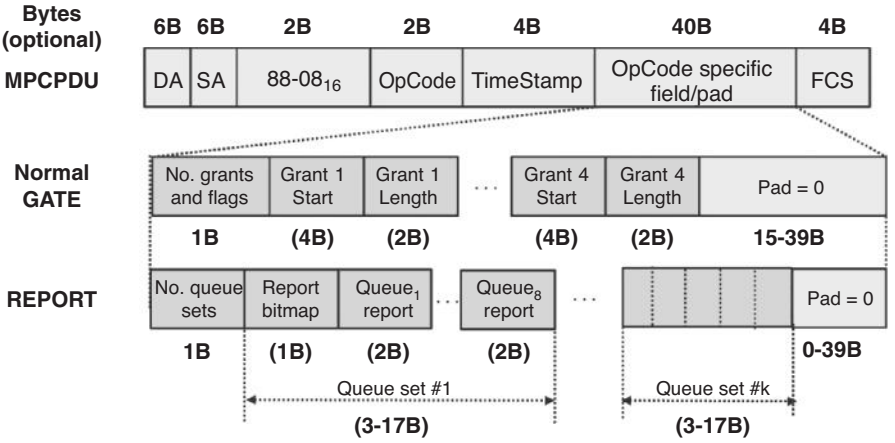


FIGURE 3.25 EPON normal GATE and REPORT message formats.

- *REPORT message*. Each REPORT message could include a variable number of queue sets. For each queue set, the ONU uses a 1-byte report bitmap to indicate the subset of eight ONU upstream queues that would be reported. The actual queue length is reported according to the number of TQ (16b)'s in a 2B field. The REPORT message could include any of 0 to 13 queue sets, depending on the number of queues reported in each queue set. For example, a REPORT message could contain two queue sets for reporting full eight queues (34B) or 13 queue sets for reporting only one queue (39B).

The reason that the REPORT message is specified to include a multiple queue set in EPON is to avoid wasting bandwidth. In particular, the OLT DBA function might grant only a portion of the bandwidth requested. Since Ethernet frames cannot be fragmented, unfit frames have to be deferred to the next time slot, resulting in significant unused bandwidth. Reporting multiple queue sets allows the ONU to report upstream queue length according to some predefined threshold. If a frame does not fit into the threshold, the ONU would choose a slot size that is smaller than the threshold and preserve the upstream bandwidth for other ONUs. The standard does not provide details on how the threshold should be determined, but it does dictate that each queue set specify the cumulative length (including overhead, i.e., 8B preamble and 16B IFG) from the head of the FIFO queue.

3.5.3 Open Implementations in EPON

The IEEE 802.3ah standard intentionally leaves many details outside the specifications except for the basic physical and data link layers. This is done to preserve flexibility in terms of implementations and to encourage vendor innovations. The objective of the former is to maintain a low-cost device and a fast time to market. For example, the physical synchronization time has a relaxed requirement to allow

a low-cost burst-mode receiver, but the standard leaves room to incorporate a better device to enhance system efficiency. Some of the details omitted include interesting problems that stimulate academic research studies in addition to vendor implementations. For example, dynamic bandwidth allocation and the provision of guaranteed service levels are two very interesting problems that have received much attention from the research community. However, EPON open implementations have also led to confusion in the market place and a lack of interoperability. For example, Japanese EPON will not work with a Chinese EPON.

The following short list summarizes some of the most interesting topics left open by the IEEE 802.3ah standard: dynamic bandwidth allocation (DBA), SLA guaranteed mechanisms, discovery slot and window assignment, protection switching architecture, downstream security and encryption, and OLT burst-mode receiver limitation. In this section we address some of the key developments in DBA and data encryption.

3.5.3.1 Dynamic Bandwidth Allocation Although details of DBA and scheduling are not specified by the standard, vendors and researchers have developed many interesting schemes to facilitate high bandwidth utilization and robust SLA provisioning. The best known DBA algorithm is called *interleaved polling and adaptive cycle time* (IPACT) DBA. IPACT utilizes a just-in-time scheduling strategy to dynamically allocate upstream and to achieve high bandwidth utilization. Figure 3.26 is a working example of how an IPACT scheme would function.

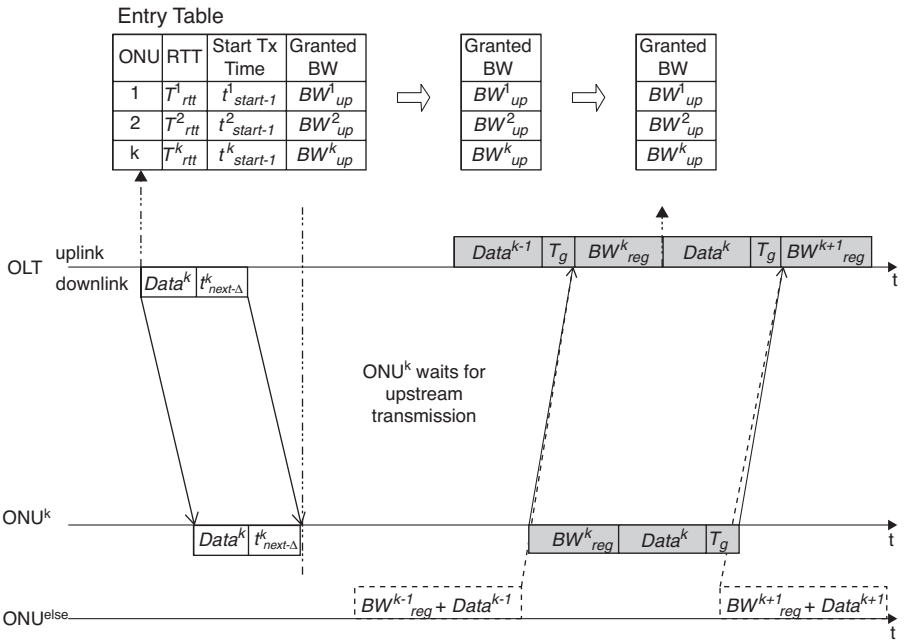


FIGURE 3.26 EPON DBA process.

IPACT DBA The objective of the IPACT scheme is to maximize bandwidth utilization by pipelining ONU requests. In IPACT, the OLT keeps active records of the bandwidth requests sent through REPORT messages from an ONU. The bandwidth requested is granted to the ONU using a GATE message in the next traffic cycle. Successive GATE messages are timed according to their respective RTT such that upstream collision is avoided. The ONU responds to the GATE message immediately (i.e., just-in-time). The IPACT scheme could schedule upstream traffic bursts in a pipelined fashion such that successive bursts are separated by the minimum specified guard time. In this way, upstream bandwidth utilization is maximized.

In the grant allocation scheme above, called *gated service*, the OLT grants exactly how much an ONU has requested in the preceding traffic cycle. Although the scheme provides robust bandwidth utilization, it does not perform well in terms of bandwidth fairness. A large request from one ONU could potentially block upstream transmission from the other ONUs, creating a large latency in other services.

IPACT could be enhanced significantly in terms of QoS and fairness by limiting the size of the grants to a fixed value. In the limited service scheme, IPACT is able to provide both robust bandwidth utilization and fairness across ONUs. The article by Kramer et al. [2] is a good reference on further details of the IPACT scheme.

System Implementation Issues During development of the EPON standard, a just-in-time DBA scheme such as IPACT has been proposed, but the eventual ratification of the standard does not adopt the just-in-time scheme. This is because Ethernet does not have frame fragmentation ability. Thus, a GATE message itself cannot be used to indicate the timing of upstream transmission because a GATE message could be blocked by a long Ethernet frame in transmission. As a result, the GATE message is defined to carry the start time in addition to just simply sending the length of the upstream slot granted.

Moreover, EPON operates in a multiservice environment and the operators need to honor their SLA agreements. An online and purely reactive type of DBA scheme such as IPACT does not always provide the performance operators needed to honor their SLA agreements. As a result, many other DBA schemes have been investigated since the introduction of IPACT to investigate off-line scheduling, hybrid online/off-line scheduling, predictor-based scheduling, and so on [3]. The objective of these DBAs is to provide class-of-service-oriented bandwidth allocation with reasonable complexity and hardware delays. Due to its pioneering nature in the field, however, IPACT is still used extensively as a performance benchmark for these new DBA schemes.

3.5.3.2 Downstream Encryption Mechanisms IEEE 802.3ah does not specify security, in particular encryption, mechanisms for EPON. As mentioned earlier, implementing encryption is important in protecting the data integrity in any P2MP PON system. As a result, many systems do implement vendor-specific encryption mechanism in the EPON chip set. The lack of standard specifications encouraged vendor innovations, but it also created a compliant and interoperability issue.

Today, most EPON chip set manufacturers adopt the AES-128 standards, similar to the standards used by GPON to encrypt downstream EPON frames. In some cases, a stronger public key infrastructure (PKI) cryptography technique is used to protect the integrity of the encryption key during protocol exchanges.

3.5.4 Unresolved Security Weaknesses

Although the AES-128 encryption technique can effectively protect against downstream eavesdropping, other known types of security weaknesses still reside in the EPON system. The article by Hajduczenia et al. [4] provides a comprehensive review of the development and existing questions concerning EPON security issues. It is important to note that these problems are generic to any P2MP-based PON system, and the solution to these security weaknesses is still an active area of research.

3.5.4.1 Upstream Denial of Service Attack A denial of service (DoS) attack could occur in PON when a malfunctioning (babbling/promiscuous) or malicious user sends an in-band message that could disrupt valid ONU services. Regardless of the source (whether it is malicious or innocent due to hardware defects), the effect of upstream disruption could be costly. In EPON, for example, operation of a keep-alive mechanism would cause the system to restart, thus bringing down an entire system with meticulously crafted security mechanisms.

To address the problem, the PON system must be able to detect the source of attack and counter the security threat. However, a PON lacks the mechanisms to either detect or counter upstream DoS attacks. If the upstream disruption blocks upstream transmission from other ONU users, it is inherently difficult for the OLT to determine the source of attack using a higher-layer detection method. If a passive power-monitoring technique is available (such as the OLT transceiver power-monitoring tool specified by G.984.2 amendment 2), the OLT could potentially determine the location of the attack based on the strength of the receiving power. Knowing the optical receiver power during the time of attack does not guarantee knowledge of the location of attack, however. Even if OLT could detect the source of attack automatically, a PON does not have any active switching comparability that can block the malicious user from disrupting the network further.

In current implementation, operators assume the occurrence of upstream disruption to be rare and use manual monitoring (optical time-domain reflectometry and power measurement) and an intervention method (physical shutdown) to resolve DoS attacks.

3.5.4.2 Masquerade or Replay Attack If a malicious ONU user can listen and decrypt traffic of another ONU user, the malicious ONU can monitor information of another targeted ONU, including LLID, MAC address, and so on. The malicious user can then easily steal the identity of the victimized ONU to disguise and masquerade as the target ONU and steal its identity. Masquerade attack is very difficult to detect once it is under way since the malicious user is seen as the valid user. Secure authentication and/or upstream encryption could potentially solve the problem.

However, it should be noted that malicious ONU users, even if they fail to decipher the collected information, can still relay the signals collected and replay them in the network to cause a disturbance with unpredictable consequences.

3.6 IEEE 802.AV-2009 10GEPON STANDARD

The IEEE EPON standard shows many distinct benefits over the BPON standard in terms of greater bandwidth efficiency and more relaxed timing requirements. However, the GPON standard addresses the bandwidth efficiency and Ethernet protocol adaptation issue successfully by introducing GEM. It also provides a much higher downstream bit rate than the one that EPON provides. On the other hand, the performances of burst-mode optical components have improved, and some of the original EPON relaxed timing requirements have become the critical bandwidth-limiting bottleneck.

In response to a demand for higher-bit-rate and burst-mode hardware equipment, the IEEE 802.3 working group in March 2006 held a call for interest (CFI) that aimed to leapfrog the 2.5-Gb/s option and opted directly for the 10-Gb/s EPON option. The IEEE P802.3av task force [5] was then formed in September 2006. In October 2009, the IEEE 802.3av standard [6] was officially ratified as an amendment to the 802.3av standard.

3.6.1 10GEPON PMD Architecture

The 802.3av standard confines the changes to the 10GEPON system primarily within the PMD sublayer. The intention is to leave the existing OAM, DBA, and network management system (NMS) unchanged, such that 10GEPON users could expect backward compatibility toward 1GEPON equipment. The 802.3av standard adds two new clauses to the IEEE 802.3 documentation that addresses PMD and extension to RS and PCS/PMA sublayers. In this subsection we note the important amendments in these two new clauses.

3.6.1.1 Bit Rates and Power Budget The 10GEPON standard specifies several power budget options. PRX options define the power budget for an asymmetrical 10.3125-Gb/s downstream and a 1.25-Gb/s upstream bandwidth combination. PR options define the power budget for symmetrical 10.3125-Gb/s transmission. Each option comes with three additional classes. Table 3.10 summarizes the 10GEPON power budget options.

It is important to note that the PR10 and PR20 classes of the power budget for PRX and PR systems are the same as those for the PX10 and PX20 systems. As explained in Section 2.3.2, the optical receiver sensitivity would suffer additional losses when receiving a higher bandwidth. To compensate for the loss budget, the 10GEPON standard recommends the use of better OLT transceiver components as well as use of the FEC to offset the losses expected.

TABLE 3.10 Bit Rate and Power Budget Options of IEEE 802.3av PMD

Rate/ Budget Type	DS Line Rate (Gbaud/s)	US Line Rate (Gbaud/s)	Power Budget (dB)	OLT Tech. (DS/US)	ONU Tech. (DS/US)
PRX10	10.3125	1.25	20	EML[a]/APD[b]	PIN/DML[b,c]
PRX20	10.3125	1.25	24	EML + SOA/APD[b]	PIN/DML[b]
PRX30	10.3125	1.25	29	EML/APD[b]	APD/Hi DML[b]
PR10	10.3125	10.3125	20	EML/APD	PIN/DML
PR20	10.3125	10.3125	24	EML + SOA/APD	PIN/DML
PR30	10.3125	10.3125	29	EML/APD	APD/Hi DML

[a] EML, electroabsorption modulated laser.
[b] The FEC option is turned off at a 1.25-Gbaud/s transmission rate.
[c] DML, direct modulated laser.

Another important note is that 10GEPON applies 64b/66b line coding to the 10.3125-Gbaud/s line rate. In this case, the actual data rate is 10 Gb/s. The choice of 64b/66b has many implications in terms of the burst overhead and FEC alignment. Some of these details are provided in a later section.

3.6.1.2 FEC and Optical Coding Gain EPON standard initially considers an RS(255,239) FEC option but eventually opts for RS(255,223) for a higher optical coding gain but at twofold higher overhead. Table 3.11 compares the difference between these two types of coding techniques.

Figure 3.27 shows about 6.4 dB of optical coding gain from a 10GEPON prototype system. In the figure, optical coding gain is defined further as the improvement in receiver sensitivity given the same 10^{-12} BER level. It is important to note that the optical coding gain is smaller than the electrical coding gain. This is because the optical gain is affected further by the noise characteristics of the optical receiver devices. In the example illustrated, an APD type of receiver is used, and thermal noise and shot noise are mixed in an APD receiver. The presence of thermal noise reduces the perceived optical gain. In this case, the results show that optical coding gain is approximately 70 to 90% of the electrical coding gain. In the case where a PIN receiver is used and thermal noise is the dominant factor in the receiver, the optical coding gain would be worse and only about half of the electrical coding gain.

TABLE 3.11 Comparison of RS(255,239) and RS(255,223) at a 10.3125-Gbaud/s line rate

FEC Code Type	Overhead (%)	US Overhead (%)	Coding Gain[a] (dB) at 10^{-12}	Input BER	Burst Tolerance (bits)
RS(255,239)	6.3	7.34	5.9	1.8×10^{-4}	57
RS(255,223)	12.5	13.53	7.2	1.1×10^{-3}	121

[a] Electrical coding gain.

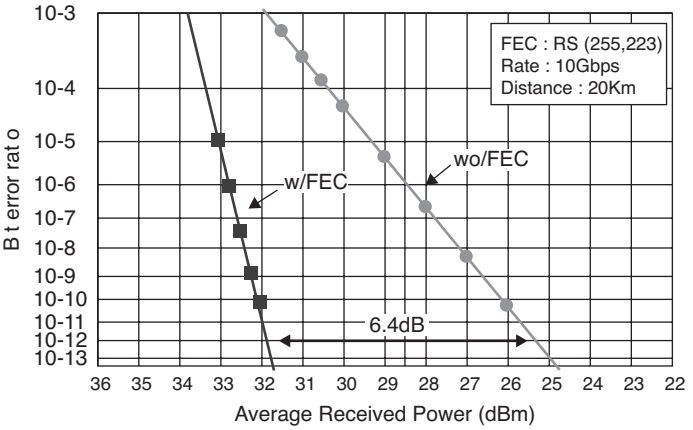


FIGURE 3.27 10GEPON RS(255,223) optical coding gain.

3.6.1.3 Wavelength Plan and Coexistence The 10GEPON standard specifies the upstream transmission wavelength in the range 1260 to 1280 nm and the downstream transmission wavelength in the range 1575 to 1580 nm. The idea behind the wavelength selection is that the future 10GEPON system is expected to coexist with the 1GEPON system using TDMA in the upstream and WDM in the downstream. Details of coexistence architecture are presented later.

As described in Section 3.1.2, FP lasers are no longer applicable in 10G systems, due to their high chromatic dispersion power penalty. As a result, the original 100-nm upstream window is no longer necessary. Thus, a narrower upstream wavelength band is preferred. Since a 20-nm transmission window is sufficient even for uncooled DFB lasers, the 10GEPON standard defines the lowest block of a 20-nm transmission window (i.e., 1260 to 1280 nm) for 10G-ONUs. The remainder of the 80-nm transmission window is reserved for future use.

In the downstream direction, the downstream wavelength is selected both to avoid collisions with legacy downstream signals and to minimize additional optical transceiver components in the ONU. Similar to the G.984.5 wavelength enhancement standard, 10GEPON ONUs are required to install a bandpass filter (BPF) to filter out legacy signals, including video signals at the 1550 to 1560-nm spectrum. As a result, the standard settles on the wavelength range 1575 to 1580 nm to minimize the cost and insertion loss caused by the BPF. Five nanometers of the transmission window is selected for a cooled 10-Gb/s DFB laser.

3.6.2 10GEPON MAC Modifications

To ensure maximum compatibility toward legacy 1GEPON systems, the IEEE 802.3av standard makes few modifications even to MAC layer operations. Thus, the structure of both upstream and downstream Ethernet framing are in principle preserved. However, the addition of the new FEC option, the 64b/66b line code, and

their combination do require corresponding changes in the framing process to ensure proper synchronization. In addition, the potential mix of 1G- and 10G-capable ONUs would also require a modified MPCPDU for the OLT to properly acquire ONU's transmission capability.

3.6.2.1 10GEPON Framing with FEC Consideration Due to 64b/66b line coding, each code block in the 10G signal is taken as 66 bits. FEC encodes the output of the line coding block and operates on each code block. Since RS(255,223) is employed, each FEC code word can operate on 27 blocks (222B + 6b) of payload and generate four blocks of parity checks.

In the downstream direction, the ONU should synchronize on the boundary of both the 66b code block and the FEC code word boundary. Since the FEC code word is aligned to the code block, finding the FEC code word would also implicitly locate the code block boundary. As a result, the ONU could rely on a simple code word synchronization process that could lock to the 31-block pattern within about 6 μs. The 6-μs synchronization time is considered trivial considering the much larger overhead by the autodiscovery process. The code word synchronization process uses a circular buffer with a 32 FEC code word and searchs for a 31-block pattern with the EPON synchronization header.

Upstream burst framing is modified to enable for OLT to lock to the FEC code word boundary on pre-FEC data. Modified upstream burst framing is illustrated in Figure 3.28. To facilitate FEC code word and 66b code block synchronization, the last code block in the synch time portion of the upstream burst is changed into a unique start of data (SOD) delimiter. However, the data scrambler would lose synchronization during locking to the SOD delimiter because the delimiter is not scrambled, in order to preserve its unique pattern. As a result, two more idle code blocks are added at the beginning of the payload section to allow the scrambler to resynchronize and the OLT to detect a start of packet (SOP) after the scrambler gains synchronization. The last two added idle code blocks effectively increase the length of the upstream payload section, and corresponding changes in the calculation of the upstream queue REPORT are also required to account for the expanded burst size. In addition, the upstream data detection function is modified to further delay the transmission of data to account for the added idle blocks.

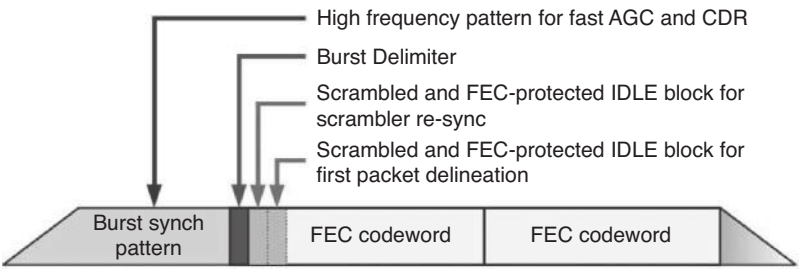


FIGURE 3.28 10GEPON upstream burst frame format.

3.6.2.2 MPCPDU Modifications Some MPCP protocol is modified in the amended 802.3av standard to support corresponding changes in the PMD sublayer. For example, 10GEPON OLT/ONU units now need to learn about system transmission capability during the autodiscovery process. For another example, the original 1GEPON standard specifies very relaxed timing requirements: 512 ns for ONU laser-on/off time and $\leq$ 832 ns for AGC, CDR, and code group alignment time. It has been observed that all existing ONU lasers in the marketplace perform much better than the timing specified. However, the original standard fixes the laser-on/off guard time and thus prevents the system from performing to its full capability. As a result, a mechanism to negotiate the laser-on/off time flexibly is added to the MPCP protocol in the 10GEPON specification.

Mixed-Mode Autodiscovery Process The autodiscovery process in a mixed-mode 1G- and 10GEPON system must be able to advertise OLT's upstream receiving capability (1G or 10G upstream) and preferably could allow different classes of ONU to register in separate windows. To facilitate the mixed-mode autodiscovery process, the following MPCPDU control messages are modified.

- Discovery GATE. A 1-byte discovery information bitmap is added after the original synchronization field to indicate OLT's capability (bit 0) and types of discovery window (bits 1 and 2). The discovery window could either be dedicated for 1G ONU (bit 1 = 1), 10G ONU (bit 2 = 1), or both (both bits set to 1). The reason for differentiating the autodiscovery window is to prevent the ONU from attempting registration during the discovery window, where it has no chance of succeeding.
- REGISTER_REQ. A 1-byte discovery information bitmap is added after the original pending grants field. This bitmap is symmetrical to the one added in the discovery GATE message and serves a similar purpose.

Adjustable Laser-On/off Time To facilitate flexible adaptation of ONU laser-on/off time into upstream burst guard time, the IEEEP802.3av task force considered two options. The first was to use OAM query through the OMCI channel, and the second was to incorporate the ONU/OLT exchange within the MPCP protocol. The latter option was selected because it allows the OLT to learn the ONU burst guard time parameter from the beginning. Since the MPCP state machine would be modified in 10GEPON, adding other minor changes would not add significantly to the cost. The modified MPCPDU control messages are:

- REGISTER_REQ. A 2-byte laser-turn-on time (1B) and laser-turn-off time (1B) by which information is added after the newly added discovery information bitmap.
- REGISTER_ACK. The OLT echoed the received laser-turn-on/off time using a symmetrical 2-byte field in the REGISTER_ACK message.

Broadcast LLID The standard specifies a different broadcast LLID for the 10G broadcast channel ($0x7FFE_{16}$) and the 1G broadcast channel ($0x7FFF_{16}$). The purpose of the unique broadcast LLID assignment is to simplify Ethernet switching in the OLT, which is expected to support a mix of 10G (XUAI–XGMII interface) and 1G channels.

3.6.3 10GEPON Coexistence Options

Future 10GEPON systems are expected to gradually replace the 1GEPON system. During the upgrading process, a mixture of 10GEPON (with mixed 1G and 10G upstream transmission capability) and legacy 1GEPON ONUs are expected to share the same optical distribution network infrastructure. As a result, a coexistence-capable option is desirable to allow both systems to serve over the same physical infrastructure.

3.6.3.1 10GEPON Coexistence Architecture The 10GEPON coexistence architecture depends on a hybrid WDM–TDMA technique. In the downstream direction, 10GEPON signals are transmitted over the range 1575 to 1580 nm and a 10G-ONU uses a BPF to filter out the 10G signal over a legacy signal. A legacy ONU would, however, also need to install a new wavelength-blocking filter to block out the 10G signal at the time of the first 10G signal upgrade.

There are more possibilities in the upstream direction because a 10GEPON ONU could transmit at either 1 or 10 Gb/s, or both. The 10GEPON PMD specifications intentionally set ONU transceiver characteristics similar to those used by 1GEPON. As a result, 10GEPON ONUs would transmit in the 1260- to 1280-nm transmission window and overlap with the original 1G signal. To multiplex the signal in the time domain, the system would use a dual-rate TDMA scheme to separate a 10G signal burst from a 1G signal burst (the operation to multiplex a new 1G signal and a legacy 1G signal is trivial). In this scenario, OLT needs to employ a dual-rate-capable APD receiver.

3.6.3.2 Coexistence Options Using a downstream WDM overlay and an upstream TDMA technique, 10GEPON can offer a number of coexistence options. Table 3.12 summarizes the plausible 1G- and 10GEPON coexistence options.

TABLE 3.12 1G- and 10G-EPON Coexistence Options

Downstream Transmission	Upstream Transmission	ONU Supported Type
2 λ	Single rate	1G-EPON 10G-PRX-EPON (1G US)
1 λ (1570–1580 nm)	Dual rate	10G-PRX-EPON (1G US) 10G-PR-EPON (10G US)
2 λ	Dual rate	1G-EPON 10G-PRX-EPON (1G US) 10G-PR-EPON (10G US)

Overall, the IEEE 802.3av standard defines the 10GEPON PMD parameters and some MAC and MPCP modifications. The objective of the specification is to encourage the use of low-cost optical components and enable architecture that could be backwardly compatible with existing 1GEPON systems. Like its IEEE 802.3ah predecessor, this first version of this IEEE 802.3av specification leaves many areas open to vendor innovation. Some of the newer issues, such as designs of a high-performance OLT dual-rate receiver and mixed-mode DBA scheduling, could create interesting new research areas for future study.

3.7 NEXT-GENERATION OPTICAL ACCESS SYSTEM DEVELOPMENT IN THE STANDARDS

Interest in pursuing a more capable next-generation optical access (NGA) network is already evident by the ratification of ITU-T G.984.5 (WDM), G.984.6 (reach extension), and IEEE 802.3av (10GEPON) standards. One of the major driving forces in NGA is further reduction in the operating expense and capital expense of optical access networks. Despite the use of different approaches, the objectives of these new standards are similar. They all look to support more users, over a longer reach, at a faster rate, and at lower cost.

3.7.1 FSAN NGA Road Map

To classify NGA development based on the expected demand and maturity of the necessary technology, FSAN further specifies two generations of NGA network. Figure 3.29 shows the FSAN NGA road map. Backward compatibility is the key to NGA-1 networks, which could be further classified into single-wavelength downstream or upstream bit-rate enhancement or WDM. ITU-T G.984.5 and IEEE 802.3av standards could all be considered as NGA-1 types of networks. In addition, FSAN also predicts the demand for such networks in the future. With the expectation that better

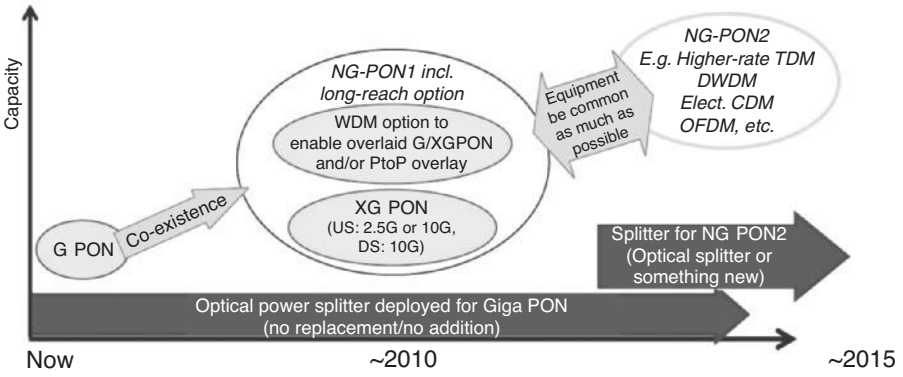


FIGURE 3.29 FSAN NGA road map.

TABLE 3.13 Summary and Comparison of ITU-T G.984.5, G.984.6, and IEEE 802.3av Standards

Characteristics	802.3av	G.984.5/6
DS bit rate (Gb/s)	10	
DS wavelength (nm)	1575–1580	1524–1635
US bit rate (Gb/s)	10/1	
US wavelength (nm)	1260–1280	1260–1360
Max. loss budget (dB)	29	28/55
Max. reach (km)	20	20/60
DS coexistence	WDM	WDM
US coexistence	TDMA	WDM
OLT requirement	Dual-rate receiver	

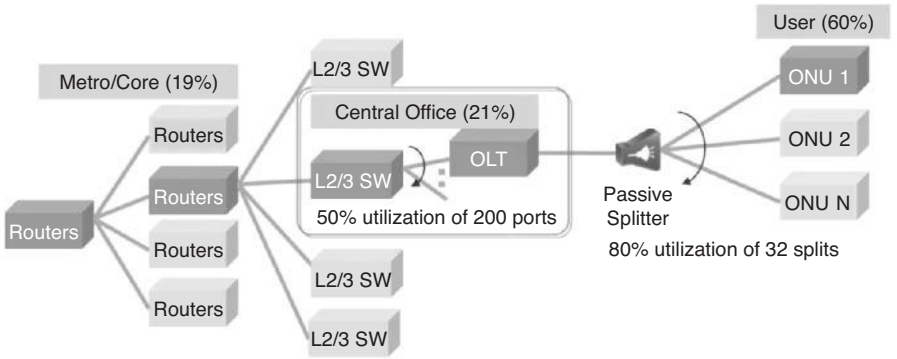
components and technology would become available in the future, NGA-2 networks are not constrained by existing ODN parameters. NGA-2 may very well use a new fiber network and, for example, employ arrayed waveguide grating (AWG) in place of a passive splitter. The ITU-T G.984.6 recommendation is considered an early type of NGA-2 network, as it is expected to use optical amplifier technology to support ODNs with up to a 60-km logical reach distance. Table 3.13 further summarizes and compares these recently ratified NGA standard options.

3.7.2 Energy Efficiency

In response to a growing awareness of the energy consumption issue in telecommunication networks, both the ITU-T and IEEE standard bodies have proposed studies of energy-saving techniques for PONs. Conserving energy is not only a noble cause for the operators but a realistic one as well. This is because a significant amount of energy is expected to be consumed by the access segment of the network in the future. Figure 3.30 illustrates the breakdown of energy consumption in the current core/metro/FTTH network hierarchy.

As the computations in the figure illustrate, the majority of energy consumed in a future network could reside in the ONU. To combat the problem, the ITU-T is exploring a dozing-mode option that would put POTS into a low-energy-consuming role but maintain operation of a 911 lifeline. The IEEEP802.3av task force also studied a ONU sleep-mode option that would put an ONU into a sleep mode when it is inactive. Recent proposals have suggested the feasibility of putting an ONU into a sleep mode even when it has active traffic, by exploring the gap in current traffic allocation cycles [7].

There is also growing (or renewed) interest in consolidating the number of central access offices because of the high energy footprints (and thus bills) related to central office buildings. This is because significant amounts of operating expense are currently being spent just to support the electrical supply and cooling of these buildings. NGA-2 long-reach PONs seem to be a promising solution that could fundamentally



	Power	Users	Power per User	Note
ONU	5 W	1	5W	EPON
OLT	15 W	N_{OLT} = 32*80%	0.59W	EPON
Aggregation	3 kW	N_{SW} = 200*50%*N_{OLT}	1.17W	200 Gb/s Switch
IP Core	3 kW*n	$N_{level-(n)}$ = 4*$N_{level-(n-1)}$	1.56W	200 Gb/s Router

FIGURE 3.30 Distribution of energy consumption in current core/metro/FTTH infrastructure.

enhance the loss budget performance in an access network and thus help reduce the number of central offices.

3.7.3 Other Worldwide Development

There are also activities that are being developed outside ITU-T and IEEE standard bodies. Korea Telecom, for example, has had the most WDM-PONs in deployment. Its WDM-PON architecture employs a proprietary solution provided by Novera and relies on coarse WDM transceivers and the use of wavelength multiplexer/demultiplexer AWG in the distribution network. China's Communications Standards Association (CCSA) recently added 2GEPON option as a specification for the country's fiber-based access network.

3.8 SUMMARY

In this chapter we present the architecture and protocol for ITU-T BPON, GPON, and IEEE EPON standards. A comprehensive and very detailed overview of each system's specifications and implementation is presented. Three recently ratified next-generation access network standards—the ITU-T G.984.5 WDM enhancement, ITU-T G.984.6 reach extension, and IEEE 802.3av 10GEPON—are presented in detail. Technical challenges from previously unresolved issues such as automatic detection

of upstream denial of service attacks, as well as new ones such as mixed-mode dynamic bandwidth allocation, are presented. Finally, we summarize the latest fronts in fiber access networks, including energy efficiency and deployment outside ITU-T and IEEE specifications.

REFERENCES

1. PON standards: ITU-T G.983 for APON and BPON, G.984 for GPON, IEEE 802.3ah for EPON, and 802.3av for 10GEPON.
2. G. Kramer, B. Mukherjee, and G. Pesavento, IPACT: a dynamic protocol for an Ethernet PON, *IEEE Commun.*, vol. 40, no. 2, Feb. 2002, pp. 74–80.
3. Y. Luo, S. Yin, N. Ansari, and T. Wang, Resource management for broadband access over time-division multiplexer passive optical networks, *IEEE Network*, vol. 21, no. 5, Sept.–Oct. 2007, pp. 20–27.
4. M. Hajduczenia, P. Inacio, H. de Silva, M. Freier, and P. Monteiro, On EPON security issues, *IEEE Commun. Survey Tutorials*, vol. 9, no. 1, 1st quarter 2007, pp. 68–83.
5. IEEE P802.3av Task Force, *10Gb/s, Ethernet Passive Optical Network Archive*, http://www.ieee802.org/3/av/.
6. 802.3av-2009, *IEEE Standard for Information Technology*, Part 3, Amendment 1, *Physical Layer Specification and Management Parameters for 10Gb/s Passive Optical Networks*, Oct. 30, 2009.
7. S.-W. Wong, L. Valcarenghi, S.-H. Yen, D. Campelo, S. Yamashita, and L. Kazovsky, Sleep mode for energy saving PONs: advantages and drawbacks, presented at the 2nd IEEE GreenComm Workshop, Nov. 2009.

CHAPTER 4

NEXT-GENERATION BROADBAND OPTICAL ACCESS NETWORKS

In Chapters 1 to 3 we presented an overview of several broadband access technologies, the basic components and systems that are used in optical communications, and the current network architectures and protocols for time-division multiplexed passive optical networks (TDM-PONs), including BPON, GPON, EPON, and 10GEPON.

In this chapter we present and discuss further development of passive optical networks. We start by taking a look at the evolution paths currently envisioned for TDM-PON networks, mostly in terms of bandwidth increase and reach extension. Then we discuss wavelength-division multiplexed passive optical network (WDM-PON) components and network architectures that have been developed in research laboratories and have been used in a few trial deployments around the world. These WDM-PONs can provide much higher bandwidths to customers by making use of many wavelengths, in contrast to TDM-PONs, which have just one wavelength for data transmission in each direction (downstream and upstream).

We then present an overview of research hybrid TDM/WDM-PON network architectures. These network architectures allow for a progressive evolution from TDM-PON to WDM-PON. By the time WDM-PON network start being deployed, a considerable number of TDM-PON deployments will be in place. Thus, WDM-PON networks would ideally make use of the outside plant deployed for TDM-PONs, minimizing the changes to this plant and disturbing the current TDM-PON customers as little as possible. The hybrid TDM/WDM-PON architectures discussed in this chapter provide a smooth upgrade path from TDM-PON to WDM-PON.

Broadband Optical Access Networks, First Edition. Leonid G. Kazovsky, Ning Cheng, Wei-Tao Shaw, David Gutierrez, and Shing-Wa Wong.

We conclude by presenting briefly protocols that allow resource sharing in WDM-PON. This is useful in making it possible to share expensive WDM resources, such as tunable lasers, among many users; this reduces the overall cost per user. We also discuss scheduling algorithms to provide QoS in WDM-PON networks. These algorithms, and the provision of QoS, are important access deployments, where the same connection may be used for multiple purposes and the traffic may combine data, video, and voice.

4.1 TDM-PON EVOLUTION

As we saw in earlier chapters, passive optical networks (PONs) lower network maintenance costs from those of to active networks by eliminating the need for power-consuming elements and supporting infrastructure in the optical distribution network (ODN). TDM-PONs are reasonably priced for access deployments since they use components that are relatively inexpensive for optical networks. TDM-PONs use the 1490-nm attenuation window for downstream transmission and the 1310-nm attenuation window for upstream transmission. Since the manufacturing technology for 1310-nm transmitters is very well developed, the components to transmit in this range are inexpensive. This allows the cost of an optical network unit (ONU), of which one is needed per customer, to be reasonable.

In the last few years, the number of TDM-PON deployments in Asia and North America has increased considerably. In Asia the preferred technology has been Ethernet PON (EPON), (sometimes referred to as GEPON in Asia), while in North America, deployments have used primarily broadband PON (BPON) and, more recently, gigabit PON (GPON) [1].

Even though the downstream and upstream data rates in EPON and GPON are relatively high (1 Gb/s for EPON and 2.5 Gb/s for GPON), this bandwidth is shared among 16 to 128 users. At present, this sharing does not affect the performance perceived by the end user. This is because, currently, most of the use of the Internet is for bursty data (e.g., HTTP browsing, SMTP e-mail, P2P file sharing, or FTP data transfers), which can be statistically multiplexed without the end user noticing it [2]. However, video and other multimedia streams are quickly becoming the predominant type of traffic in terms of both use and traffic percentage. This is particularly so for home subscribers, who are increasingly using online video services. Multimedia streams require high bandwidths (e.g., around 20 Mb/s are needed for a high-definition video channel). This required bandwidth cannot be statistically multiplexed, since they are continuous data streams and can easily saturate the capacity of current TDM-PON technologies.

Several enhancements have been proposed for TDM-PONs, which we review in this section. The main concern of these enhancements has been to increase the downstream bandwidth to the customer. This can be achieved in many ways, such as by increasing the line bit rate for everyone, which requires the simultaneous upgrade of all the customers in the network, or by increasing the line bit rate just for

new customers through electrical multiplexing or spectral line coding, enabling the coexistence with legacy TDM-PON users.

Another important concern of research has been to extend the reach of TDM-PONs so that the area covered, and thus the number of users serviced by the same network, increases without needing to deploy new networks. We also present and discuss the mechanisms proposed for doing this.

4.1.1 EPON Bandwidth Enhancements

The most direct approach to enhancing a TDM-PON network capacity is to increase the line bit rate. As discussed earlier, the P802.3av 10GEPON task force was created to achieve speeds in the 10-Gb/s range on EPON. From a technical perspective, one of the main challenges at this data rate is to support burst-mode transmission in the upstream direction at the optical line terminal (OLT). It is difficult to achieve this with current conventional integrated-circuit technology [3]. Even if this technical difficulty is overcome or an alternative technology such as silicon germanium is developed to meet this requirement, the cost might be too high for access networks. Because of this, the 802.3av task force plans to standardize an asymmetrical version with 1-Gb/s data rate in the upstream direction.

In addition to the 10-Gb/s burst-mode transmission issue, service providers need to ensure that the enhanced rate signals can coexist with the existing signals if they plan to gradually upgrade their existing subscriber base to the higher rate services. An abrupt service upgrade to all existing subscriber bases would require a new ONU for every existing user before the enhanced rate service could be deployed.

It has been shown that the signals can be effectively separated using electric multiplexing or spectral line coding [4]. Both approaches, however, have disadvantages. Electrical multiplexing, which inserts higher-data-rate frames between existing frames, requires each new ONU to be equipped with a multirate clock recovery circuit. In addition, scrambling coding standards and forward error correction (FEC) schemes need to be extended. The effects of fiber dispersion at 10 Gb/s also need to be considered.

Due to the disadvantages of same-wavelength overlaying, enhanced rate signals could be launched at new wavelengths outside of the upstream and downstream basic bands. This, in effect, would mean that WDM techniques would be needed.

4.1.2 GPON Bandwidth Enhancements

WDM enhancement is a promising way to make better use of the deployed fiber capacity and provide new network services. Service providers can provide premium services to a subset of users on additional CWDM or DWDM wavelengths, different to the upstream and downstream basic bands (1260 to 1360 and 1480 to 1500 nm, respectively). The preferred wavelength range for WDM enhancements is 1530 to 1605 nm, since in this window there is low attenuation in commonly deployed standard single-mode fiber (ITU-T G.652 and it corresponds to the erbium-doped fiber amplifiers' (EDFAs) gain window.

The ITU G.983.3 amendment specifies two options for the enhancement band: 1539 to 1565 nm for digital services or 1550 to 1560 nm for video distribution. For example, in the United States, Verizon uses the 1550-nm overlay to provide RF broadcast video signals [5].

In order to allow coexistence, WDM enhancements in GPON require wavelength blocking filters to avoid interference with the downstream basic band at the ONUs. In anticipation of future WDM enhancements, the full service access network (FSAN) working group is contemplating standardization through the ITU of such blocking filters to make legacy GPON ONUs blind to new WDM wavelengths [6].

4.1.3 Line Rate Enhancements Research

Service providers who have already invested in PON deployments require TDM-PON enhancement plans that will minimize upgrading costs and avoid disturbing existing customers. Table 4.1 summarizes the two most likely techniques for line bit-rate enhancements without disturbing current customers. Line bit-rate enhancements are those that increase data transmission speeds over the links. In contrast, WDM enhancements, which we discuss in the following sections, use new optical wavelengths for data transmission.

The electrical multiplexing technique consists of making use of the empty downstream windows that are not being used at the current rate (e.g., 1 Gb in EPON) to transmit data at a much higher rate (e.g., 10 Gb). In this case, legacy ONUs are supported, since they simply will not be able to receive the higher-bit-rate transmission. The issue with the new OLT and ONUs to support this technique is that they require having the ability to clock and data recovery at the two different rates.

The spectral line coding technique refers to making use of the available electrical signal spectrum to simultaneously transmit legacy bit-rate and higher bit-rate streams as discussed in the following section.

4.1.3.1 SUCCESS-LCO Service Overlay In this section we discuss the service overlay technique to increase the bandwidth of passive optical networks, as described by Hsueh et al. [4]. Consider the network architecture shown in Figure 4.1, in which a new OLT and new ONUs share the same infrastructure with the existing equipment. It is not trivial to smoothly upgrade the downstream service. In conventional TDM PON, existing ONUs typically have a broadband transmit/receive optical filter (1300 nm/1500 nm), as opposed to a narrowband wavelength-selective filter.

TABLE 4.1 TDM-PON Line Bit-Rate Enhancements

Technique	Advantages	Disadvantages
Electrical multiplexing	Supports legacy ONUs	New OLT/ONUs require multirate clock and data recovery.
Spectral line coding	Supports legacy ONUs; simple software implementation	Lower efficiency due to coding

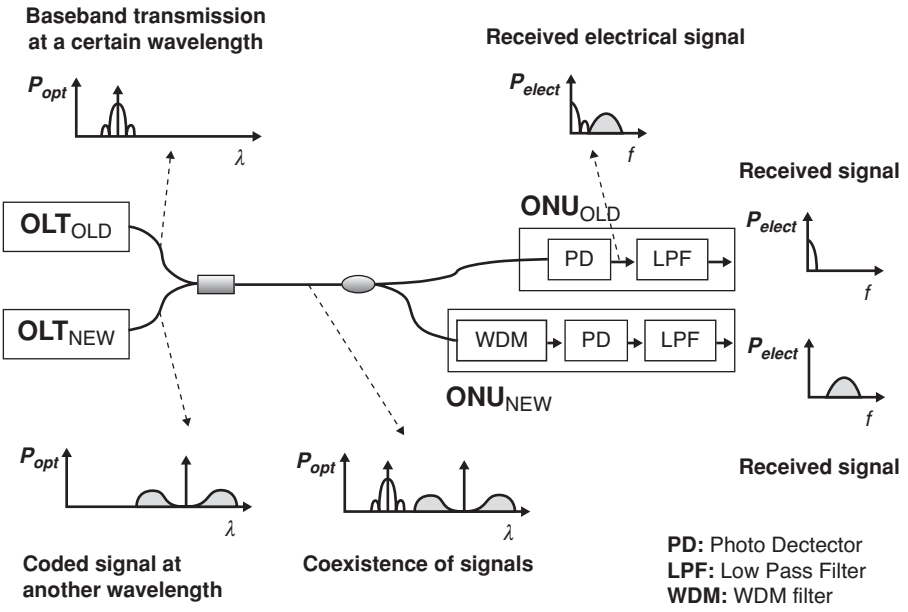


FIGURE 4.1 Upgrading existing TDM-PON using an overlay.

Therefore, any enhancements must take into account that existing ONUs will see all wavelengths that make it onto their leg of the distribution fiber.

To enable a smooth upgrade and service coexistence, three techniques are discussed in [4]: (1) launching new services in new wavelengths and retrofitting the existing equipment and/or infrastructure; (2) using a subcarrier multiplexing (SCM) technique to move the spectrum of the new service out of the baseband of existing ONU; and (3) using spectral line coding to shape the spectrum of the new service to minimize its interference with existing users. Compared to the spectral line coding technique, the first two methods are less favorable, due to the cost of hardware replacement and optical/electrical components. In principle, using spectral line code enables a graceful upgrading path without replacing existing PON equipment. During the upgrading period, only end users demanding enhanced services are required to replace their equipment. Line codes are flexible and can be reconfigured whenever necessary. Therefore, researchers focused on investigating the spectral line code technique to allow for the coexistence of the existing lower-bit-rate signal and higher-bit-rate overlay signal.

In an effort to capitalize on the advantages of both SCM and WDM, novel line codes were developed that suppress low-frequency components while maintaining dc balance [4]. The goal is to push the frequency content of the new high-bit-rate stream out of the baseband of the existing signal with minimal overhead. Figure 4.1 illustrates the spectral-shaping line codes assisted with WDM to provide a graceful upgrade for PONs. The existing OLT sends its traffic in nonreturn-to-zero (NRZ)

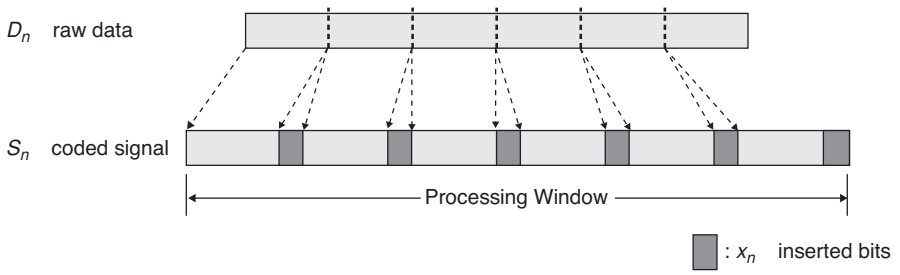


FIGURE 4.2 Coding hierarchy, consisting of a processing window, cells, data bits, and inserted bits.

format at a certain downstream wavelength. The new OLT, on the other hand, transmits its traffic at another wavelength, and the spectral power in the baseband region is avoided on purpose.

On the ONU side, the high-bit-rate signal from the new OLT is rejected by the low-pass filter (LPF) in the existing ONUs, while the signal from the existing OLT is filtered out by the WDM filter in the new ONUs. One clear advantage of this scheme is that the existing OLT, ONUs, and fiber plant remain unchanged. As subscribers upgrade their ONUs, they become members of the high-bit-rate PON. After all subscribers upgrade their ONUs, the old OLT is removed, and the new OLT can be reconfigured to maximize the data throughput.

The basic idea of the proposed line-coding technique is periodical insertion of extra bits to tailor the spectral shape of the coded signal. The data stream is divided into fixed-length processing windows of L bits (e.g., $L = 256$), which consist of small cells of C bits (e.g., $C = 8$). Figure 4.2 shows the coding hierarchy. A long processing window ensures long-term spectral properties, while small cells simplify bit processing. The processing windows can be partially overlapped, but the difference between overlapping and nonoverlapping windows is insignificant for reasonable window sizes. A cell contains data bits from the raw data stream and the inserted bits whose values are to be determined. Data bits and inserted bits are located in predefined positions in each cell. Decoding on the receiver side is straightforward: simply extracting data bits from their predefined positions in each cell. During network power-up or upon system reconfiguration, cell-level synchronization and signaling are necessary for the receiver to acquire correct positions of data bits in the bit stream. Details regarding the algorithm and analysis of the spectral line code are provided in the article by Hsueh et al. [4].

Finally, Figure 4.3 depicts the experimental setup used to compare four different overlay schemes, each with different efficiency and spectral overlap. The authors chose to compare codes at the same baud rate (10 Gbaud/s) because the baud rate is determined by the hardware. Under the same hardware constraint, one can easily tune the data rate by reconfiguring the line code in firmware or software. In all cases, the 1-Gbaud/s baseband signal source emulating existing service is a bit-error-rate tester (BERT) BERT1, which modulates the distributed-feedback (DFB) laser DFB1. The

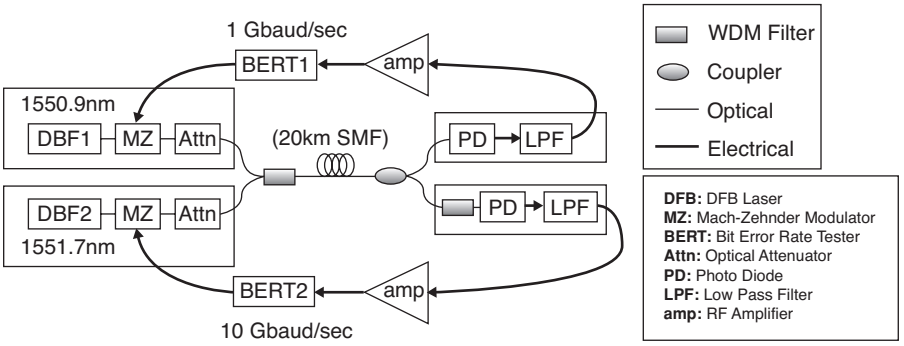


FIGURE 4.3 Experimental setup.

interferer or new service is realized by programming different line codes in BERT2, which modulates DFB2 at 10 Gbaud/s; the actual data rate depends on the code rate. DFB1 and DFB2 operate at DWDM wavelengths 1550.9 and 1551.7 nm, respectively. In the first test, a pseudorandom bit sequence (PRBS $2^{23} - 1$) is modulated onto the interferer in NRZ format. Since this interferer is not coded at all, it has a code rate of 1, but the spectral overlap in the baseband of the existing ONU is unacceptable. This scenario results in the worst performance. The other three codes are SUCCESS-LC68, SUCCESSLC58, and Manchester coding, with code rates of 6/8, 5/8, and 1/2, respectively. The measurement results of the BER vs. the signal-to-interference ratio (SIR) are discussed by Hsueh et al. [4]. They show that for a specific tolerable interference level, the optimal line code can easily be determined so that it maximizes the data throughput. The service overlay by the line-coding technique provides an elegant approach to smooth upgrading of existing PONs.

4.2 WDM-PON COMPONENTS AND NETWORK ARCHITECTURES

WDM allows making use of the enormous bandwidth that fibers have by transmitting simultaneously on multiple wavelengths. WDM has been used successfully in long-haul networks, but it has been relatively expensive for access network deployments. In long-haul networks, the cost of the components is shared among many thousands (possibly millions) of users, while in access networks, the amount of users per network is usually less than a hundred.

Several novel WDM-PON architectures have been proposed in recent years. We review briefly the technologies that these architectures have used.

WDM-PONs attempt to provide much higher average bandwidth capacities per user than TDM-PONs. In WDM-PONs, the capacity of the network is many times higher than that of current TDM-PONs, since multiple wavelengths can be used.

Figure 4.4 illustrates a basic WDM-PON. On the left, the optical line terminal (OLT), which resides at the central office (CO), has an array of transmitters and

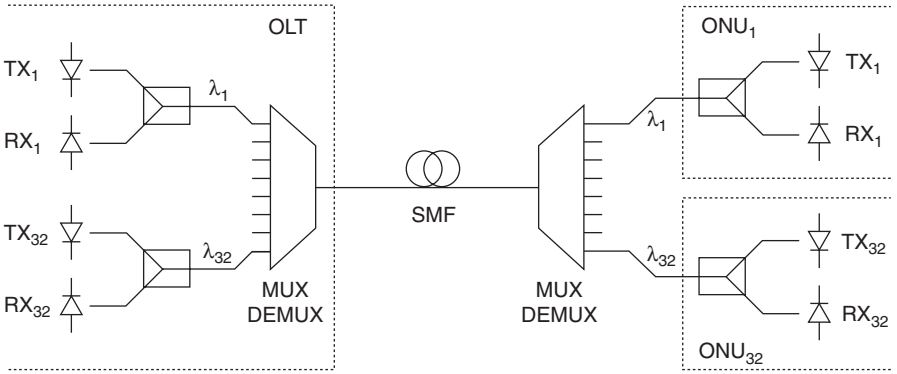


FIGURE 4.4 Basic WDM-PON.

receivers. Each transmitter–receiver pair is set at the wavelength band of the port of a multiplexing device to which the pair is connected. In this case, the multiplexing device is an arrayed waveguide grating (AWG). In this example, ONU_n transmits and receives on wavelength λ_n. A single-mode fiber (SMF) connects the OLT AWG to another AWG in the distribution network. Each port of the second AWG is connected to a different ONU. Each ONU has a passive splitter that is connected on one end to a transmitter and on the other to a receiver. In this particular architecture, the receiver need not be wavelength selective; however, the transmitter needs to be set at the particular wavelength of the AWG port to which it is connected. It is worth noting that the distribution network may include multiple AWGs cascaded to increase flexibility, as discussed by Maier et al. [7].

For optical transmitters, conventional laser diodes such as Fabry–Parot (FP) lasers, distributed feedback (DFB) or distributed Bragg reflector (DBR) lasers, and vertical cavity surface emitting lasers (VCSELs) are normally used under direct modulation. Optical receivers generally use PIN detectors or avalanche photodiodes (APDs) for photodetection. Passive splitters can be implemented with fused fibers or planar light-wave circuits (PLCs), and athermal AWGs are commonly used as multiplexers/demultiplexers.

In the following sections we discuss some of the most important enabling technologies for WDM-PON and the corresponding network architectures.

4.2.1 Colorless ONUs

Architecture such as the one described in Section 4.4 is quite expensive for access networks, and having a separate ONU for each PON customer creates an inventory problem. Cost-effective implementation of WDM light sources at the ONU side is necessary for WDM-PONs. They also need to be *colorless*, that is, wavelength independent, so that the equipment is the same across all customers. If the ONUs are not colorless, the service provider would need to have multiple types of replacement ONUs in stock and would need to arrange a particular type of ONU for each customer.

For example, customer 7 might only be able to use ONUs that operate on λ_7 instead of being able to use a generic colorless ONU. This creates additional installation, operational, and maintenance expenses for the service provider and, ultimately, for the end user.

Several approaches have been proposed and demonstrated for the implementation of colorless ONUs in WDM-PONs, such as a spectral sliced broadband light source, an injection-locked Fabry–Parot (FP) laser, or a reflective semiconductor optical amplifier (RSOA). The wavelengths of these light sources are not determined by the gain media but by external factors: optical filters or injected signals. As a result, the wavelengths are easier to manage without worrying about temperature changes or aging effects in the ONUs.

A lot of research has gone into making RSOAs, currently the technology of choice for colorless ONUs [8]. We discuss RSOAs in a centralized light sources architecture below. Some of the main issues that RSOAs have for this purpose are their noise levels and the speed at which they can be modulated. The reach that they can provide is also more limited than in technologies with a light source in the ONU, since they have to deal with back-reflection Rayleigh backscattering in the upstream transmission [9]. These issues may eventually justify the use of tunable lasers at the ONU. Tunable lasers are also a colorless ONU option, since all the ONUs can be identical, with each one tuning to a different wavelength. However, the main issue in this case is the cost of the components. Other options that have been considered are electroabsorption modulators (EAMs) and reflective electroabsorption modulators (REAMs) [10].

4.2.2 Tunable Lasers and Receivers

We now discuss the feasibility of using tunable lasers and receivers for WDM-PON and the resulting network architecture.

4.2.2.1 Tunable Lasers Tunable light sources provide flexibility and reconfigurability for network provisioning, minimize production cost, and reduce the backup stock required. Commonly used options for tunable lasers (TLs) are external cavity lasers, multisection DFB/DBR lasers, and tunable VCSELs. Due to cost concerns, it is desirable that tunable lasers be modulated directly, not through external modulators.

An external cavity laser is usually tuned by changing the characteristics of the external cavity, which consists of a grating or FP etalon. The tuning ranges of external cavity lasers are extremely wide, covering a few hundred nanometers. However, the long cavity length prevents high-speed direct modulation, so external cavity lasers are not suitable for fiber optic communications. Tuning speed and stability are also issues with external cavity lasers.

Traditional DFB or DBR lasers can support high-speed direct modulation and be thermally tuned over a few nanometers. However, the tuning speed is limited to the millisecond range. Multisection DFB/DBR lasers usually consist of three or more sections: an active (gain) section, a phase control section, and a Bragg section. Wavelength tuning is achieved by adjusting the currents in the phase-control and Bragg sections. Using multisection DFB or DBR lasers, the tuning speed can reach

nanoseconds by current injection, and tuning ranges over tens of nanometers can be achieved. Some multisection DFB/DBR lasers with sampled gratings can be tuned over 100 nm [11]. The disadvantages of multisection DFB/DBR lasers are mode hopping and complicated electronic control.

VCSELs emerged as a new type of semiconductor lasers in the 1990s. A typical VCSEL consists of a layer of gain medium sandwiched between two oppositely doped distributed Bragg reflectors. VCSELs have a potential for low-cost mass production because of simple one-step epitaxy and on-chip testing. However, the development of long-wavelength VCSELs has been hindered by the unsatisfactory optical and thermal properties of InP-based group III to V materials. New designs using different materials and dielectric mirrors have led to successful development of 1.3- and 1.5-μm VCSELs. A primary tuning mechanism for VCSELs uses a MEMS structure that changes the cavity length through electrostatic control. The tuning speed can be a few microseconds and the tuning range can reach 10 to 20 nm [12]. As its fabrication matures, VCSELs will be a strong candidate for access networks.

4.2.2.2 Tunable Receivers A tunable receiver can be implemented using a tunable optical filter and a broadband photodiode. Although this approach provides a simple solution, it can be bulky and expensive. As an alternative, CMOS-controlled tunable photodetectors with an integrated filtering mechanism in the photon-detection process have been reported [13]. The device consists of a few metal–semiconductor–metal detectors laid side by side with interleaved metal fingers. As multiple wavelengths impinge on the photodetector, different wavelengths form different fringe patterns. When a specific biasing pattern is applied to the biasing fingers, the photodetector can select one wavelength while canceling out the rest. Since the wavelength of this device is set electronically, the tuning time is essentially limited by the electronics switching time, which is on the order of nanoseconds. The channel spacing is limited by the coherence length of the laser system, and can be much less than the standard 50-GHz DWDM requirement. The detecting algorithm for the device can be applied to either 1550- or 850-nm optical systems. In addition, this device allows spectral shaping to adapt to specific system applications.

The disadvantage of this device is a limited scalability of the integrated interferometer. In order to have satisfactory spectral responses, the device needs to have an integrated interferometer that generates a few harmonic interference patterns simultaneously. This interferometer will become more complicated as the wavelength number in the access network scales up. It thus prevents a single device from being applied in a large access network. This limitation can be relaxed by the proper design of access network architectures.

4.2.2.3 Network Architectures Figure 4.5 shows what a WDM-PON with tunable lasers and receivers would look like. It would be very similar to Figure 4.4, except that in this case each laser and receiver can be tuned to whatever wavelength has been assigned to that ONU. That is, each ONU is colorless using tunable components. The TLs would need to cover a wide range of DWDM channels, as many as the range of the AWG MUX/DMUX. The wavelength output of the TL needs to be precise in

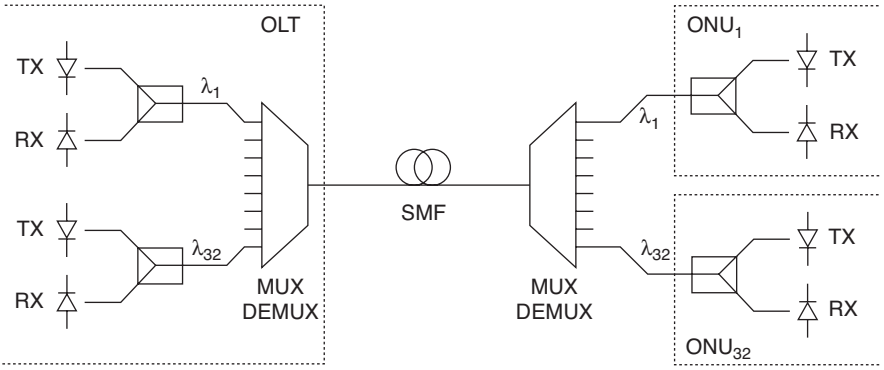


FIGURE 4.5 WDM-PON with tunable lasers and receivers.

order to stay within the AWG port band to which it connects. Thus, tunable laser diodes require expensive temperature-stabilization control circuitry. Even though having identical ONUs solves the inventory problem, the cost of these tunable devices at the moment is too high to be cost-effective for access. On the other hand, employing TLs in the CO can be economically feasible when they are time-shared among users.

4.2.3 Spectrum-Sliced Broadband Light Sources

A comb of optical signals, each with a unique wavelength, can be achieved by slicing the spectrum of a broadband light source. The broadband light source can be a superluminescent light-emitting diode [14], EDFA [15], or FP laser [16]. For wavelength selection, an AWG, usually deployed at the remote node, will slice a narrow spectrum of the broadband optical signals, and different wavelengths will be selected for different ONUs. If a tunable filter is used, the wavelengths of the optical comb can be tuned.

Network Architectures The spectral slicing technique was first used [17] for downstream transmission, and later [14,18,19] for upstream transmission. Figures 4.6 and 4.7 illustrate the downstream and upstream transmission with a broadband light source (BLS) using a spectral slicing approach. We describe the upstream case; the downstream case is analogous, just in the reverse direction. Note that the AWG has cyclic properties, so if both systems were to coexist (i.e., BLS with spectrum slicing was used in both directions), the wavelength range could be different in each direction to avoid issues with backreflections and Rayleigh backscattering in the same wavelength.

In the upstream case, each ONU has a BLS that is modulated with upstream data. Upon reaching the WDM MUX/DEMUX, the spectrum of each ONU is "sliced." After that, multiple data streams can be multiplexed onto the distribution network fiber. At the OLT, the signals are demultiplexed and a separate receiver is used for each ONU.

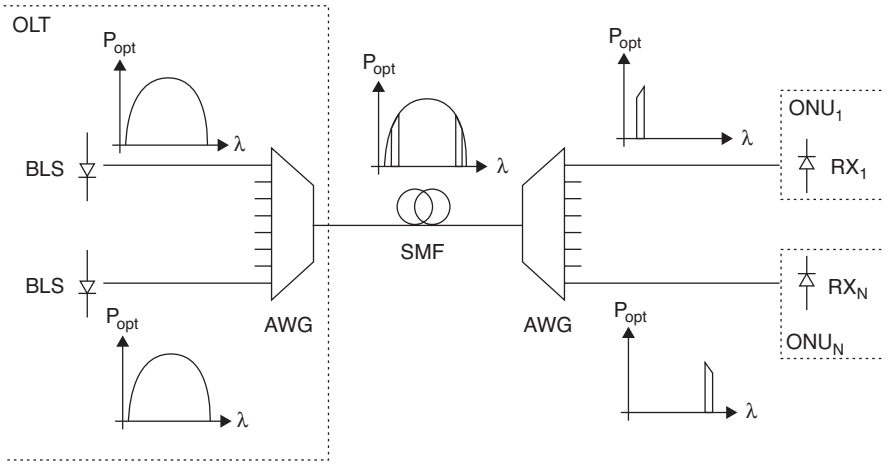


FIGURE 4.6 Broadband light sources WDM-PON: downstream transmission.

Compared to the WDM MUX/DMUX, the BLS spectrum should be sufficiently wide and its output power large enough so that even after being spectrally sliced (which leads to significant optical power loss, e.g., 10 dB [18]), the remaining power will still be enough to traverse the feeder fiber (about 20 km by current standards). This, added to the coupling, connectors, and devices losses, may induce up to 20 to 30 dB of additional loss after spectral slicing.

A sharp filtering profile is necessary in the WDM MUX/DMUX to reduce the crosstalk between adjacent channels. Another issue with the AWG is its thermal sensitivity. Since these devices are placed in the outside plant and not powered,

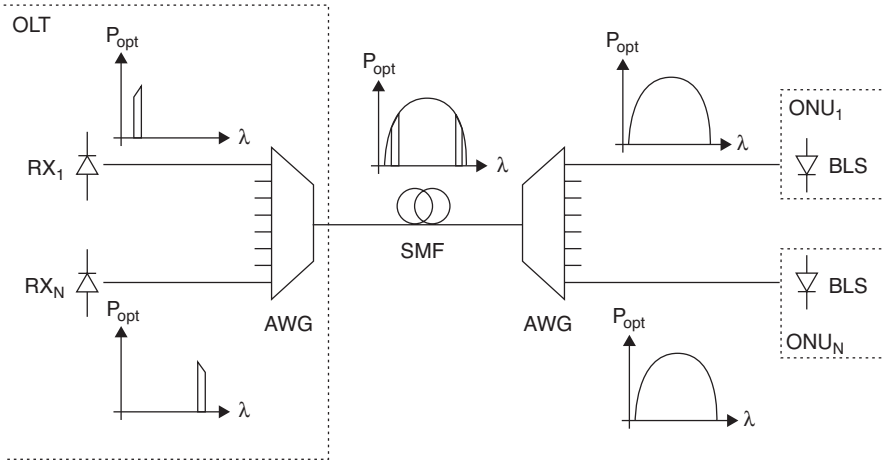


FIGURE 4.7 Broadband light sources WDM-PON: upstream transmission.

their wavelength properties should be insensitive to temperature changes. Theoretical analysis and experimental results [20] have shown that the adjacent-channel crosstalk can be a problem for this technique.

The main advantages of spectral slicing are simple implementation and low cost. The disadvantages are limited modulation speed, low power, and incoherent output that limits the transmission span.

A BLS with spectral slicing introduces some limitations to the network topology as well. In particular, the signals from different ONUs share the same spectrum. Therefore, they cannot be combined until after they have been spectrally sliced at the AWG; otherwise, upstream traffic collisions may occur. This reduces the flexibility of the network, since some TDM-PON deployments have more than just one splitting point.

4.2.4 Injection-Locked FP Lasers

When an external narrowband optical signal is injected into a multiple-longitudinal-mode laser such as an FP laser, the lasing mode can be locked to a single mode. The external optical signal acts as a seed for laser oscillation in the FP cavity, and the mode that is the nearest the peak wavelength of the injected optical signal will be locked to the injected light and other modes will be suppressed, as shown in Figure 4.8 [21]. The injected optical signal can be a spectral sliced broadband light

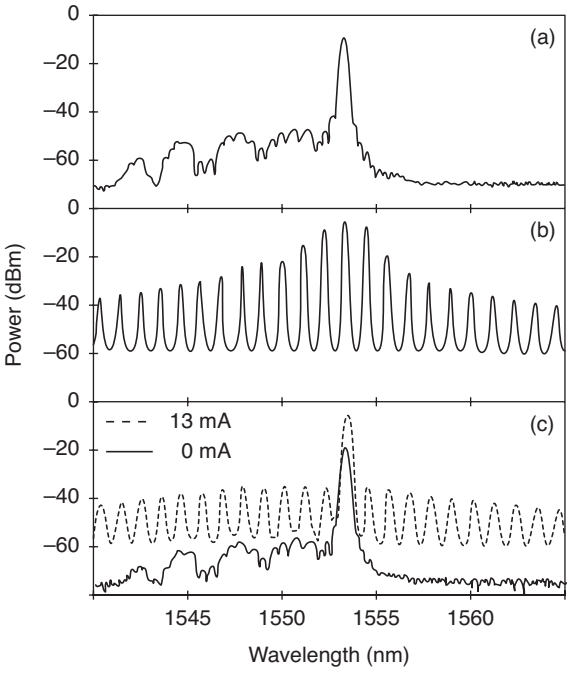


FIGURE 4.8 Injection locking a Fabry–Perot laser. (From [25].)

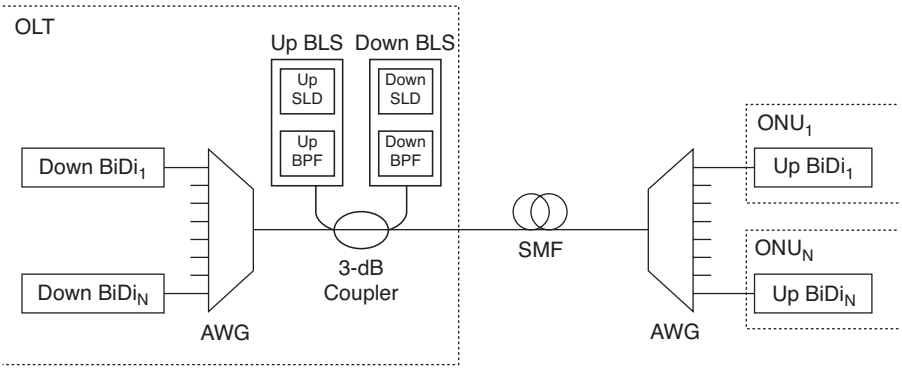


FIGURE 4.9 Injection-locked FP laser WDM-PON. (From [26].)

source (e.g., ASE) [22] or a modulated downstream signal [23]. The advantages of using injection-locked FP lasers are simplicity and low cost, and the modulation speed can be relatively high. The disadvantages are the limited wavelength locking range, high power required, and questionable stability. The modulation index, laser bias current, and power of external optical excitation must be chosen carefully to maximize the efficiency of injection-locked FP lasers [24].

Network Architectures Several network architectures have been reported based on injection-locked lasers [16,26], and even an actual deployment by Novera and Korea Telecom in Korea uses this technology [27]. Figure 4.9 shows an example in both the downstream and upstream directions. At the CO, the upstream BLS generates optical power that is sent downstream. This optical power is sliced by the remote AWG and is used to injection-lock the ONU FP laser. The locked FP laser is then modulated and the data are sent in the upstream direction.

In this architecture, higher optical injection power is required for a higher data rate to achieve colorless operation [26]. Since the signal is transmitted on the same wavelength as the injection seed light, the BLS power cannot arbitrarily increase because of Rayleigh backscattering and reflections. The scalability of this architecture needs further investigation.

4.2.5 Centralized Light Sources with RSOAs

Another alternative for colorless ONUs is to use centralized light sources (CLSs) at the OLT, as shown in Figure 4.10. In this approach, no light source is deployed at the ONUs; the OLT provides optical signals to ONUs, where the optical signals are modulated with the upstream data, possibly amplified, and sent back to the OLT. Since all light sources are located at the OLT, the requirements for wavelength provisioning and management are relaxed due to a centralized environment. The OLT manages on its own which wavelengths to use; there is no need for a wavelength assignment protocol between the OLT and the ONUs.

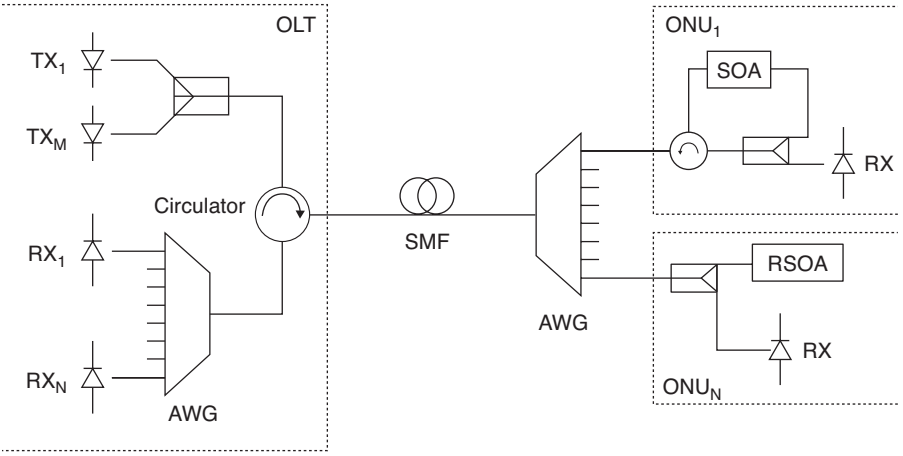


FIGURE 4.10 Centralized light sources: WDM-PON.

There are several options for modulating at the ONU, including Mach–Zehnder modulators, semiconductor optical amplifiers (SOAs), and reflective SOAs (RSOAs), which can all be used for upstream modulation. SOAs and RSOAs can modulate and amplify the signal simultaneously, making them preferable to an amplifier + modulator configuration, which would require two separate devices and thus extra manufacturing cost.

To transmit both downstream and upstream in a fully duplex manner in the same wavelength, various data modulation schemes have been proposed, including ASK-ASK using different power offset values, FSK-ASK, and SCM-SCM. These different schemes provide different capabilities and limitations [8].

There are some disadvantages to the CLS approach as well. Since the transmission in both directions is done on the same wavelength, there is some noise, particularly in the upstream direction, due to Rayleigh backscattering or the Brillouin effect. Another disadvantage is the polarization sensitivity of the RSOAs or optical modulators. Creating RSOAs that are less polarization dependent has been an important issue for manufacturers in the past few years.

Currently, RSOAs are still expensive for access networks, since they need a thermoelectric cooler (TEC) in the same package. Manufacturers are currently working on RSOAs based on new materials that can provide good performance without the TEC control mechanism, making them less expensive and power consuming. This would also allow the RSOA to fit in a TO-CAN package, allowing for a smaller footprint within the ONU [28]. In such a case, there would be no manufacturing limitations to prevent these devices from having very low costs, adequate for an ONU, when mass-produced.

Network Architectures CLS architectures were first proposed by Frigo et al. [29] and developed further by Kani et al. [30], Takesue and Sugie [31], and in the

SUCCESS-HPON architecture, which we discuss below. There has been increased interest in these architectures recently, due to the possibility of low-cost RSOAs. An actual deployment of this technology in Korea was recently announced [32].

Figure 4.10 illustrates a generic CLS architecture. In this case, all the light sources are at the OLT. These provide optical power for both downstream and upstream transmission. The downstream transmission case is straightforward from the OLT, then the ODN and the ONU.

Upstream data transmission works as follows: The transmitters at the OLT provide a carrier at the particular wavelength of an ONU; the ONU amplifies and modulates this carrier with its queued data, and sends it back upstream. A passive circulator at the OLT directs these upstream data onto an AWG followed by an array of receivers, one for each ONU. Since traffic in opposite directions is transmitted on the same wavelength and the same fiber, Rayleigh backscattering and reflections need to be minimized throughout the network. To achieve the goals, the optical power of the carrier needs to be high enough for upstream transmission, but limited to prevent significant Rayleigh backscattering interference.

Clean connectors throughout the distribution network are required to minimize the reflections. This is particularly important in this type of network, since both the continuous wave to be modulated and the signal travel through the same fiber. Backreflections can create issues in this type of setup.

Figure 4.10 also illustrates two possible ONU designs. ONU_1 has a circulator and a splitter. Incoming data are received by a receiver that does not need to be tunable while the incoming carrier is modulated and sent back. An SOA is used as an amplifier and modulator due to its potential for integration with the electronics of the ONU [30]. ONU_N illustrates a simpler design using a one-port RSOA. Again, the downstream data are received by a photodiode receiver while the upstream data are modulated onto a CW by the RSOA. In a paper by Prat et al. [33], the RSOA acts as a detector as well by measuring the voltage difference produced by the change in carrier density inside the cavity.

4.2.6 Multimode Fiber

There are several architectures that employ a centralized light sources approach (e.g., [29–31,33,34]) and SUCCESS-HPON, which we discuss later in the chapter. Most of these architectures have been developed on the premise that single-mode fiber (SMF) is being used in the ODN. While SMF shows high signal integrity in the long-haul network, it takes a large amount of time and money to install SMF in the field. On the other hand, multimode fiber (MMF) is easy to install and has already been installed on many campuses and local area networks. Therefore, the possibility of using MMF in an optical access network is of interest despite the fact that the bandwidth–distance product is limited by the modal dispersion in MMF. There are several efforts to increase the bandwidth–distance product. As a result, data transmission through a few kilometers at more than 1 Gb/s is already achievable, as reported [35–37]. In this section we discuss a CLS access network that includes MMF, as presented by Kim et al. [38,39]. Using physical mode confinement, the

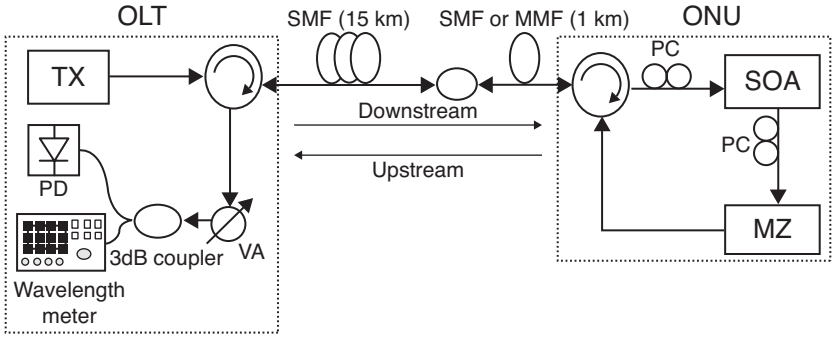


FIGURE 4.11 CLS + MMF experimental setup.

MMF link shows a performance comparable to that of the SMF link. The results were verified using BER measurements. Considering the fact that the MMF currently prevails in many local area networks, this architecture can expedite and complement PON deployments.

As shown in Figure 4.11, the network architecture follows the conventional frame of other centralized light source networks. To transmit a signal, a DFB laser is located at the OLT. Although the OLT supplies a carrier to the all ONUs, since the incoming signal will decrease as it goes through the downstream link, we need an amplifier in the ONU to compensate for the power budget of the overall signal path. Therefore, each ONU has an SOA and an intensity modulator to transmit the upstream signal. The upstream traffic passes through the same fiber link as does the downstream traffic and arrives at the receiver in the OLT. In bidirectional transmission, it has been reported that the upstream signal is more degraded than the downstream signal because the former is more exposed in Rayleigh scattering [33]. Therefore, in this experiment the authors investigate only the worst case (i.e., upstream transmission).

Generally, a passive splitter is located near the ONU in the PON access network. Hence, the authors divide the entire data link into two segments. One is from the OLT to the splitter, and the other is from the splitter to the ONU. A 15-km-long SMF is placed in segment 1 and 1-km-long MMF in segment 2. The MMF has a graded index and a 50-μm core diameter. To connect the single-mode component, both ends of the MMF are core aligned and spliced with a small piece of SMF pigtailed by a conventional physical contact adaptor. Since the authors observe the performance of a single ONU, the passive splitter does not give a significant effect except contributing to the power budget. Because of this, the authors do not include the passive splitter in the experiment setup: instead, using less SOA gain. Table 4.2 summarizes the loss of each component in the data link. Compared to an SMF of the same length, the MMF has additional 1.1 dB of loss caused by the core misalignment between the SMF and the MMF connection as well as a higher-order mode filtering at the interface from the MMF to the SMF connection.

To create upstream data, the DFB laser in the OLT again provides a CW at 1550 nm that is remotely modulated at the ONU with NRZ-coded $(2^{23} - 1)$-word

TABLE 4.2 CLS + MMF Loss of Components

System	OLT	Segment 1 (15 km)	Segment 2 (1 km)		ONU		
Component	Circulator	SMF	MMF	SMF	Circulator	PC	MZ
Loss (dB)	1.42	3.1	2	0.9	1.3	1	6

PRBS at 1.25 Gb/s through the intensity modulator. The overall power budget of the round trip signal path is 23.64 dB (before the 3-dB coupler). Therefore, a 2-dBm CW is launched from the OLT and the gain of the SOA is set to 10 dB. The gain of the SOA can be increased with the amount of the loss in the passive splitter, which depends on the number of ONUs. The signal received is coming from the isolator in the OLT and passing through an optical attenuator (HP 8156A) and a power splitter. One branch of the power splitter goes into an error analyzer (HP 70843A) after a SMF pigtailed commercially available photodiode, while the other is monitored by a light-wave multimeter (HP 8153A). As the power received is changed by varying the value of the optical attenuator, the bit error rate (BER) is measured along the varying levels of power received for three separate cases. First, a baseline without a fiber link is measured, and then the SMF (15 km) + MMF (1 km) link is measured. Replacing segment 2 with 1-km-long SMF, a third measurement is made. Comparing the three cases, the degree of difference caused by the MMF link is clear.

Figure 4.12 shows the results for the three cases. Comparing the baseline to case 3, there is a 3-dB power penalty. Although the effect of the chromatic dispersion

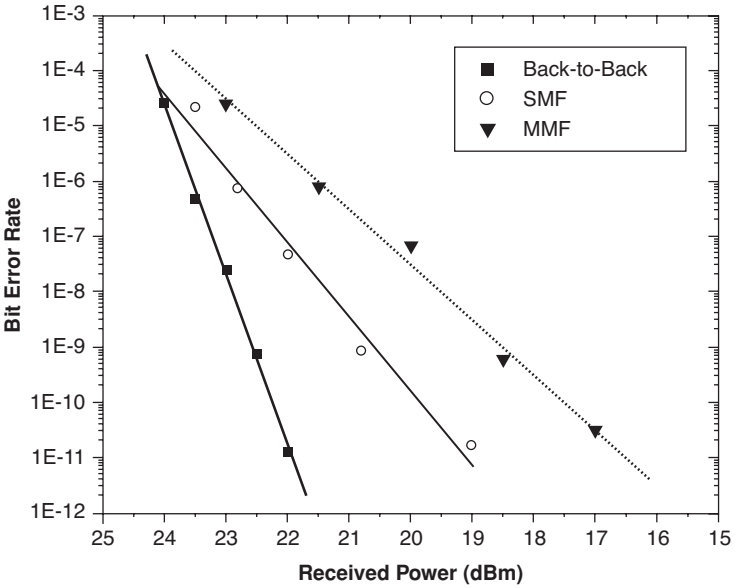


FIGURE 4.12 CLS + MMF BER measurements at the OLT.

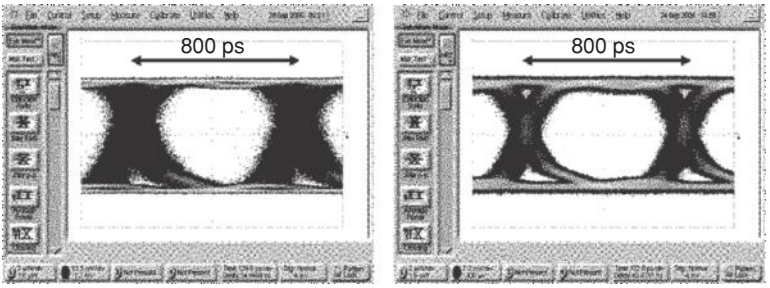


FIGURE 4.13 CLS + MMF eye diagrams.

for the 16-km-long fiber link at 1.25 Gb/s may be ignored, the signal can still be affected by Rayleigh backscattering. Furthermore, there can be other noise sources, such as imperfect isolation of the circulators and noise coming from the SOA in the ONU. When segment 2 is changed to a 1-km-long MMF, the signal can be degraded further, due to the modal dispersion. However, since the upstream signal is launched with the polarization controlled, the signal feasible to the upstream data link can be optimized. With finely tuned polarization control, a BER performance of up to 10^{-11} can be achieved with an additional 2-dB power penalty compared to case 3.

The eye diagram for case 2 at low and high BERs are shown in Figure 4.13. The left figure has a vague eye, and it becomes closed as time goes on, while a clear eye can still be seen on the right.

From this experiment, it is determined that CLS transmission in a SMF + MMF configuration is a pausible promising access network configuration. Using a core aligned connection and polarization control, high BER performance can be achieved with just a 2-dB-higher power penalty than that of the SMF link. Therefore, MMF can play a complementary role in the deployment of WDM-PON. For further details on this approach, readers are referred to a paper of Yam et al. [38].

4.3 HYBRID TDM/WDM-PON

In hybrid TDM/WDM PONs, TDM and WDM components coexist in the same ODN. Usually, some customers are served by the TDM infrastructure and others by the WDM infrastructure. The reason for this type of coexistence is that it may be desirable to make use of the same infrastructure deployed for TDM-PON to provide WDM-PON capabilities. As we will see next, this has to do with the evolution from TDM-PON to WDM-PON.

4.3.1 TDM-PON to WDM-PON Evolution

The WDM-PON architectures discussed earlier do not take into consideration the previous deployment of TDM-PON ODN. By the time that WDM-PON deployments

begin in earnest, a huge number of TDM-PON ODNs will have been deployed. Because of this, network operators will be unwilling to incur large new infrastructure costs and would prefer a smooth transition from TDM-PON to WDM-PON in terms of both cost and service. This migration from TDM-PON to WDM-PON must happen in a way that satisfies several operational, cost, and flexibility requirements. Some of these requirements are the following:

- Service disruptions to the current users must be avoided.
- Service levels to current users should not be downgraded.
- The network should support legacy TDM infrastructure and equipment when possible (e.g., fiber deployments, legacy ONUs).
- The cost of new WDM equipment should be minimized and reasonable for access networks.
- If possible, the cost of the network should increase slowly as new bandwidth demands are satisfied.
- If possible, the new WDM-PON architecture must be flexible enough to accommodate different WDM-PON ONU technologies.

The first two service-level requirements are quite obvious. The support for legacy infrastructure and equipment comes from the fact that it would be difficult to change equipment and technologies instantaneously across the ODN. Thus, ideally, a time of coexistence between legacy and new technologies should be allowed. Managing the cost of deploying the new WDM-PON technology is also an important issue. Finally, the last requirement is given because, as we have seen, there are several proposals for WDM-PON ONUs, and currently it is not clear which technology, convention, or protocol will be the most widely deployed. Because of this, ideally, the topology and architecture should be such that they will support multiple WDM-PON ONU technologies.

The technology used in the new WDM-PON should also satisfy some requirements, for it to be worthwhile to go through a migration from TDM-PON:

- Provide much higher bandwidths (e.g., 10 times as much) per user than those of current TDM-PONs, to justify the investment.
- Be flexible and scalable for future bandwidth demand increases.
- For small business users, enhanced security and protection features are desirable.

If the new WDM-PON technology is just marginally superior to TDM-PON in end-user bandwidth capacity, the upgrade will not be worth it. The new architecture should be flexible, so that it can adapt to new technologies that may crop up in the WDM-PON space and scalable so that it can support many users. In addition, many small business users will make use of the good bandwidths supported by TDM-PON and WDM-PON. For these users, enhanced security and protection features (such as continuity in the event of a fiber cut) are very important.

The authors' research group at Stanford University has proposed two smooth migration architectures from TDM-PON to WDM-PON developed under the Stanford University aCCESS (SUCCESS) project umbrella; these are the hybrid TDM/WDM PON (SUCCESS-HPON) and the dynamic wavelength allocation (SUCCESS-DWA) architectures, which are discussed in subsequent sections, along with other approaches and evolutions.

4.3.2 Hybrid Tree Topology Evolution

As we discussed before, current TDM-PONs have been deployed using a tree network topology. Such is the case for the three major types of deployments: BPON, GPON, and EPON. The standards allow for flexibility in the topology regarding the number of splitters and the splitting ratio of each, setting only the minimum and maximum attenuations supported (i.e., the attenuation range). For example, the ODN classes A, B, B+, and C, as defined in the standards, correspond to the 5 to 20, 10 to 25, 13 to 28, and 15 to 30 dB ranges. These attenuation ranges, in turn, limit the total number of users and the maximum distance between the OLT and the ONUs.

Even though many splitters could be throughout the ODNs, most current TDM-PON deployments in North America and Asia usually have only one 1 : 32 splitter [5,40] and a maximum reach of 20 km, as shown in Figure 4.14. In Japan, most deployments have a 1 : 4 splitter at the CP and 1 : 8 splitters at the remote nodes. Currently, there is as well a convergence of PON deployments toward the class B+ ODN (i.e., 13 to 28 dB range and 20 km reach) [40]. Since most of these PONs have been deployed from scratch, the fiber optic cables used in these deployments usually have several extra backup or dark fibers between the CO and the splitter and a few extra fibers from the splitter to the home (i.e., they are fiber-rich deployments, as we explain below). The number of fibers that have been included in these deployments is higher than the number needed even in the case of a high take rate; since the cost of redeploying extra fibers in the future for additional services is much greater, operators

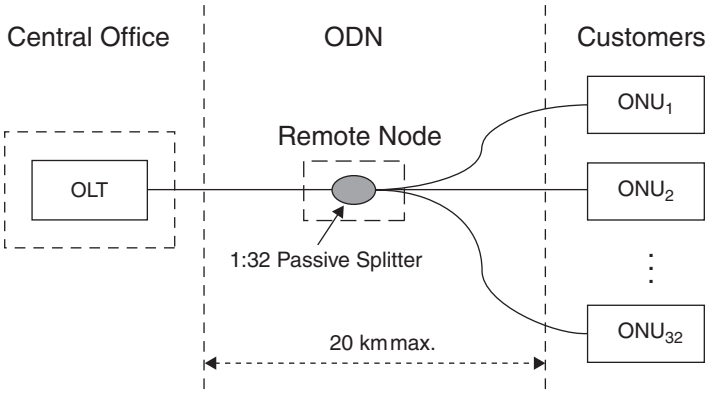


FIGURE 4.14 Typical TDM-PON deployment.

have included them already. The precise numbers depend on the expected take rate for an area, the cost of adding fibers, and the economic objectives for the following years.

We discuss now two alternative evolution paths for tree topologies from TDM-PON to WDM-PON. The first evolution path is for fiber-lean scenarios, where no extra fibers are available, and thus TDM-PON and WDM-PON wavelengths must coexist in the same fiber. The second evolution path is for fiber-rich scenarios, such as the one just mentioned above, where at least one extra fiber is available from the CO to the splitter and from the splitter to each ONU.

4.3.2.1 Fiber-Lean and Fiber-Rich Deployments

We can classify TDM-PON deployments into two categories: *Fiber-lean* deployments are those in which no new unused fibers are available in the ODN; in contrast, in *fiber-rich* deployments, backup or dark fibers were installed initially and are available for further network development (Table 4.3).

Fiber-Lean Deployments WDM enhancements can take place in both fiber-lean and fiber-rich deployments. In fiber-lean deployments, since there are no additional unused fibers, we need to make use of the currently used fibers. There are two natural ways of introducing new wavelengths on this type of deployment.

1. By making use of wavelength blocking filters, the TDM-PON ONUs can eliminate the interference cause by new wavelengths introduced. What makes a big difference on whether or not using wavelength-blocking filters makes economical sense is when they are deployed and installed. If it is part of the initial deployment, the cost of installing them is minimized. However, if their use is an afterthought and are added afterward, the cost of installing them throughout the network for every customer will be quite high. Another issue is that they

TABLE 4.3 TDM-PON WDM Enhancements

	Advantages	Disadvantages
Fiber-lean techniques		
Wavelength blocking filters	Inexpensive a priori filtering at ONU	Expensive a posteriori filtering at ONU; may affect power budget and reach
Remote node modification	Wavelength selectivity without making changes to existing ONUs	Requires ODN changes; may affect power budget and reach
Fiber-rich techniques		
Bundled fiber	Simple deployment by adding wavelength MUX/DEMUX in ODN	Extra bundled fiber is not always available
New fiber deployment	Completely new infrastructure for enhanced PON	Very expensive and unlikely to happen

may require additional power to reach the ONU from the OLT, thus affecting the power budget or reach specifications.

2. An easier approach, especially in the case in which the changes are being done as an afterthought, is to modify the remote node. This allows all the modifications to take place in a single place (the remote node) instead of at hundreds or thousands of places (the ONUs), thus making the logistics and practicality of this approach much better. Modifying the remote node allows for wavelength selectivity without making changes to the existing ONUs. Again, as in the wavelength-blocking filters approach, because of the filtering and changes to the remote node, which is passive, the changes may affect the power budget or reach specifications.

Fiber-Rich Deployments In the case of fiber-rich deployments, where unused fibers that were installed during the initial deployment of the ODN are available to make these enhancements, the approach is more straightforward. These unused bundled fibers can be used by a WDM overlay to provide WDM enhancements over the same ODN. We discuss this and provide an example in the following sections.

Of course, the last fiber-rich technique is to add new fibers to the ODN simply but expensively. This approach is very expensive and thus unlikely to happen, but we would like to suggest considering it as an option as well. It is worth noting that, in general, the attainable increase in network throughput is much greater with WDM enhancements than with the line rate enhancements discussed previously, since completely new wavelength channels become available.

4.3.2.2 Fiber-Lean Evolution In the fiber-lean scenario, the new WDM-PON wavelengths must coexist with the TDM-PON wavelengths in the same fiber. To do this, two waveband combiner/splitters (WCS) are installed, one at the CO and the other at the RN, and a new 1 : 32 athermal AWG MUX/DEMUX is installed at the RN, as shown in Figure 4.15. The legacy TDM-PON ONUs are not changed at all.

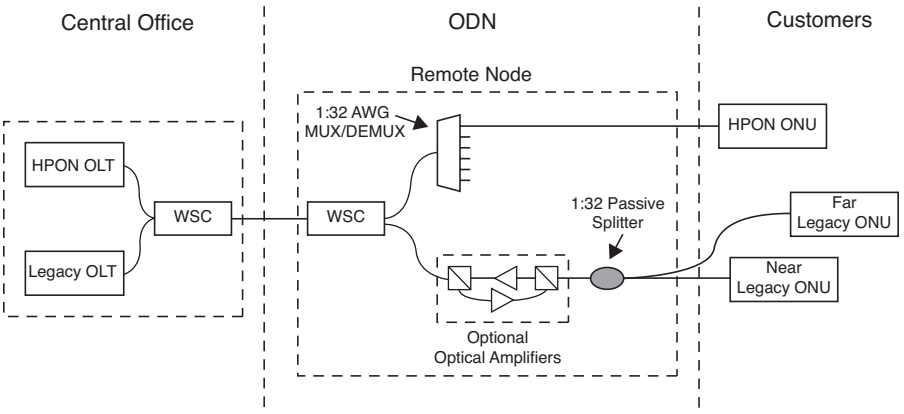


FIGURE 4.15 Fiber-lean tree upgrade with optional amplifiers.

New WDM-PON users are connected to the ports of the AWG. Since the two fused WCSs introduce an additional loss of about 1 dB, optical amplifiers *may* be needed for the TDM-PON service in the upstream direction. In the downstream direction, this 1-dB loss can easily be compensated with higher transmission powers from the OLT, without creating optical return loss (ORL) disturbances. Assuming a SMF loss of 0.35 dB/km for the 1310-nm wavelength, those TDM-PON ONUs that are approximately in the distance range 17 to 20 km from the CO may need upstream amplification. In such a case, the optional optical amplifiers indicated in Figure 4.15 are included; note that since it is impossible to introduce upstream amplifiers without creating some insertion loss in the downstream direction, the amplification needs to be bidirectional.

The preference for SOAs is that they are appropriate for the 1310- and 1490-nm basic bands and the fact that their transient effects are short and therefore good for burst-mode upstream transmissions (unlike EDFAs).

Alternatively, these distant TDM-PON ONUs can be upgraded to new WDM-PON ONUs, as shown in Figure 4.16, which might be less expensive than providing power for the optical amplifier at the RN (i.e., this decision would be an economic one). Note that the loss in the WDM-PON path is actually less than the one in the TDM-PON path, since the AWG loss is typically around 5 dB, while the passive splitter/couplers typically have a loss of at least 12, 15, or 18 dB for 16, 32, and 64 users, respectively.

As we have mentioned before, most of the current EPON and GPON deployments are converging toward the class B+ ODN (i.e., 13 to 28 dB, 20 km, with 32 users). In such a TDM-PON network, the 1 : 32 passive splitter accounts for at least 15 dB of the loss. The athermal AWG incurs a loss of 5 dB (instead of 15 dB, a difference of 10 dB) and the WCSs a loss of 0.5 dB each (total of 6 dB instead of 15 dB, a difference of 9 dB), the new power budgets in the same ODN are 3 to 18 dB for the fiber-rich scenario and 4 to 19 dB for the fiber-lean scenario. Note that this loss is in just one direction. This is important for the CLS approach, where the downstream carrier is modulated at the ONU and then sent upstream. In the CLS + RSOA approach, the RSOA amplifies the signal for the upstream direction, but introduces some noise due to the high noise figure of the device.

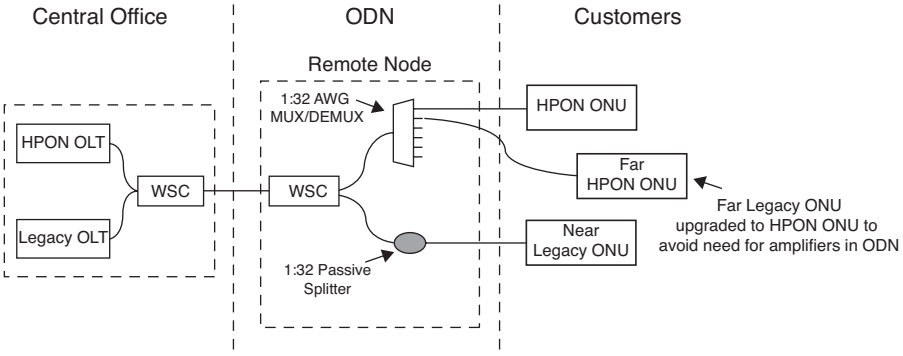


FIGURE 4.16 Fiber-rich tree upgrade without amplification.

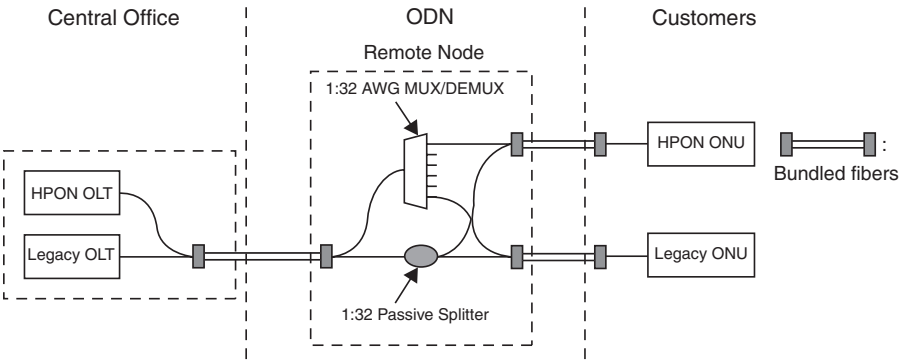


FIGURE 4.17 Fiber-rich tree upgrade.

4.3.2.3 Fiber-Rich Evolution In the fiber-rich scenario, an athermal AWG MUX/DEMUX is installed at the remote node (RN), collocated with the splitter, as shown in Figure 4.17. The MUX/DEMUX capabilities of this AWG would be the same as the ratio of the splitter (e.g., typically 1 : 32). On the network side, one extra fiber available between the CO and the RN is used, while on the users' side, each port of the AWG is connected to one extra fiber that reaches each customer. Thus, each customer now has two fiber paths to the CO: a legacy path for TDM-PON and a new path for WDM-PON. The addition of the athermal AWG at the RN is performed by a skilled technician. Depending on how the extra fiber to the home was terminated during the initial deployment, and the access that the user has to the ONU (in North America, for example, the ONU cabling can usually only be accessed by the operating company) installation of WDM-PON equipment in the home can be done by either a skilled technician or the user. The cost of using the extra fibers is quite low, since the deployment cost was assumed during the initial rollout. A green field WDM-PON ODN is also assumed to be minimally different (AWG instead of splitter) to a TDM-PON ODN and thus is not much more expensive.

4.3.2.4 SUCCESS-DWA SUCCESS-DWA PON is a novel optical access network architecture designed for the next-generation access networks, which employs dynamic wavelength allocation (DWA) [41]. As is shown later, the scalability of the SUCCESS-DWA PON allows the network to easily bridge the large gap between TDM and WDM-PONs. In addition, the architecture provides excellent cost efficiency and network performance by sharing bandwidth across multiple physical PONs. Existing arbitrary field-deployed PON infrastructures remain intact when brought together in a SUCCESS-DWA PON.

Downstream Transmission Figure 4.18 shows the architecture of SUCCESS-DWA PON. Fast tunable lasers (TLs), cyclic AWG, and thin-film WDM filters constitute the key components of the system. The TLs and AWG reside in the CO, while the WDM filters are within the ONU. Note that the field-deployed infrastructure is

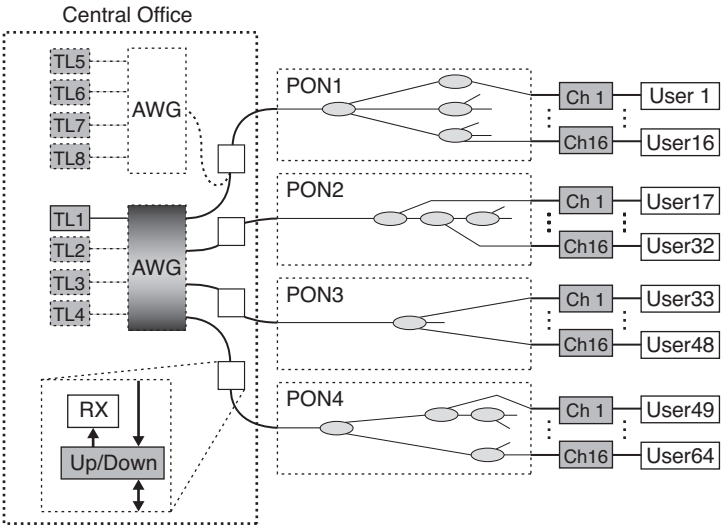


FIGURE 4.18 SUCCESS-DWA downstream architecture.

compatible with TDM-PONs. In the basic architecture, the cyclic AWG multiplexes the TLs and routes the TL outputs to different physical PONs depending on the wavelength. Each ONU within a single PON contains a unique fixed-wavelength filter and a burst-mode receiver. The key lies in the fact that the passband of the ONU filter encompasses the free spectral range of the AWG. For example, 200-GHz ONU filters would work with a 4×4 cyclic 50-GHz AWG. The relative filter shapes are illustrated in Figure 4.19. In this architecture, any tunable laser can individually address any ONU across separate physical PONs at any given time [41].

For each downstream frame, the TLs tune to the appropriate wavelengths and transmit the data to the corresponding end users. The transmission durations for end users are globally managed to achieve optimal network performance. The traffic scheduling allows free moving of bandwidth among physical PONs, which is very useful to accommodate bursty Internet traffic. All TLs share the load, shifting bandwidth across the separate physical PONs as necessary. The authors refer to this technique as dynamic wavelength allocation (DWA).

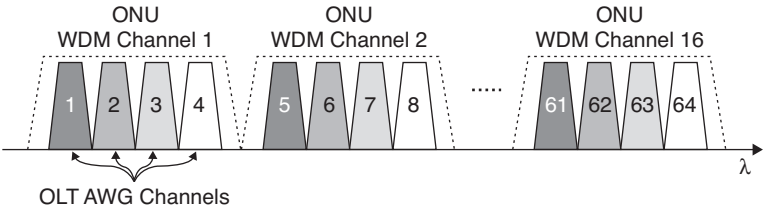


FIGURE 4.19 Wavelength bands for AWG channels and thin-film WDM filter.

To illustrate the flexibility of this architecture, compare an initial deployment of four TDM-PONs to a SUCCESS-DWA PON that spans four physical PONs. In the initial deployment stage, the first several subscribers would probably scatter across multiple PONs, and in the worst case they may be spread across all four PONs. In the four TDM-PON cases, then, all four OLTs (lasers) must be activated, despite the fact that some OLTs may only be serving a few subscribers. With the SUCCESS-DWA PON, on the other hand, it is sufficient to deploy only one TL and AWG in the CO, and the subscribers across multiple PONs are all serviced by the single TL. As demand grows, additional TLs can be added to the AWG. When the subscription rate is high enough, the two scenarios seem to converge; both have four transmitters serving all subscribers. However, the SUCCESS-DWA PON enjoys the benefit of statistical multiplexing over a larger customer base, so its performance will exceed that of the four TDM-PONs. For the performance of statistical multiplexing gain, readers are referred to a paper by Hsueh et al. [42].

When demand dictates, the SUCCESS-DWA PON can be scaled beyond the conventional TDM-PONs. The darkly shaded AWG in Figure 4.18 illustrates the concept. By connecting some of the physical PONs to additional AWG + TLs sets, the architecture shifts from four TLs serving four PONs to four TLs each serving a single PON. If 8×8 AWGs are utilized, 16 end users can be served by eight TLs, which results in a very high-performance network, close to a full WDM-PON. The excellent scalability provides a smooth and graceful upgrade from a TDM-PON toward a WDM-PON. The step-by-step system upgrade easily tracks user demands, and the initial overhead can be even lower than that of conventional TDM-PONs. In addition, the TLs provide protection for each other, maintaining service to all physical PONs in the event of a failure.

Upstream Transmission The measured upstream traffic rates are strongly related to the downstream traffic rates in access networks. Users enjoying high-speed downloads also expect a commensurate data rate in the upstream. This suggests that a high-performance upstream scheme is desirable. Depending on the performance requirements, there is a wide range of possible scenarios for upstream transmission. Hsueh et al. [41] investigated five upstream schemes for SUCCESS-DWA PON and evaluated them in terms of cost, performance, and scalability, as summarized in Figure 4.20.

For scheme A, users within a group transmit on the same upstream wavelength and time-share the same PD in the OLT. It is static in wavelength allocation and provides a baseline for comparison with the other schemes. Schemes B and C are distributed DWA schemes in which the tunability is spread across the ONUs. The distributed schemes allow more flexibility since the deployment of tunable devices can be judged by user demands. On the other hand, schemes D and E are centralized DWA schemes in which the tunability resides in the OLT. The tunable DeMux is described in the literature [42]. Qualitatively, full tunability results in the best performance. Scheme E exhibits full tunability but at the cost of an expensive, centralized tunable DeMux. Scheme C can provide equal performance if all users are equipped with TLs. Whereas the centralized schemes subject all ONUs to the high cost of the tunable device, the

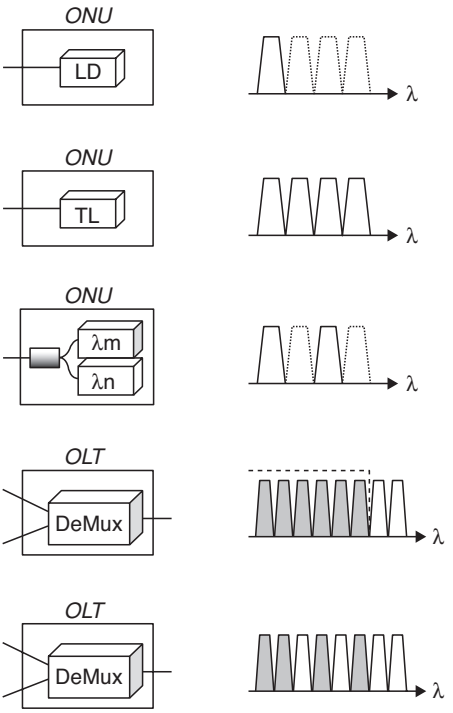


FIGURE 4.20 Five upstream schemes for SUCCESS-DWA.

distributed schemes require only those ONUs that demand high performance to be upgraded. Therefore, the added design flexibility of distributed schemes B and C makes them preferable.

A complete upstream SUCCESS-DWA PON architecture is shown in Figure 4.21. Four physical PONs are connected to an OLT, and an AWG functions as the DeMux for routing the incoming wavelengths to the corresponding PDs. It is worth noting that the AWG does not require cyclic features, and the upstream and downstream AWGs pass completely different wavelengths and require different channel spacings. Hence, separate AWGs are necessary for up/downstream. Similar to the downstream, only one PD + receiver module is activated in the initial deployment. To cover all ONUs residing on different physical PONs, the first several subscribers from different PONs are assigned specific upstream wavelengths (i.e., the first several subscribers from PON1 are assigned to λ_1, the first several subscribers from PON2 are assigned λ_2, and so on). When the number of users increases, a second PD + receiver module can be installed in the OLT. For new subscribers to be served by the second PD, they are assigned λ_2 on PON1, λ_1 on PON2, and so on. Similarly, the upstream architecture can scale up as the downstream. When demand dictates, additional AWG + PD/RX sets can be added (illustrated by the AWG in Figure 4.21) to shift from four PDs serving four PONs to four PDs serving each PON. For the bandwidth-demanding users,

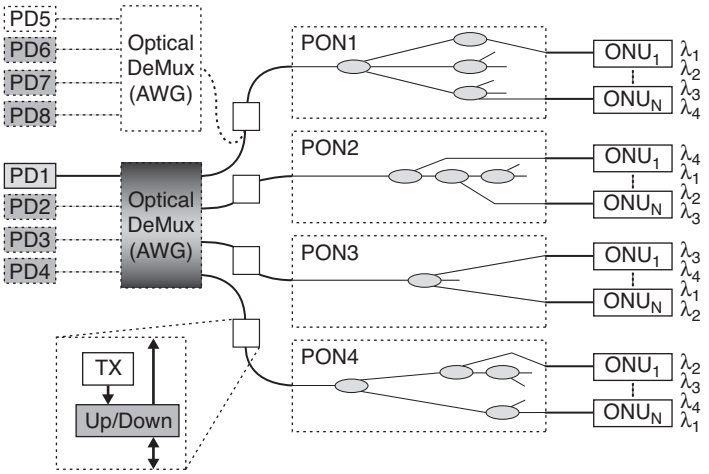


FIGURE 4.21 SUCCESS-DWA upstream architecture for users across multiple PONs.

tunable devices can replace the fixed-wavelength transmitters, as in schemes B and C. However, unlike the downstream architecture, which requires fast TLs (10 ns tuning times) to achieve DWA on the order of microseconds, the upstream DWA schemes perform slower wavelength reallocations (on the order of milliseconds) due to the more involved communications between the OLT and ONUs for transmission scheduling.

Experimental Testbed Hsueh et al. [41] described the details of design and implementation issues of the key building blocks of SUCCESS-DWA PON, including fast tunable lasers, burst-mode receivers, and scheduling algorithms with QoS support. A testbed is constructed and video streaming experiments in the downstream are performed to show the feasibility of the integrated system. As shown in Figure 4.22,

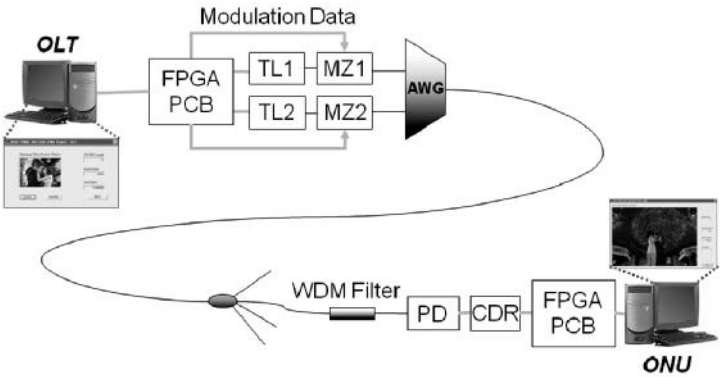


FIGURE 4.22 SUCCESS-DWA experimental testbed.

the experimental testbed consists of two fast TLs coupled by an AWG. Each TL is modulated externally at 1.25 Gb/s by a Mach–Zehnder modulator. Four user wavelength channels are defined, and 16 subscribers share bandwidth in both the wavelength and time domains. The WDM filter in the OLT separates the downstream and upstream traffic occupying different wavelength bands. The WDM filter in the ONU not only separates the up/downstream traffic but also provides the necessary filtering corresponding to the ONU's user wavelength channel. Burst-mode level recovery and CDR are implemented in both the OLT and ONUs. In the ONU, a laser driven by a dc-coupled amplifier realizes a burst-mode transmitter. Processors are necessary to perform the framing procedure, line coding, and scheduling algorithm. We designed PCBs with field-programmable gate arrays (FPGAs) as processors in the OLT and ONUs. The average packet latency and jitter of a realistic SUCCESS-DWA PON system are also measured on the testbed. For further details, readers are referred to the literature [41].

4.3.3 Tree to Ring Topology Evolution

All currently deployed TDM-PON networks are based on a tree topology, with a feeder fiber, a splitter at the remote node, and distribution fibers to each customer. In the long term, the access topology can go from multiple trees to a ring topology. This has been considered in developed markets to provide protection and restoration. In emerging markets, it has also been considered since ring topologies are sometimes a more cost-efficient way to provide service to many users with less infrastructure. The feeder fiber in this case can be shared among many more users than in a tree topology. In developed markets, even though this would require a higher investment in deploying the extra fiber required for the ring, this topology allows for greater resource sharing and alternative paths for protection and restoration.

There are two ways to achieve protection and restoration: (1) a particular high-end business customer can be connected to two remote nodes (i.e., be multihomed) and switch from one to the other in case of a fiber cut; (2) semipassive RNs with switches that sense traffic timeouts can shift from one side of the ring to the other after a certain timeout period, similar to what happens in other well-known ring networks, such as SONET.

Given these advantages, some researchers have proposed the use of ring topologies even for initial PON deployments in the greenfield [43,44].

4.3.3.1 SUCCESS-HPON Figure 4.23 illustrates the SUCCESS-HPON network migration scenario from a tree topology to a ring topology. Figure 4.23(a) shows the existing tree TDM-PONs connected to the same CO. Each TDM-PON has its own cabling and OLT inside CO. Figure 4.23(b) shows the first migration step of the existing network infrastructure. The passive couplers of the PONs are replaced with CWDM RNs. The feeder fibers of PONs are replaced by a single fiber ring that strings the RNs served by this CO. Note that distribution fibers are untouched during this migration. From an ONU's point of view, the functionality of the optical access network is exactly the same; only a short downtime for upgrade is needed.

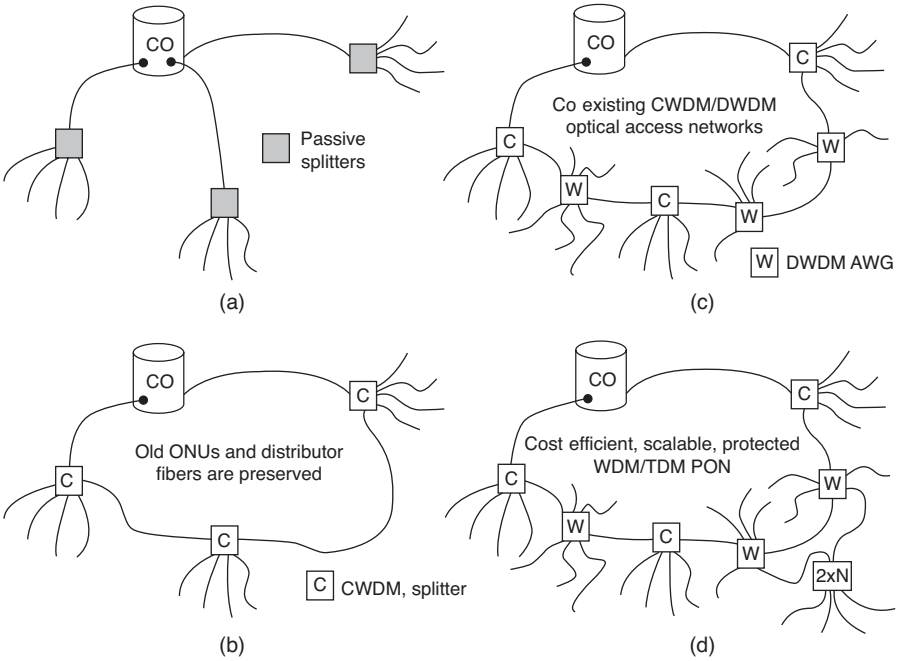


FIGURE 4.23 SUCCESS-HPON migration from tree TDM topology to ring WDM topology.

Therefore, existing TDM-PON ONUs can virtually work the same as before without any upgrade.

Figure 4.23(c) describes the second phase of migration. As more users demand highbandwidth for future broadband applications, DWDM RNs are inserted in the network. In this case, there is a dedicated DWDM channel between each ONU and the OLT. If protection and restoration functionality is implemented in the existing RNs using semipassive switches, inserting a new RN in the network will not disturb the network operation in general. Figure 4.23(d) shows the possible extension of the network. Since there is a dedicated wavelength at the output of the AWG, it is possible to use the collector ring as a backhaul for the PON with tree topology. The two feeder fibers of the PON can connect to different RNs to form a protection path. To upgrade the capacity even further, the RSOA modulator can be replaced by a stabilized laser source to perform full-duplex operation.

Figure 4.24(a) illustrates the RN design for the CWDM RNs and Figure 4.24(b) that of the DWDM RNs. A DWDM RN has a pair of CWDM band splitters to add and drop wavelengths for upstream and downstream transmissions, respectively. A DWDM RN has one CWDM band splitter, adding and dropping a group of DWDM wavelengths within a CWDM grid, and a DWDM MUX/DEMUX device such as an arrayed waveguide grating (AWG). Each WDM ONU has its own dedicated

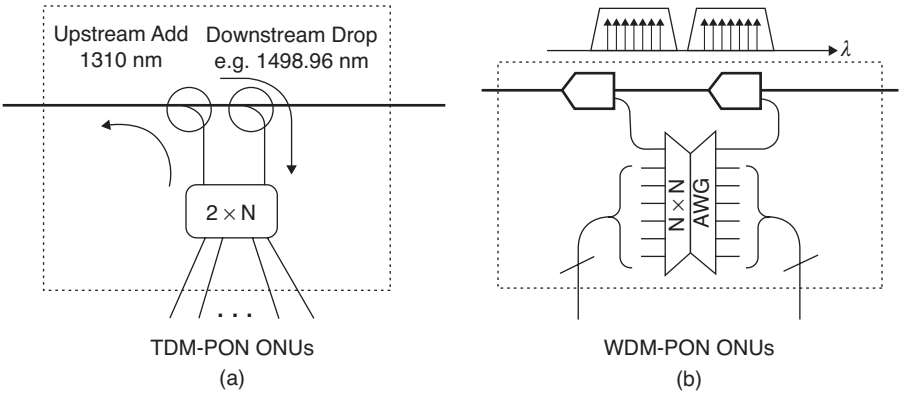


FIGURE 4.24 SUCCESS-HPON remote nodes for ring topology: (a) CWDM RN and (b) DWDM RN.

wavelength for both upstream and downstream transmissions on a DWDM grid to communicate with the OLT. Since the insertion loss of an AWG is roughly 6 dB regardless of the number of ports, one with more than eight ports can be used to enjoy a better power budget than that of a passive splitter. Each RN generally links 16 to 64 WDM-PON ONUs.

The overall resulting architecture of SUCESS-HPON, including TDM-PONs and WDM-PONs as its subnetworks, is shown in Figure 4.25. A single-fiber collector ring with stars attached formulates the basic topology. The collector ring strings up the RNs, which are the centers of the stars. The ONUs attached to the RN on the west side of the ring talk and listen to the transceiver on the west side of the OLT, and likewise for the ONU attached to the RNs on the east side of the ring. There is a point-to-point WDM connection between the OLT and each RN. No wavelength is reused on the collector ring. If extra reliability is desired, optional semipassive RNs may be used to sense fiber cuts and flip the orientation.

For more details than provided here on the SUCCESS-HPON architecture, readers are referred to the literature [45–49].

OLT Figure 4.26 shows a logical block diagram for the SUCCESS-HPON OLT. Tunable components such as fast tunable lasers and tunable filters are employed. Since the average load of access networks is generally low, using tunable components minimizes the transceiver count and thus minimizes the total system cost. This arrangement also has excellent scalability: As more users join the network, or their traffic increases, more tunable lasers and receivers are added at the OLT. Upstream optical signals are separated from the downstream signals by circulators. The scheduler controls the operation of both tunable transmitters and tunable receivers. Note that the tunable transmitters at the OLT generate both downstream frames and CW optical bursts to be modulated by ONU for upstream data. With this configuration, half duplex communication is possible at the physical layer between each ONU and

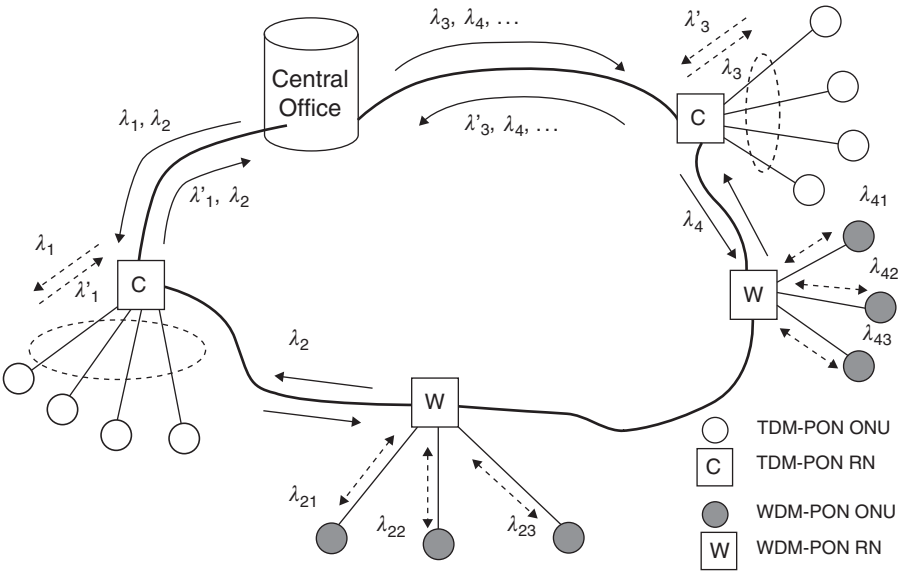


FIGURE 4.25 Ring topology SUCCESS-HPON.

the OLT. Compared to a similar architecture [30] with a two-fiber ring, two sets of light sources, and two sets of MUX/DEMUX to perform full-duplex communications, the SUCCESS-HPON architecture lowers costs dramatically. As a trade-off, it needs carefully designed MAC protocol and scheduling algorithms to provide efficient bidirectional communication. The authors have developed both, and they are discussed in the next section.

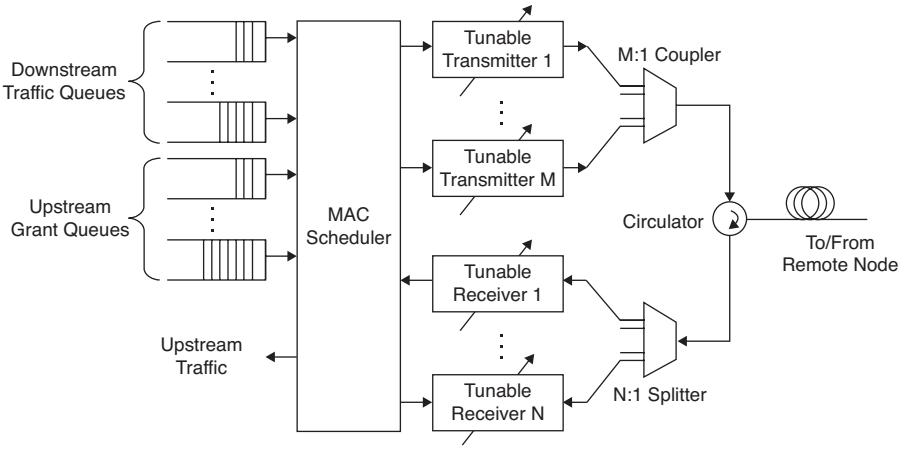


FIGURE 4.26 SUCCESS-HPON OLT.

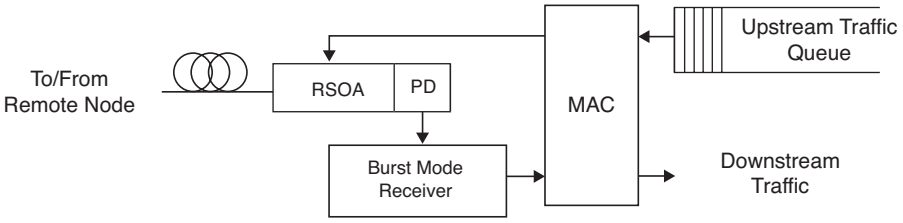


FIGURE 4.27 SUCCESS-HPON colorless ONU.

Colorless ONU Figure 4.27 illustrates the logical block diagram for the WDM-PON ONU. The ONU has no local optical source and uses instead an optical modulator to modulate optical CW burst received from the OLT for its upstream transmission. An RSOA can be used as an amplifier/modulator for this purpose with the assumption that its integration with electronics would decrease its production costs when mass-produced. Note that the ONU does not need a tunable receiver. The WDM-PON RN removes extraneous wavelengths and allows only a specific wavelength to reach each WDM-PON ONU. The receiver at the ONU just needs to have enough optical spectral bandwidth to receive any DWDM channel used in the network. Note as well that the MAC block in the ONU not only controls the switching between upstream and downstream transmissions but also coordinates with the scheduler at the OLT through polling and reporting mechanisms.

Experimental Testbed A SUCCESS-HPON testbed was implemented to (1) demonstrate the feasibility of the bidirectional transmission of upstream and downstream traffic on the same wavelength for access networks, (2) demonstrate the possibility of modulating upstream data onto CWs provided by the OLTs, (3) demonstrate the functionality of the MAC protocol and scheduling algorithms, and (4) explore possible SUCCESS-HPON implementation issues.

The testbed diagram is shown in Figure 4.28. The key components used in this testbed are:

- *OLT transmitting end:* two tunable lasers (Agility 4245), two Mach–Zehnder modulators (MZ, SDL 2.5 Gb/s), one pattern generator (HP70843A), one programmable board (not shown in the diagram, developed in-house with an Altera APEX 20K FPGA), one coupler, and one circulator.
- *OLT receiving end:* one circulator, one coupler, one erbium-doped fiber amplifier (EDFA, developed in-house), one optical bandpass filter, one photodiode (HP11982A), and one oscilloscope (HP54120).
- *Distribution network:* standard single-mode fiber (SMF), two AWGs (Lucent X1450F), and four thin-film add/drop WDM filters.
- *Each ONU:* one circulator, one coupler, one photodiode, one SOA (Genoa LOA), one MZ modulator (as above), and one programmable board (not shown, as above).

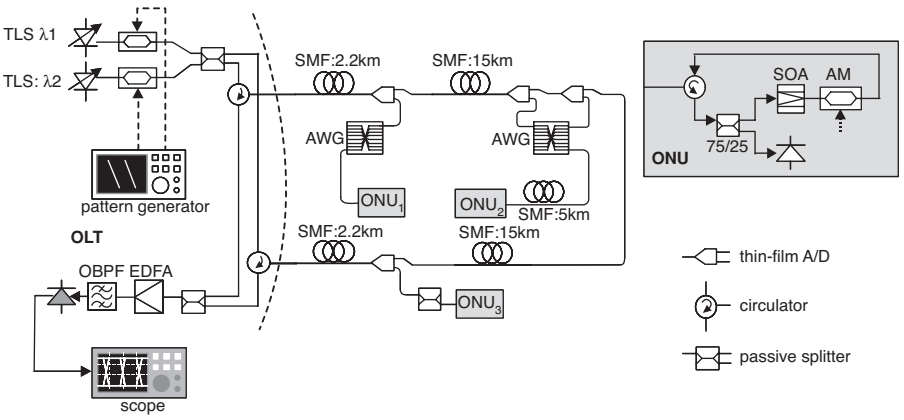


FIGURE 4.28 Testbed diagram.

At the OLT transmitting end, two tunable lasers use different wavelengths to communicate with ONU_1 and ONU_2; these wavelengths are determined by the distribution network components as explained below. The pattern generator and the programmable board control the MZ modulators to generate downstream data at a 1.25-Gb/s rate and/or CW bursts. Downstream traffic and CW bursts pass through ports 1 and 2 of the circulators to enter the ring.

The collector ring is composed of four standard single-mode fiber (SMF) sections of 2.2, 15, 15, and 2.2 km. Each remote node has at least one thin-film CWDM add/drop filter, and two of them have an AWG with a channel spacing of 100 GHz. Since no athermal AWGs were available at the time, we used conventional temperature-stabilized AWGs. The combined wavelength characteristics of the add/drop filters and the AWGs at each remote node determine the wavelength assigned to a particular WDM-PON ONU and used by the OLT to communicate with it. A single wavelength for each ONU, as mentioned before, is used for both downstream and upstream communication. The total distances from the OLT to the ONUs are 2.2, 22.2, and 2.2 km, respectively.

Each ONU connects to its fiber through port 2 of its own circulator. Downstream data and CWs are received through port 3, while upstream data (that have been modulated on the CW) is sent through port 1. An optical power splitter is used: 25% goes to the receiver and 75% goes to the SOA and MZ. The power budget for the downstream signal is enough to achieve error-free transmission. The splitter allocates more power for the upstream transmission since it needs to travel back to the OLT. Since the electronic driver circuitry to modulate the SOA is currently unavailable and the SOA is not designed for fast switching, the SOA only amplifies the CW in our testbed. The modulation is then performed by an external MZ modulator that has a 6-dB loss. This MZ modulator again is controlled by a programmable board, different from the one at the OLT. When the ONU is receiving downstream data, the modulator is turned off to prevent upstream interference.

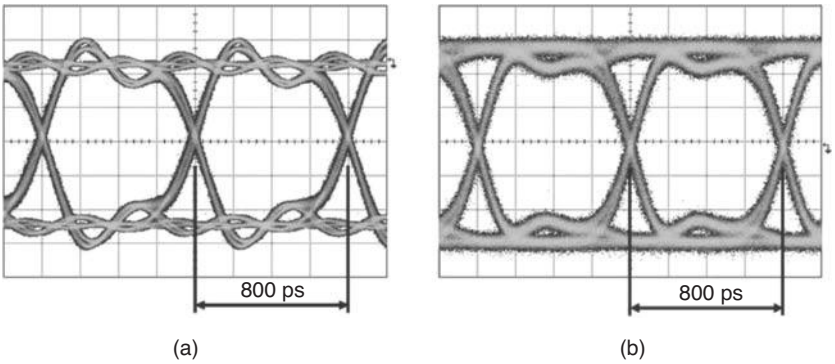


FIGURE 4.29 Downstream (a) and upstream (b) transmission eye diagrams.

Upstream traffic enters the receiving end of the OLT through ports 2 and 3 of the circulators. Note that this traffic comes from CWs sent by the OLT transmitting end, which were then amplified by the ONUs, modulated, and sent back. The receiving end at the OLT has an erbium-doped fiber amplifier (EDFA) as a preamplifier, an optical bandpass filter to remove the noise of the EDFA, and a photodiode.

We first send continuous data from the OLT to ONU_2 at a 1.25-Gb/s rate. The laser's output optical power is set at 5 dBm and a pseudorandom bit sequence (PRBS) with a $2^{23} - 1$ word length is used for MZ modulation. After leaving the OLT, the signal traverses 2.2 km of SMF, an add/drop filter, 15 km of SMF, a second add/drop filter, an AWG, 5 km of SMF, and reaches ONU_2. The total power loss from the OLT's laser to ONU_2 is approximately 20 dB. Figure 4.29(a) shows the downstream eye diagram for the data received at ONU_2. The ripples (distortion) of the eye diagram are due to the frequency response of the postdetection electrical filter.

The transmitter at the OLT is then set to generate a CW optical carrier for ONU_2, also at an optical power of 5 dBm. The CW reaches the ONU following the same path described above, with an optical power of approximately -15 dBm. The circulator, 75%/25% coupler, and MZ modulator at ONU_2 add an additional loss of approximately 7.5 dB, but the SOA amplifies the CW by 20 dB, making the final output power approximately -2.5 dBm. The MZ units modulate the data onto the CW, again using a $2^{23} - 1$ word PRBS sequence. The modulated data are sent back through the ONU's circulator and follow a similar upstream path to the OLT, where they reach the OLT's receiving end. The loss caused in the upstream direction reduces the received power level to approximately -22.5 dBm. Figure 4.29(b) shows an eye diagram for the upstream data. As can be seen, both parts of the figure show clear eye diagrams. The eye diagram for downstream data is clearer than the one for upstream data mostly because of the high noise figure of the SOA.

Now we demonstrate the possibility of modulating upstream data onto CWs provided by the OLTs. For this, we use two separate ONUs, which communicate to the OLT through different wavelengths. For both downstream and upstream data, we use a $2^{23} - 1$ word PRBS sequence.

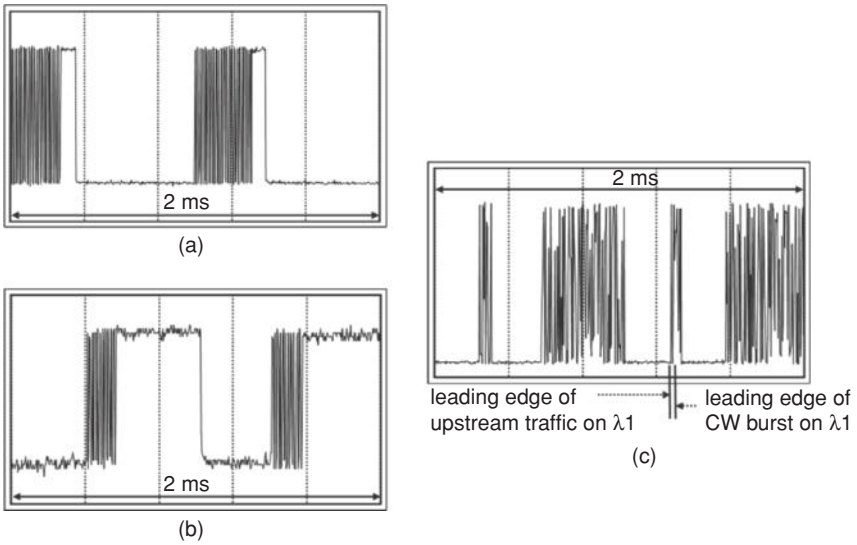


FIGURE 4.30 Downstream and upstream transmission timing diagrams.

The two lasers at the OLT transmit at the 1550.92-nm (λ_1) and 1550.12-nm (λ_2) wavelengths, assigned to ONU_1 and ONU_2 respectively. The MZ modulators at the OLT are controlled by a programmable board, generating downstream traffic and CW bursts piggybacked together within a l-ms frame, as shown in Figure 4.30(a) and (b).

The add/drop filters and AWG assignments at the RNs ensure that λ_1 is received by ONU_1 and λ_2 by ONU_2. The MZ modulators at the ONUs are again controlled by a programmable board which turns them off while receiving downstream traffic (or in the idle state) and turns them on when receiving the CW to modulate the data onto it.

The receiver at the OLT detects the upstream traffic on both λ_1 and λ_2. In this experiment, since the CWs in two different wavelengths do not overlap, there is no need for tunable receivers. Figure 4.30(c) shows the upstream traffic pattern retrieved by the oscilloscope at the OLT receiving end. Note that the timing of downstream and upstream traffic on λ_2 is aligned; however, there is a forward time shift of the upstream traffic on λ_1 compared with the downstream data on λ_1. The reason is that the distance between the OLT and ONU_1 is shorter; therefore, so is the corresponding round-trip time (RTT). This factor is taken into account by the MAC protocol.

4.4 WDM-PON PROTOCOLS AND SCHEDULING ALGORITHMS

Just as in TDM-PON, in some WDM-PON network architectures resources are shared. These resources can, for example, be particular wavelengths or tunable resources. Whenever there is a shared resource in data transmission, a media access

control protocol is required to allocate use of this resource to multiple parties. In this section we discuss some MAC protocols that have been developed for WDM-PON.

Another important issue is the scheduling of these resources according to some goal. For example, if the same tunable laser is used among many users, how should it be scheduled in a way that can provide service to all of them, and, if possible, in a fair manner. Or, if a single laser is used for multiple types of traffic, each in a separate queue, what scheduling algorithm should be used to guarantee higher priority to real-time traffic over data traffic. We also discuss some WDM-PON scheduling algorithms in this section.

4.4.1 MAC Protocols

Media access control (MAC) protocols are used to regulate the access of entities wishing to transmit data to the media that transmits them. For example, the well-known Ethernet standard uses carrier-sense multiple access with collision detection (CSMA/CD) to regulate access to the wire media.

4.4.1.1 SUCCESS-HPON MAC Protocol The SUCCESS-HPON architecture described in Section 4.3 is a good example of a WDM-PON network that requires a MAC protocol to share the transmission medium. The tunable lasers at the OLT need to allocate their time among competing ONUs. This needs to be done for both downstream and upstream traffic, since the ONUs transmit upstream data by modulating a continuous wave (CW) sent by the OLT. In turn,an ONU needs to know when it can modulate the CW grant that it has received from the OLT to transmit upstream data, and for how long.

Figure 4.31 illustrates the frame formats that are used. All frame formats have an 8-bit preamble and a 16-bit delimiter to allow for clock synchronization and to indicate the beginning of the frame, respectively. The uppermost frame format is used for downstream data transmission: a payload of multiple Ethernet frames. The next downstream frame is used to provide a continuous wave grant to the ONU. The 8-bit flags section of the downstream frame is used to indicate whether the downstream frame is for downstream data or is a grant to be used for upstream data. They also indicate whether the ONU should report its queued traffic in the upstream frame. This information is transmitted from the OLT to the ONU in the downstream frame that is sent for upstream transmission.

Ethernet frames or IP packets can be carried in the payload part of SUCCESS-HPON frames. Each ONU reports the amount of traffic waiting in its upstream traffic queue in octets in the 16-bit report field in an upstream frame. The OLT uses the grant field to indicate the actual size of each grant (also in octets).

Using the frames above, and similar to GPON and EPON systems [50], the SUCCESS-HPON OLT polls to check the amount of upstream traffic stored at the ONUs and sends grants followed by CWs to allow the ONUs to transmit upstream traffic. Since there is neither a separate control channel nor a control message embedding scheme using escape sequences as described by Kramer et al. [51], the MAC

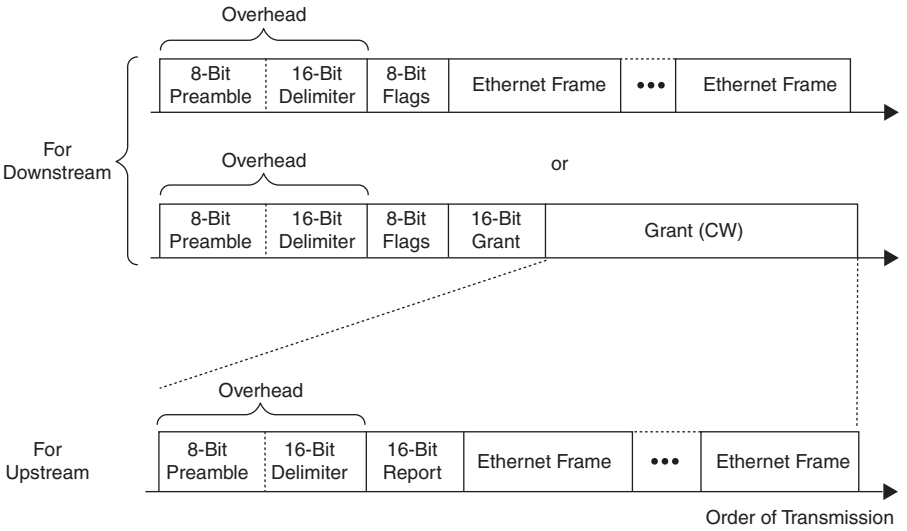


FIGURE 4.31 SUCCESS-HPON frame formats.

protocol has to rely on in-band signaling, where the report and grant fields are defined for the polling process. For further details on the specific values of the flags used in the frames, readers are referred to the literature [45,48].

4.4.2 Scheduling Algorithms

Scheduling algorithms in WDM-PON are used to schedule the use of shared resources (e.g., tunable lasers) among multiple competing users or among different types of traffic. The details of different scheduling algorithms (e.g., how they calculate when to take turns, how they package and unpackage the data payload) determines the performance and convenience of an algorithm. A good scheduling algorithm can dramatically minimize the number of shared resources that are needed to serve a given number of end users. In this section we present some of the scheduling algorithms that have been developed for WDM-PON architectures.

4.4.2.1 SUCCESS-HPON Resource Scheduling The SUCCESS-HPON architecture requires scheduling algorithms that are able to assign all the resources available, including tunable lasers, tunable receivers, and wavelengths, to avoid conflicts among them and to provide efficient bidirectional data communication between the OLT and the ONUs. This discussion is based on results presented in the literature [45,47,48].

The scheduling algorithms and MAC protocols for SUCCESS-HPON are based on intelligent scheduling of tunable components over multiple wavelengths [52]. Like current GPON and EPON systems [51], the OLT polls the ONUs to check the

amount of upstream data queued stored inside them and sends grants, along with the corresponding CWs, to allow the ONUs to transmit their upstream traffic. The SUCCESS-HPON protocol, however, is distinctive from previous protocols in that:

- It uses the TCM technique [53] (also known as "Ping-Pong") to provide half-duplex bidirectional transmission between the OLT and the ONUs.
- There is no separate control channel or frames.
- There is no embedding of control messages with escape sequences like those proposed by Jia et al. [52].

To schedule all the resources in SUCCESS-HPON efficiently, we have developed novel algorithms that take into account the demands and allocate the resources in time, just like DBA algorithms do in TDM-PONs. In this section we discuss these algorithms.

We consider a SUCCESS-HPON system with W wavelengths (corresponding to W ONUs), M tunable lasers, and N tunable receivers. Note that because the tunable lasers are shared by both upstream and downstream traffic whereas tunable receivers are not, generally $W \geq M \geq N$. This is also a consequence of the fact that usually there will be more downstream than upstream traffic. A guard band between successive frames of G nanoseconds takes into account not only the tuning time of lasers and receivers at the OLT, but also the effect of unstable local ONU clock frequencies. Each ONU_i has a round-trip time RTT_i. Since we do not equalize end-to-end delays among ONUs, the round-trip times usually differ from one another. Two more parameters are of importance: *ONU_TIMEOUT*, a timer per ONU, reset whenever a grant frame is sent downstream to the corresponding ONU, and *MAX_GRANT*, a limit on the maximum size of a grant (i.e., the payload part of the CW burst) for ONU upstream traffic.

The OLT manages and stores the following MAC protocol global status variables for resource scheduling purposes:

- CAT: array of channel available times. $\text{CAT}[i] = t$, where $i = 1, 2, \ldots, W$, means that the wavelength λ_i will be available for transmission after time t.
- TAT: array of transmitter available times. $\text{TAT}[i] = t$, where $i = 1, 2, \ldots, M$, means that the ith tunable transmitter will be available for transmission after time t.
- RAT: array of receiver available times. $\text{RAT}[i] = t$, where $i = 1, 2, \ldots, N$, means that the ith tunable receiver will be available for reception after time t.
- RTT: array of round-trip times (RTTs) between the OLT and the ONUs. $\text{RTT}[i]$ denotes the RTT between the OLT and the ith ONU.

The scheduling algorithms developed make use of all the parameters discussed above to schedule the resources. We describe them briefly here; for a detailed

explanation of each, readers are referred to the literature [45,47,48]. The three algorithms are:

- *Algorithm 1: Sequential scheduling*. In this algorithm, as requests for downstream and upstream data are received, the next available tunable transmitter, receiver, and channel are scheduled and the request is addressed as soon as possible.
- *Algorithm 2: Batch earlier departure first* (BEDF). In this algorithm the idea of batching a small queue of requests and addressing them as a group is used as shown in Figure 4.32. The major concern is to provide room for scheduling optimization by forming a task set consisting of multiple frames by a batching process.
- *Algorithm 3: Sequential scheduling with schedule-time framing* (S^3F). This is an improved sequential scheduling algorithm; it can encapsulate multiple payload packets in the same transmission. The scheduling is done at the end of each frame transmission. For the purpose of granting downstream traffic, we maintain a downstream transmission counter per downstream VOQ. When granting upstream traffic based on a request received from an ONU, we also grant downstream traffic as well based on the VOQ status at the time of the arrival of the report message. Granting downstream traffic is done by setting the said grant counter to the minimum of the queue length of the VOQ and *MAX_GRANT*. When scheduling downstream transmission, the grant counter value controls the number of Ethernet frames to be scheduled and transmitted in one SUCCESS frame through the procedure shown in the algorithm below.

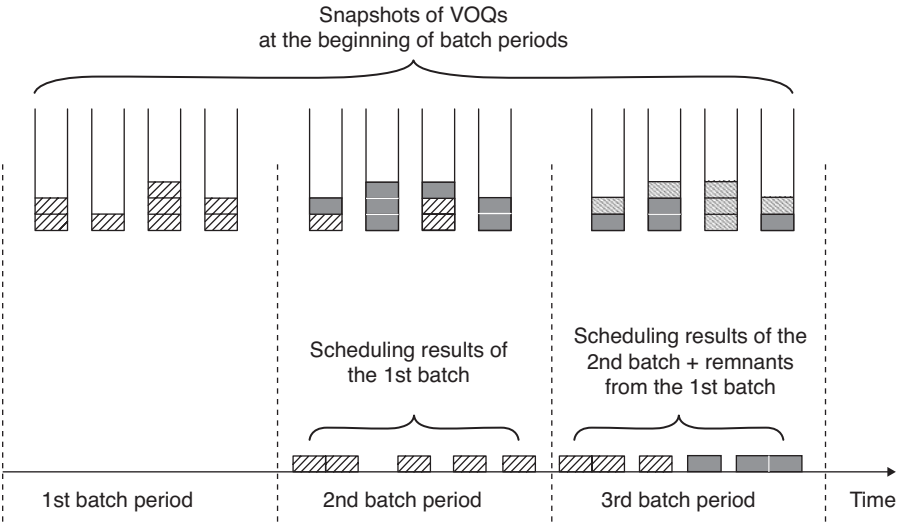


FIGURE 4.32 Scheduling example showing the interleaving of scheduling and transmission phases in BEDF.

The sequential scheduling algorithm has two main problems. First, downstream traffic is given an implicit priority over upstream traffic; when the network is heavily loaded, downstream traffic goes through whereas upstream traffic does not. The second problem is that the sequential scheduling algorithm saturates itself at loads of about 35% of the maximum capacity.

In S^3F, the benefit of granting and schedule-time framing of downstream traffic is three-fold. First, by encapsulating multiple Ethernet frames in one SUCCESS-HPON frame as in upstream transmission, the overhead due to the SUCCESS-HPON framing and the guard bands can be reduced. Second, the waste of tunable transmitters and channels can also be minimized and therefore minimize scheduling delays by preventing the spread of smaller frames over multiple transmitters and channels. This is illustrated in the examples in Figures 4.33 and 4.34, where it is assumed that for both framing schemes, there are three Ethernet frames at t_1 in the VOQ for channel 1, four Ethernet frames at t_2 for channel 2, and one Ethernet frame at t_3 for channel 3. t_i' denotes the resulting scheduled transmission time of the first frame for channel i, so the corresponding scheduling delay is given by $t_i' - t_i$. There is some inefficient use of transmitters in the arrival-time framing, where each incoming Ethernet frame is encapsulated in a SUCCESS-HPON frame at the moment of its arrival. The impact of this inefficiency will become clear when the scheduling delay of the frame for channel 3 in Figure 4.33 is compared to that of the schedule-time

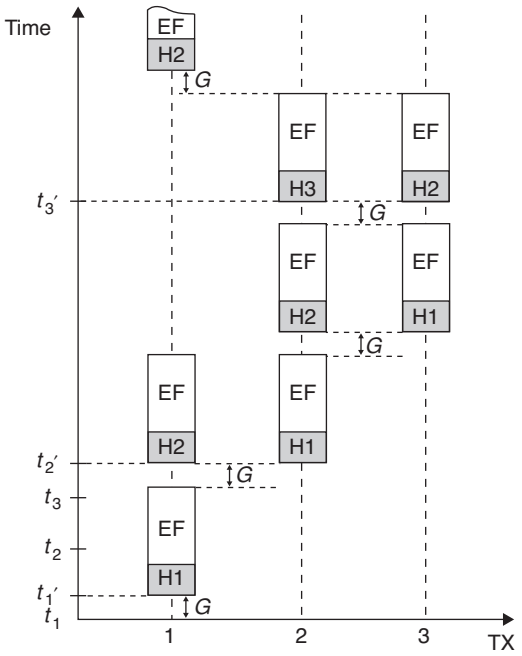


FIGURE 4.33 Arrival-time framing used in the sequential algorithm.

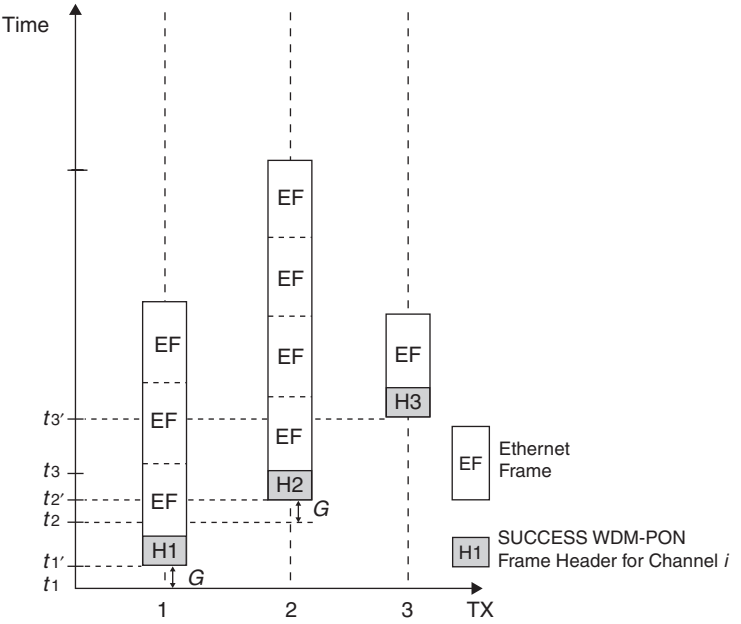


FIGURE 4.34 Schedule-time framing used in the S^3F algorithm.

framing in Figure 4.34. Third, by integrated and intelligent granting of both upstream and downstream traffic, the traffic flows can be better controlled to guarantee fairness and to support QoS in the future.

The authors have developed a simulation model for the performance evaluation of the algorithms mentioned using an objective modular network testbed in C++ (OMNeT++) [54]. These algorithms can provide efficient and fair bidirectional communications between the OLT and the ONUs in the SUCCESS-HPON architecture.

A summary of the results discussed in the literature [45,47,48] is shown in Figures 4.35, 4.36 and 4.37. In Figure 4.35, we can see the throughput for a given network load. The network load has been normalized; 1.0 would mean that the incoming traffic rate is the same as that of the transmitters, and 0.5 would mean that the incoming traffic rate is half the rate of the transmitters. As soon as the throughput line deviates from a 45° angle, the network is being saturated. As we can see in the figure, algorithm 1 saturates at around 30%, algorithm 2 around 85%, and algorithm 3 around 90%. It is clear from this how the design of such algorithms can seriously affect the performance of the network.

Figure 4.36 also shows that the algorithms are progressively better. The graph illustrates the delay experienced by an incoming packet, in miliseconds, given a network load. Whenever the delay grows vertically, the network is considered to be saturated and the packet delays become intolerable. Algorithms 3 saturates, creating considerable delay, only at normalized network loads of around 0.9.

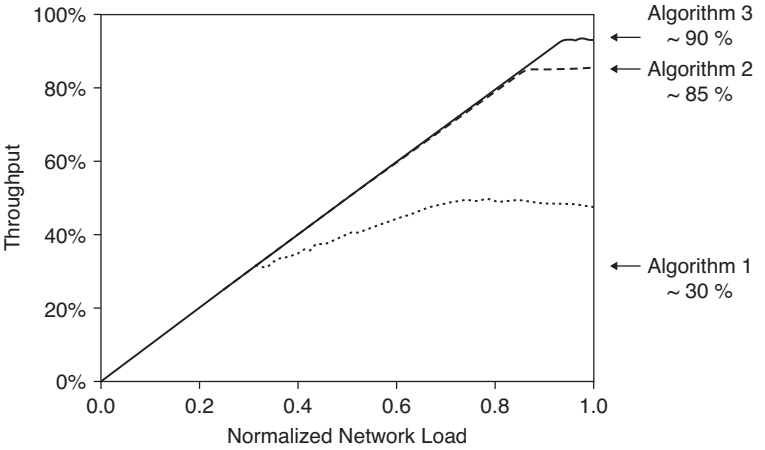


FIGURE 4.35 Scheduling algorithms throughput.

Figure 4.37 illustrates the scalability of the SUCCESS-HPON approach, in which shared tunable transmitters are used. In comparison to GPON or EPON networks at 2.5 or 10 Gb/s, where the maximum incoming traffic rate will not go beyond these values, the shared tunable transmitters approach can incrementally increase the number of such transmitters as the load on the network increases. That is, as more and more users sign up to use the network, more tunable resources can be installed to satisfy higher demands. In the example in Figure 4.37, with two tunable transmitters at 10 Gb/s each, an overall network thoughput of almost 20 Gb/s can

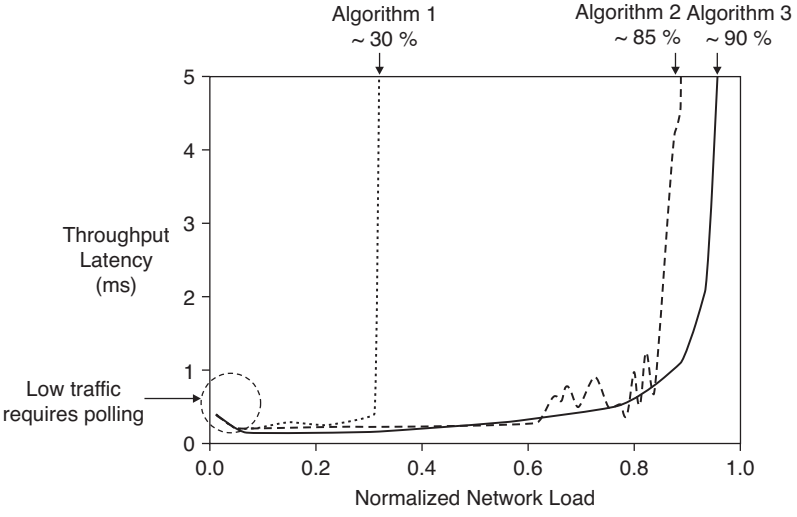


FIGURE 4.36 Scheduling algorithms delay.

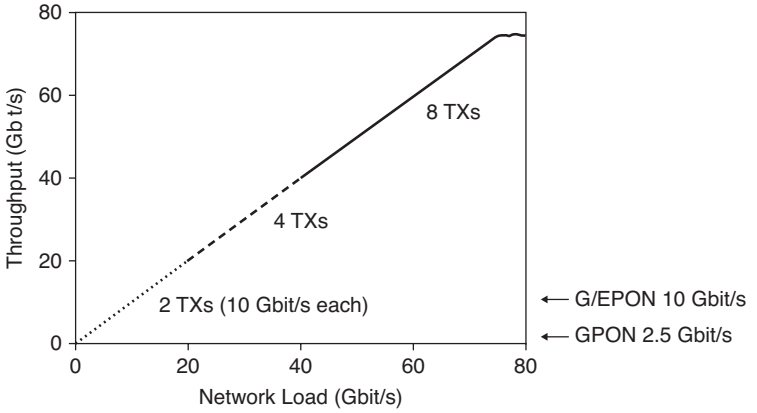


FIGURE 4.37 Sharing resources for scalability.

be supported; similarly for four tunable transmitters (almost 40 Gb/s) and for eight tunable transmitters (almost 80 Gb/s).

4.4.2.2 SUCCESS-DWA QoS Traffic Scheduling Multiple services may be flowing simultaneously in an access network. For example, voice, video, Web data, and e-mail may all be using the same distribution fibers at the same time. To ensure that higher-priority traffic such as real-time video is given priority over traffic that is not so time-sensitive, such as e-mail download, quality of service policies need to be applied and enforced. This policies determine how the priority is handled and in some cases the implementation as well. In SUCCESS-DWA, a buffering scheme was used to provide two different priorities to flowing traffic.

To support QoS on SUCCESS-DWA PON, strict-priority traffic scheduling algorithms were investigated [41]. The traffic is categorized into high-priority (HP) and best effort (BE) traffic. Specifically, to improve the scalability and reduce the complexity of the buffering strategy in an OLT, we propose a TL-buffering scheme for the OLT of the SUCCESS-DWA PON. The block diagram of the TL-buffering scheme is shown in Figure 4.38. The incoming traffic is divided and stored in

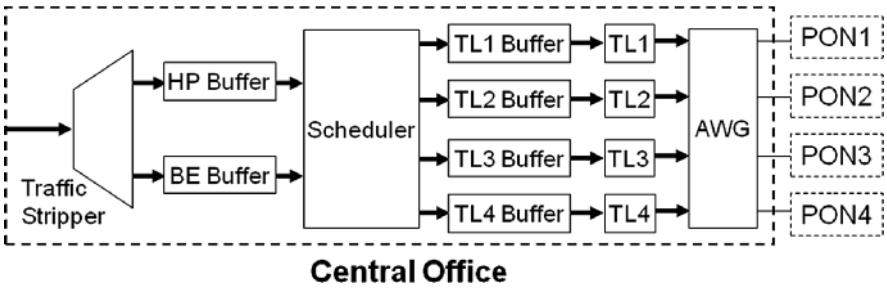


FIGURE 4.38 Block diagram of the TL-buffered scheme in an OLT.

separate HP and BE buffers before scheduling. Depending on the packet arrival time and priority and the availability of TLs, the HP traffic can be serviced at the earliest possible time and the prioritized traffic is highly differentiated. The performance of the TL-buffering scheme is evaluated in comparison with a conventional virtue-output-queuing (VOQ) scheme, and the results show that the average latency and jitter of HP traffic outperforms that of the VOQ buffering strategy [42].

4.5 SUMMARY

In this chapter we review the technologies and network architectures for the future development of TDM-PON. We present bandwidth enhancements that have been proposed for EPON and GPON and line rate enhancements that have been researched for TDM-PON in general. Extending the reach of TDM-PON is another important enhancement in which research and standardization efforts are ongoing.

The natural progression for TDM-PON is to move toward WDM-PON, where the enormous bandwidth of multiple wavelengths can be used. We discuss the most promising technologies and network architectures for WDM-PON, including spectrum-sliced broadband light sources, injection-locked Fabry–Perot lasers, and RSOAs in a centralized light sources network. Since tunable lasers may also be used as colorless ONUs, we discuss them as well.

Service providers would like to find a way to move progressively from TDM-PON deployment to WDM-PONs, using the same ODNs as those that deployed for TDM-PONs. We present the requirements for such an evolution and the possible evolution paths from TDM-PON to WDM-PON for fiber-rich and fiber-lean deployments. We discuss evolutionary paths for tree topology as well as possible migration paths from tree to ring topology to provide protection and restoration services.

WDM-PONs that make use of shared resources such as tunable transmitters and receivers require specialized MAC protocols and scheduling algorithms. We review some of these protocols and algorithms, which allow these networks to grow in a scalable manner as demand increases.

REFERENCES

1. D. Gutierrez, K. S. Kim, S. Rotolo, F.-T. An, and L. G. Kazovsky, FTTH standards, deployments and research issues (invited paper), presented at the Joint Conference on Information Sciences 2005, Salt Lake City, UT, July 2005.
2. N. K. Shankaranaryanan, Z. Jiang, and P. Mishra, User-perceived performance of Web browsing and interactive data in HFC cable access networks, *Proceedings of the IEEE International Conference on Communications*, Helsinki, Finland, June 2001.
3. L. G. Kazovsky, Burst-mode metro and access networks (invited paper), presented at the Optical Fiber Communication Conference OFC 2007, Anaheim, CA, Mar. 2007.
4. Y.-L. Hsueh, M. S. Rogge, W.-T. Shaw, J. Kim, S. Yamamoto, and L. G. Kazovsky, Smooth upgrade of existing passive optical networks with spectral-shaping line-coding service overlay, *IEEE/OSA J. Lightwave Technol.*, vol. 23, no. 8, Sept. 2005, pp. 2629–2637.

5. M. Abrams, P. C. Becker, Y. Fujimoto, V. O'Byrne, and D. Piehler, FTTP deployments in the United States and Japan: equipment choices and service provider imperatives, *IEEE/OSA J. Lightwave Technol.*, vol. 23, no. 1, Jan. 2005, pp. 236–246.
6. Calix press release, Calix to host full service access network group fall meeting, Sept. 2006.
7. G. Maier, M. Martinelli, A. Pattavina, and E. Salvadori, Design and cost performance of the multistage WDM-PON access networks, *IEEE/OSA J. Lightwave Technol.*, vol. 18, no. 2, Feb. 2000, pp. 125–143.
8. C. Arellano, C. Bock, J. Prat, and K.-D. Langer, RSOA-based optical network units for WDM-PON, presented at the Optical Fiber Communication Conference OFC 2006, Anaheim, CA, Mar. 2006.
9. Y. J. Lee, K. Y. Cho, A. Murakami, A. Agata, Y. Takushima, and Y. C. Chung, Reflection tolerance of RSOA-based WDM PON, presented at the Optical Fiber Communication/ National Fiber Optic Engineers Conference, Feb. 2008.
10. B. Schrenk, J. A. Lazaro, and J. Prat, Employing feed-forward downstream cancellation in optical network units for 2.5G/1.25G RSOA-based and 10G/10G REAM-based passive optical networks for efficient wavelength reuse, presented at the International Conference on Transparent Optical Networks, June 2009.
11. P.-J. Rigole, S. Nilsson, L. Backbom, T. Klinga, J. Wallin, B. Stalnacke, E. Berglind, and B. Stoltz, Access to 20 evenly distributed wavelengths over 100 nm using only a single current tuning in a four-electrode monolithic semiconductor laser, *IEEE Photon. Technol. Lett.*, vol. 7, no. 11, Nov. 1995, pp. 1249–1251.
12. C. J. Chang-Hasnain, Tunable VCSEL, *IEEE J. Sel. Top. in Quantum Electron.*, vol. 6, no. 6, Nov.–Dec. 2000, pp. 978–997.
13. R. Chen, H. Chin, D. A. B. Miller, K. Ma, and J. S. Harris, Jr., MSM-based integrated CMOS wavelength tunable optical receiver, *IEEE Photon. Technol. Lett.*, vol. 17, no. 6, June 2005, pp. 1271–1273.
14. S. S. Wagner and T. E. Chapuran, Broadband high-density WDM transmission using superluminescent diodes, *IEEE Electron. Lett.*, vol. 26, no. 11, May 1990, pp. 696–697.
15. D. K. Jung, H. Kim, K. H. Han, and Y. C. Chung, Spectrum-sliced bidirectional passive optical network for simultaneous transmission of WDM and digital broadcast video signals, *IEEE Electron. Lett.*, vol. 37, no. 5, Mar. 2001, pp. 308–309.
16. S. L. Woodward, P. P. Reichmann, and N. C. Frigo, A spectrally sliced PON employing Fabry–Perot lasers, *IEEE Photon. Technol. Lett.*, vol. 10, no. 9, Sept. 1998, pp. 1337–1339.
17. M. H. Reeve, A. R. Hunwicks, W. Zhao, S. G. Methley, L. Bickers, and S. Hornung, LED spectral slicing for single-mode local loop applications, *IEEE Electron. Lett.*, vol. 24, no. 7, Mar. 1988, pp. 389–390.
18. M. Zirngibl, C. H. Joyner, L. W. Stulz, C. Dragone, H. M. Presby, and I. P. Kaminow, LARnet, a local access router network, *IEEE Photon. Technol. Lett.*, vol. 7, no. 2, Feb. 1995, pp. 215–217.
19. D. K. Jung, S. K. Shin, C.-H. Lee, and Y. C. Chung, Wavelength-division-multiplexed passive optical network based on spectrum-slicing techniques, *IEEE Photon. Technol. Lett.*, vol. 10, no. 9, Sept. 1998, pp. 1334–1336.
20. R. D. Feldman, Crosstalk and loss in wavelength division multiplexed systems employing spectral slicing, *IEEE/OSA J. Lightwave Technol.*, vol. 15, no. 10, Oct. 1997, pp. 1823–1831.

21. S. Kobayashi, J. Yamada, S. Machida, and T. Kimura, Single-mode operation of 500 Mb/s modulated AlGaAs semiconductor laser by injection locking, *IEEE Electron. Lett.*, vol. 16, no. 19, Sept. 1980, pp. 746–748.
22. S.-M. Lee, K.-M. Choi, S.-G. Mun, J.-H. Moon, and C.-H. Lee, Dense WDM-PON based on wavelength-locked Fabry–Perot laser diodes, *IEEE Photon. Technol. Lett.*, vol. 17, no. 7, July 2005, pp. 1579–1581.
23. L. Y. Chan, C. K. Chan, D. T. K. Tong, F. Tong, and L. K. Chen, Upstream traffic transmitter using injection-locked Fabry–Perot laser diode as modulator for WDM access networks, *IEEE Electron. Lett.*, vol. 38, no. 1, Jan. 2002, pp. 43–45.
24. F.-T. An, K. S. Kim, Y.-L. Hsueh, M. Rogge, W.-T. Shaw, and L. G. Kazovsky, Evolution, challenges and enabling technologies for future WDM-based optical access networks (Invited paper), presented at the Joint Conference on Information Systems JCIS 2003, Research Triangle Park, NC, Sept. 2003.
25. H. D. Kim, S.-G. Kang, and C.-H. Le, A low-cost WDM source with an ASE injected Fabry–Perot semiconductor laser, *IEEE Photon. Technol. Lett.*, vol. 12, no. 8, Aug. 2000, pp. 1067–1069.
26. D. J. Shin, Y. C. Keh, J. W. Kwon, E. H. Lee, J. K. Lee, M. K. Park, J. W. Park, Y. K. Oh, S. W. Kim, I. K. Yun, H. C. Shin, D. Heo, J. S. Lee, H. S. Shin, H. S. Kim, S. B. Park, D. K. Jung, S. Hwang, Y. J. Oh, D. H. Jang, and C. S. Shim, Low-cost WDM-PON with colorless bidirectional transceivers, *IEEE/OSA J. Lightwave Technol.*, vol. 24, no. 1, Jan. 2006, pp. 158–165.
27. H.-J. Park, H. Yoon, T. Park, S.-J. Park, and J. H. Kim, Recent research activities of WDM-PON in Korea, presented at the Optical Fiber Communication and the National Fiber Optic Engineers Conference, Mar. 2007.
28. CIP Photonics press release, Ground breaking technology deployed by consortium to enable uncooled operation of advanced photonic devices, Feb. 2006, http://www.wordsun.com/cip9.htm.
29. N. J. Frigo, P. P. Iannone, P. D. Magill, T. E. Darcie, M. M. Downs, B. N. Desai, U. Koren, T. L. Koch, C. Dragone, H. M. Presby, and G. E. Bodeep, A wavelength-division multiplexed passive optical network with cost-shared components, *IEEE Photon. Technol. Lett.*, vol. 6, no. 11, Nov. 1994, pp. 1365–1367.
30. J. Kani, M. Teshima, K. Akimoto, N. Takachio, H. Suzuki, K. Iwatsuki, and M. Ishii, A WDM-based optical access network for wide-area gigabit access services, *IEEE Commun.*, vol. 41, no. 2, Feb. 2003, pp. S43–S48.
31. H. Takesue and T. Sugie, Wavelength channel data rewrite using saturated SOA modulator for WDM networks with centralized light sources, *IEEE/OSA J. Lightwave Technol.*, vol. 21, no. 11, Nov. 2003, pp. 2546–2556.
32. J. H. Lee, S.-H. Cho, H.-H. Lee, E.-S. Jung, J.-H. Yu, B.-W. Kim, S.-H. Lee, J.-S. Koh, B.-H. Sung, S.-J. Kang, J.-H. Kim, K.-T. Jeong, and S. S. Lee, First commercial deployment of a colorless gigabit WDM/TDM hybrid PON system using remote protocol terminator, *J. Lightwave Technol.*, vol. 28, no. 4, Feb. 15, 2010.
33. J. Prat, C. Arellano, V. Polo, and C. Bock, Optical network unit based on a bidirectional reflective semiconductor optical amplifier for fiber-to-the-home networks, *IEEE Photon. Technol. Lett.*, vol. 17, no. 1, Jan. 2005, pp. 250–252.
34. S. B. Tridandapani, B. Mukherjee, and G. Hallingstad, Channel sharing in multihop WDM lightwave networks: Do we need more channels? *IEEE/ACM Trans. Network.*, vol. 5, no. 5, Oct. 1997, pp. 719–727.

35. X. Zhao and F. S. Choa, Demonstration of 10-Gb/s transmissions over 1.5-km-long multimode fiber using equalization techniques, *IEEE Photon. Technol. Lett.*, vol. 14, no. 8, Aug. 2002, pp. 1187–1189.

36. S. S.-H. Yam, F.-T. An, S. Sinha, M. E. Marhic, and L. G. Kazovsky, 40 Gb/s transmission over 140 m of 62.5 μm multimode fiber using polarization-controlled launch, presented at the Conference on Lasers and Electro-Optics CLEO 2004, San Francisco, May 2005.

37. T. Itoh, H. Fukuyama, S. Tsunashima, E. Yoshida, Y. Yamabayashi, M. Muraguchi, H. Toba, and H. Sugahara, 1 km transmission of 10 Gb/s optical signal over Legacy MMF using mode-limiting transmission and incoherent light source, presented at the Optical Fiber Communication Conference OFC 2005, Anaheim, CA, Mar. 2005.

38. S. S.-H. Yam, J. Kim, D. Gutierrez, and F. Achten, Optical access network using centralized light source, single mode fiber + broad wavelength window multimode fiber, *IEEE/OSA J. Opt. Network.*, vol. 5, no. 8, Aug. 2006, pp. 604–610.

39. J. Kim, S. S.-H. Yam, D. Gutierrez, and L. G. Kazovsky, BER performance on access network using centralized light sources and single mode + multi mode fiber, presented at the Optical Fiber Communication Conference OFC 2006, Anaheim, CA, Mar. 2006.

40. R. Davey, J. Kani, F. Bourgart, and K. McCammon, Options for future optical access networks, *IEEE Commun.*, vol. 44, no. 10, Oct. 2000, pp. 50–56.

41. Y.-L. Hsueh, M. S. Rogge, S. Yamamoto, and L. G. Kazovsky, A highly flexible and efficient passive optical network employing dynamic wavelength allocation, *J. Lightwave Technol.*, vol. 23, no. 1, Jan. 2005, pp. 277–286.

42. Y.-L. Hsueh, M. S. Rogge, W.-T. Shaw, S. Yamamoto, and L. G. Kazovsky, Quality of service support over SUCCESS-DWA: a highly evolutional and cost-effective optical access network, presented at the Optical Fiber Communication Conference 2005, Mar. 2005.

43. C. Bock, J. Prat, and S. D. Walker, High-density ring-tree advanced access topology delivering bandwidth on demand over a resilient infrastructure, presented at the European Conference on Optical Communications ECOC 2006, Cannes, France, Sept. 2006.

44. J. J. Lepley, M. P. Thakur, I. Tsalamanis, S. D. Walker, K. Habel, K.-D. Langer, G.-J. Rijckenberg, A. Ng'oma, A. M. J. Koonen, and J. S. Wellen, Interoperability of last mile access technologies over a passive optical ring network: results of the MUSE demonstration, presented at the European Conference on Optical Communications ECOC 2006, Cannes, France, Sept. 2006.

45. F.-T. An, K. S. Kim, D. Gutierrez, S. Yam, E. Hu, K. Shrikhande, and L. G. Kazovsky, SUCCESS: a next-generation hybrid WDM/TDM optical access network architecture, *IEEE/OSA J. Lightwave Technol.*, vol. 22, no. 11, Nov. 2004, pp. 2557–2569.

46. F.-T. An, D. Gutierrez, K. S. Kim, J. W. Lee, and L. G. Kazovsky, SUCCESS-HPON: a next-generation optical access architecture for smooth migration from TDM-PON to WDM-PON, *IEEE Commun.*, vol. 43, no. 11, Nov. 2005, pp. S40–S47.

47. K. S. Kim, D. Gutierrez, F.-T. An, and L. G. Kazovsky, Batch scheduling algorithm for SUCCESS WDM-PON, presented at the IEEE Global Telecommunications Conference GLOBECOM 2004, Dallas, TX, Nov. 2004.

48. K. S. Kim, D. Gutierrez, F.-T. An, and L. G. Kazovsky, Design and performance analysis of scheduling algorithms for WDM-PON under SUCCESS-HPON architecture, *IEEE/OSA J. Lightwave Technol.*, vol. 23, no. 11, Nov. 2005, pp. 3716-3731.

49. D. Gutierrez, K. S. Kim, F.-T. An, and L. G. Kazovsky, SUCCESS-HPON: migrating from TDM-PON to WDM-PON, presented at the European Conference on Optical Communications ECOC 2006, Cannes, France, Sept. 2006.
50. K. S. Kim, On the evolution of PON-based FTTH solutions, *Inf. Sci.*, vol. 149, no. 1–2, Jan. 2003, pp. 21–30.
51. G. Kramer, B. Mukherjee, and G. Pesavento, IPACT: a dynamic protocol for an Ethernet PON (EPON), *IEEE Commun.*, vol. 40, Feb. 2002, pp. 74–80.
52. F. Jia, B. Mukherjee, and J. Iness, Scheduling variable-length messages in a single-hop multichannel local lightwave network, *IEEE/ACM Trans. Network.*, vol. 3, no. 4, Aug. 1995, pp. 477–488.
53. B. Bosik and S. Kartalopoulos, A time compression multiplexing system for a circuit switched digital capability, *IEEE Trans. Commun.*, vol. 30, no. 9, Sept. 1982, pp. 2046–2052.
54. A. Varga, OMNeT++: discrete event simulation system, Version 3.2, http://www.omnetpp.org/, Oct. 2005.

CHAPTER 5

HYBRID OPTICAL WIRELESS ACCESS NETWORKS

Optical and wireless networks were developed initially for different communication purposes and applied in different scenarios. Optical network technology was developed for long-distance and high-bandwidth communications, and wireless network technology for short-distance communication systems that do not require high bandwidth but do require flexibility. In the last two decades, however, due to the insatiable growth of bandwidth demand and the desire of end users for ubiquitous Internet access, optical and wireless communication technologies have been employed in last-mile connections to enhance bandwidth and to enable flexibility and mobility, respectively. Today we are witnessing the convergence of optical and wireless technologies at the access segment of the Internet hierarchy (Figure 5.1).

To date, various optical and wireless access solutions have been developed to address various challenges of access networks, such as broadband service, cost-efficiency, and demand for mobility. For example, passive optical networks (PONs) [1,2] have emerged to replace copper wire access networks for bandwidth enhancement, and IEEE 802.16 (WiMAX) [3,4] has been developed to provide cost-effective and flexible Internet access. Since current optical and wireless access technologies aim to address different issues, it is difficult for any single technology to resolve all the challenges in access networks. For example, albeit optical access technologies enable broadband services, the dedicated infrastructure to a user's house leads to significant deployment cost, and network availability is confined within a residential and business unit. Similarly, despite its ubiquitous and flexible connectivity, the limited

Broadband Optical Access Networks, First Edition. Leonid G. Kazovsky, Ning Cheng, Wei-Tao Shaw, David Gutierrez, and Shing-Wa Wong.

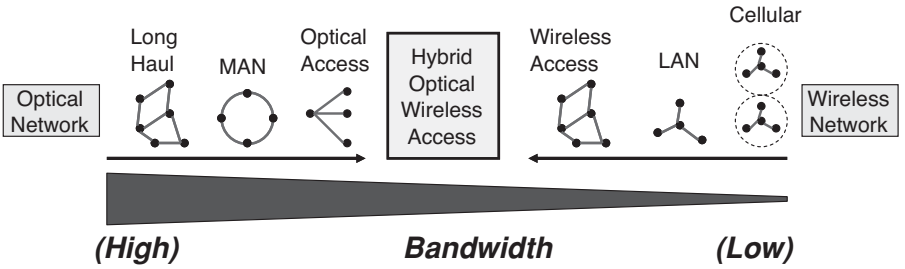


FIGURE 5.1 Convergence of optical and wireless networks at the access segment.

bandwidth of wireless access technologies prevents simultaneous broadband access among a large number of users.

In light of the complementary characteristics of optical and wireless technologies, a hybrid optical wireless access network may provide a desirable compromise among bandwidth, cost, and network availability issues. Specifically, a hybrid optical wireless access architecture consists of wireless networks in a user's proximity and optical networks as the backhaul. In the front end, wireless networks are deployed to penetrate into a user's vicinity, facilitating ubiquitous connectivity and minimizing the deployment cost. In the back end, the broadband optical backhaul networks are deployed to connect the wireless network and the central office that manages the entire network. Compared with wireless access solutions that are limited in bandwidth, a hybrid optical wireless architecture should be able to facilitate bandwidth upgrade in the wireless segment by deploying more resources in the optical segment. On the other hand, compared with optical access solutions that require deploying infrastructure to a user's proximity, a hybrid optical wireless architecture should be able to reduce the service cost by replacing the last/first hundreds of feet of infrastructure on the optical segment with ubiquitous wireless links. As a result, a hybrid optical wireless access network should enable scalable bandwidth provisioning according to the demand in a cost-effective manner. Since in-depth knowledge of optical access networks has been provided in earlier chapters, in the following sections we introduce two emerging wireless access technologies: WiMAX and Wi-Fi mesh networks, as background knowledge, before we dive into hybrid optical–wireless access networks.

5.1 WIRELESS ACCESS TECHNOLOGIES

5.1.1 IEEE 802.16 WiMAX

The mobile worldwide interoperability for microwave access (WiMAX) is an emerging wireless access technology designed to replace wireline access technologies such as coaxial cable and digital subscriber loop (DSL) and to support mobile communications. It is based on the IEEE 802.16 2004 air interface standard [3], specifying

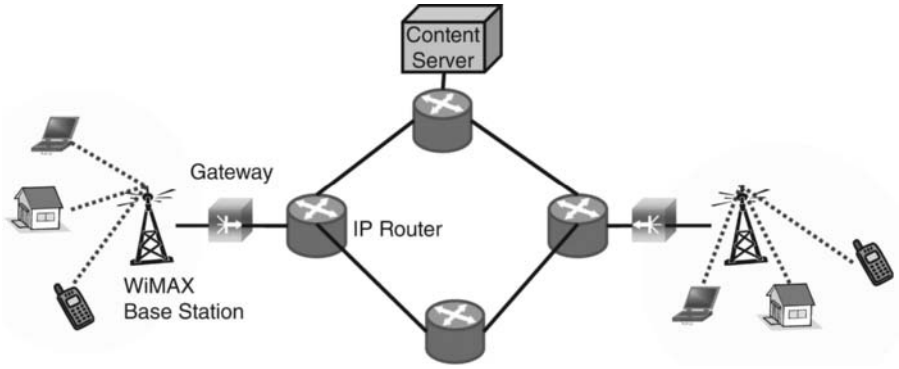


FIGURE 5.2 Mobile WiMAX network architecture.

the basic technologies and performance requirements in the physical (PHY) and medium access control (MAC) layers, and on the IEEE 802.16e amendment [4], an amendment to add features and attributes that are necessary to support mobility in the WiMAX network.

The typical mobile WiMAX network illustrated in Figure 5.2 serves fixed, nomadic, and mobile users. A WiMAX base station covers a cell with radius up to 1 km in urban areas, 1 to 3 km in a suburban area, and more than 10 km in a rural area. For mobile users a mobile WiMAX can support up to 120 km/h with handoff. The target applications of a WiMAX network include voice over Internet protocol (VoIP), videoconferencing, streaming media, interactive real-time gaming, Web browsing, instant messaging, and multimedia downloading. Some of these applications consume significant bandwidth and/or require short end-to-end latency. Connections in mobile WiMAX are based on an Internet protocol (IP) end-to-end network, which is an integrated data-communications network that uses an IP address for the end-to-end connection. This architecture separates the last-mile access network from the IP connectivity service and supports global roaming and interworking with other systems.

Concerning system capacity, the peak downlink throughput of a mobile WiMAX network can achieve up to 60 Mb/s and uplink 28 Mb/s (given a 10-MHz channel bandwidth allocation; the channel bandwidth allocation will be explained later). Like most other wireless systems, the capacity of mobile WiMAX depends heavily on the geographical distribution of users. The maximum throughput occurs when all users are close to the base station and with the line-of-sight wireless link. Where most users are at the edge of a cell, the downlink throughput decreases to 2 Mb/s and the uplink to 400 kb/s. In more general cases, where users are evenly distributed within a cell, the throughout varies depending on dynamic user demand. To achieve fair bandwidth allocation among users, a scheduler with a proper fairness algorithm should be employed in a mobile WiMAX network. In the following sections we introduce the key features and enabling technologies of a mobile WiMAX system.

5.1.1.1 RF Bands Specified in Mobile WiMAX

Multiple radio-frequency bands are specified in the mobile WiMAX standard [3]. These bands are either licensed or license-exempt.

10 to 66-GHz Licensed Bands In 10 to 66-GHz bands, due to the high radio frequency, a line-of-sight (LOS) propagation environment is required, and the multipath effect is negligible. In this band, channel bandwidths of 25 or 28 MHz are typically assigned for high throughput. With raw data rates in excess of 120 Mb/s, transmission at these bands is well suited for point-to-multipoint (PMP) access serving applications from small and home offices to medium-sized and large office applications.

Frequencies Below 11 GHz Frequencies below 11 GHz do not require an LOS propagation environment, and the multipath effect may be significant due to the longer wavelength. The ability to support near-LOS and non-LOS (NLOS) scenarios requires additional PHY functionality, such as the support of advanced power management techniques, interference mitigation and coexistence, and multiple antennas.

License-Exempt Frequencies Below 11 GHz (Primarily 5 to 6 GHz) For license-exempt bands below 11 GHz, the LOS propagation environment and multipath effects are similar to those of licensed bands below 11 GHz. Nonetheless, the license-exempt nature introduces additional interference and coexistence issues, whereas regulatory constraints limit the radiated power allowed. Note that this band is used exclusively in a mobile WiMAX network to support mobility.

5.1.1.2 Orthogonal Frequency-Division Multiplexing in WiMAX

As in most other broadband wireless communications systems, a multicarrier modulation (MCM) scheme is used in mobile WiMAX, which is a powerful and practical technique to achieve high data rates over frequency-selective channels [5,6]. In typical MCM systems, various transmission parameters, such as modulation scheme, coding rate, and transmitter power, can be adjusted to optimize the overall performance according to the channel characteristics, if they are available. In particular, rather than allocating equal power and data rate among the subcarriers, one can adapt to the propagation environment by dynamically allocating power and data rate across subcarriers to optimize system throughput in an MCM system [7,8].

Among different MCM schemes, orthogonal frequency-division multiplexing (OFDM) is an extended, multiuser orthogonal frequency-division multiplexing that can accommodate many users at the same time on the same channel [9]. OFDM is used in a mobile WiMAX system because it provides tolerance to multipath, frequency-selective fading, scalable channel bandwidth, and high compatibility with advanced antenna technology [10]. As shown in Figure 5.3, in an OFDM system the RF spectrum is divided into multiple-frequency subcarriers to facilitate multiplexing. The subcarriers in a WiMAX system are categorized into three types, depending on their functions (as in Figure 5.4): (1) data subcarriers for data transmission, (2) pilot subcarriers for channel estimation and system synchronization, and (3) null subcarriers for guard bands and dc carriers. The input data stream for each user is

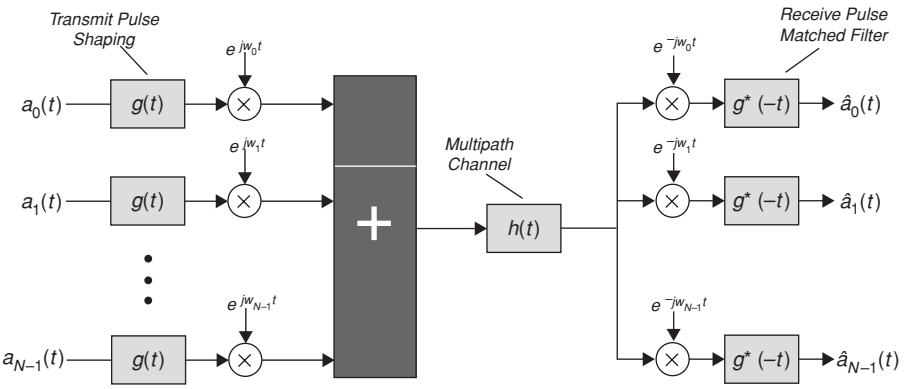


FIGURE 5.3 OFDM system architecture.

divided into several parallel data paths or substreams, each at a reduced data rate, and the substreams are modulated and transmitted on separate orthogonal subcarriers. Given these subcarriers, resource/bandwidth allocation can be achieved in both the time and frequency domains by means of symbols and subcarriers, respectively; thus, the bandwidth for each user can be adjusted dynamically and flexibly according to demand. Due to the reduced data rate in each subcarrier, the symbol duration is increased, which in turn improves the robustness to delay spread that often occurs in a non-line-of-sight wireless channel. By introducing a guard time that is greater than the delay spread for each OFDM symbol, the intersymbol interference (ISI) due to delay spread can almost be eliminated. The frequency diversity of the multipath channel is exploited by coding and interleaving the information across the subcarriers prior to transmission. By using frequency-selective scheduling, the right subcarriers are selected for each user to minimize the impact of frequency-selective fading.

WiMAX supports various modulation schemes and coding rates: BPSK, QPSK, 16QAM, and 64QAM, and the coding efficiency ranges from 1/2 to 5/6, to further boost the level of system capacity. The combinations of modulations and code rates provide a fine resolution of data rates, as shown in Figure 5.5, in the 5- and 10-MHz

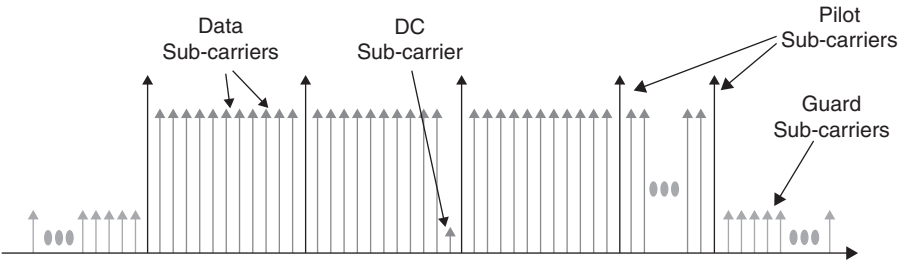


FIGURE 5.4 OFDMA subcarrier in a WiMAX system.

Parameter	Downlink	Uplink	Downlink	Uplink
System Bandwidth	5 MHz		10 MHz	
FFT Size	512		1024	
Null Sub-Carriers	92	104	184	184
Pilot Sub-Carriers	60	136	120	280
Data Sub-Carriers	360	272	720	560
Sub-Channels	15	17	30	35
Symbol Period, T_s	102.9 microseconds			
Frame Duration	5 milliseconds			
OFDM Symbols/Frame	48			
Data OFDM Symbols/Frame	44			

Mod.	Code Rate	5 MHz Channel		10 MHz Channel	
		Downlink Rate, Mbps	Uplink Rate, Mbps	Downlink Rate, Mbps	Uplink Rate, Mbps
QPSK	1/2 CTC, 6x	0.53	0.38	1.06	0.78
	1/2 CTC, 4x	0.79	0.57	1.58	1.18
	1/2 CTC, 2x	1.58	1.14	3.17	2.35
	1/2 CTC, 1x	3.17	2.28	6.34	4.70
	3/4 CTC	4.75	3.43	9.50	7.06
16QAM	1/2 CTC	6.34	4.57	12.07	9.41
	3/4 CTC	9.50	6.85	19.01	14.11
64QAM	1/2 CTC	9.50	*6.85*	*19.01*	*14.11*
	2/3 CTC	12.67	*9.14*	*26.34*	*18.82*
	3/4 CTC	14.26	*10.28*	*28.51*	*21.17*
	5/6 CTC	15.84	*11.42*	*31.68*	*23.52*

Rate ID	Modulation rate	Coding	Information bits/symbol	Information bits/ OFDM symbol	Peak data rate in 5 MHz (Mb/s)
0	BPSK	1/2	0.5	88	1.89
1	QPSK	1/2	1	184	3.95
2	QPSK	3/4	1.5	280	6.00
3	16QAM	1/2	2	376	8.06
4	16QAM	3/4	3	568	12.18
5	64QAM	2/3	4	760	16.30
6	64QAM	3/4	4.5	856	18.36

FIGURE 5.5 Modulations and code rates for mobile WiMAX. (From [10].)

channels. Each frame has 48 OFDM symbols, with 44 OFDM symbols available for data transmission. The amount of error tolerated depends on the reliability and QoS requirements of each connection.

In an OFDM system, the data can be modulated using inverse fast Fourier transform (IFFT), which enables bandwidth scalability by proportioning the FFT size according to the channel bandwidth. In this way, not only is the implementation complexity reduced, but also a mobile WiMAX system is transparent to the higher layers and

is compatible with various RF spectrum regulations in different countries to achieve worldwide interoperability, as its name suggests.

Compatibility with advanced antenna technology is a significant strength of an OFDM system, because it slices a wideband frequency-selective channel into a number of narrowband subcarriers. This allows signal processing of advanced antenna technology over flat subcarriers without the complex equalizers that are often required in a wideband, frequency-selective fading channel. Currently, mobile WiMAX supports space-time coding, spatial multiplexing, and smart antenna beamforming, which are powerful techniques to improve spectral efficiency and system capacity.

5.1.1.3 Duplex Scheme and Frequency Reuse in a WiMAX System

WiMAX standards [3] support time-division-duplex (TDD), frequency-division-duplex (FDD), and half-duplex FDD operation. TDD requires only one channel for transmitting downlink and uplink subframes at two distinct time slots, whereas FDD requires two distinct channels for transmitting downlink and uplink subframes in the same time slot. In half-duplex FDD operation, a user device can only transmit or receive at any given time. Currently, the initial release of a WiMAX certification profile will consider the TDD scheme only for the following reasons: (1) TDD enables flexible bandwidth allocation between downlink and uplink to support an efficient asymmetric traffic pattern; (2) TDD assures channel reciprocity for better support of link adaptation, MIMO, and other closed-loop advanced antenna technologies; (3) unlike FDD, TDD requires only a single channel for both downlink and uplink, so that it provides greater flexibility for adaptation to varied global spectrum allocations; and (4) transceiver designs for TDD implementations are less complex and therefore less expensive. To counter interference issues, however, TDD requires systemwide synchronization. With ongoing releases, FDD profiles will be considered by the WiMAX Forum to address specific market opportunities where local spectrum regulatory requirements either prohibit TDD or are more suitable for FDD deployment.

Due to the scarce spectrum resources, the frequency channels must be reused throughout a WiMAX system by controlling the transmission power of the base station and user devices, as in cellular systems. To maximize spectral efficiency, WiMAX supports a frequency reuse of 1 (i.e., all cells operate on the same frequency channel). However, due to heavy co-channel interference, users at the cell edge may suffer degradation in connection quality in frequency reuse of 1 deployment. Since OFDMA divides the frequency channel into subchannels, the interference at the cell edge can easily be addressed by reconfiguring the subchannels. The reuse pattern of subchannels can be configured so that all the subchannels can be used to serve users close to the base station, while a fraction of the subchannels are allocated to serve users at the cell edge. Figure 5.6 shows an exemplary reuse pattern, where F1, F2, and F3 are three different subchannels. In this configuration, a full-load frequency reuse of 1 is maintained for users close to the base station, to maximize spectral efficiency, whereas fractional frequency reuse is used to improve user connection quality and throughput at the cell edge.

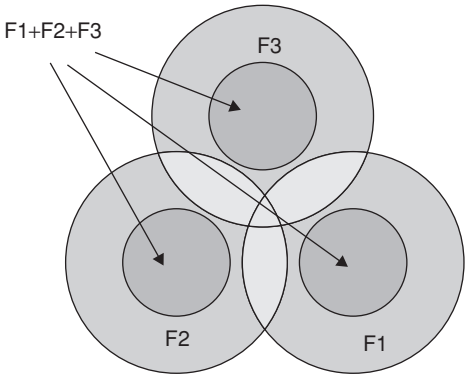


FIGURE 5.6 Subchannel reuse pattern source.

5.1.1.4 MAC Protocol of WiMAX In wireless network systems, medium access control (MAC) plays a key role in bandwidth efficiency and user experience. In IEEE 802.11, for example, the MAC protocol uses contention-based access; that is, all subscriber stations that wish to pass data through a wireless access point are competing for the channel on a random-interrupt basis. As traffic loading increases, the contention will cause a large amount of interference among the subscriber stations, which not only reduces bandwidth efficiency but also leads to significant performance degradation such as lower throughput, longer packet delay, and even packet drop. This performance degradation challenges services such as voice over IP or video streaming, which are sensitive to quality of service (QoS), to be supported in Wi-Fi networks.

In contrast, the MAC protocol of IEEE 802.16 uses a scheduling algorithm for which the subscribers just need to contend once (for initial entry into the network). After a full round of contention, each subscriber is allocated an access time slot by the base station. The resource allocated to one subscriber station by the MAC scheduler can vary from a single time slot to the entire frame, thus providing a large range of throughput to a specific subscriber station at any given time. Furthermore, for each user the resource allocation information is conveyed at the beginning of each frame, so the scheduler can effectively change the resource allocation on a frame-by-frame basis to adapt to the bursty nature of the traffic. Besides improving bandwidth efficiency, the scheduling algorithm also allows the base station to control QoS parameters by balancing the time-slot assignments among the application needs of the subscriber stations. In fact, the WiMAX standard was developed at the outset for the delivery of broadband services, including voice, data, and video. The applications include IPTV, VoIP, videoconferencing, Web services, and computer data services. It is designed to support bursty data traffic with high peak-rate demand [11] while simultaneously supporting streaming video and latency-sensitive voice traffic over the same channel. In the next section we review in more detail the QoS supported by WiMAX.

5.1.1.5 Quality of Service Support As mentioned, the fundamental premise of the WiMAX MAC protocol is to support QoS. With a fast air interface, fine bandwidth granularity, and a flexible resource allocation mechanism, QoS requirements can be met for a wide range of data services, such as voice and video. As defined in the MAC protocol, QoS is provided via service flows with a particular set of QoS parameters [3]. Before providing a certain type of data service, the base station and user terminal first establish a unidirectional logical link between the peer MACs called a *connection*. The outbound MAC then associates packets traversing the MAC interface into a service flow to be delivered over the connection. The service flows can be mapped to fine granular IP sessions or coarse differentiated service code points to enable end-to-end IP-based QoS. Additionally, subchannelization and medium access protocol (MAP)-based signaling schemes provide a flexible mechanism for optimal scheduling of broadcast and unicast traffic using space, frequency, and time physical resources over the air interface on a frame-by-frame basis. The QoS parameters associated with the service flow define the transmission ordering and scheduling on the air interface. The connection-oriented QoS can therefore provide accurate control over the RF interface. The service flow parameters can be managed dynamically through MAC messages to accommodate the dynamic bandwidth demand. The service flow–based QoS mechanism applies to both downlink and uplink, to improve QoS in both directions.

5.1.1.6 Mobile Multihop Relay WiMAX Networks The WiMAX network defined in the IEEE 802.16-2004 air interface standard [3] and IEEE 802.16e amendment [4] is a single-hop wireless network. To improve the network coverage and flexibility of the current mobile WiMAX standard with limited wired connections to WiMAX base stations, a new task group, IEEE 802.16j [12], was established in 2006 to support mobile multihop relay (MMR) operation in the current IEEE 802.16e standard. A topology envisioned for a future IEEE 802.16j MMR network is depicted in Figure 5.7. To accommodate the multihop transmission between WiMAX stations, numerous technical challenges, such as the OFDMA frame structure, routing path management, handoff, security, and relay reliability need to be addressed while maintaining compatibility with the current WiMAX standard. In fact, using multihop

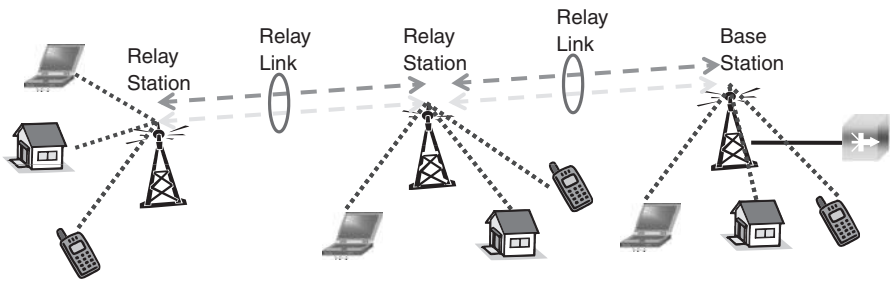


FIGURE 5.7 IEEE 802.16j: mobile multiple relay WiMAX network.

communications to enhance network coverage and flexibility with a limited wired connection is not a new idea. For example, mobile ad-hoc networks (MANETs) [13] are infrastructure-less, decentralized wireless networks. In a MANET every network entity can move and function as a router that discovers and maintains routes to other entities in the network. Historically, MANETs have been developed for special applications such as on battlefields, in emergency networks, and in wireless sensor networks. Another example is the wireless mesh network (WMN), an infrastructure-based, hierarchical multihop network. In a WMN, wireless mesh routers, deployed as network infrastructure, interface with a subscriber's device and relay traffic to and from other mesh routers. The goal of WMNs is to provide an affordable and scalable wireless access solution for community and metropolitan areas [14]. Given the rapid advance in wireless communications, WMNs have been commercialized using high-speed, cost-effective wireless interfaces and technologies. In some U.S. cities, for example, WMNs using IEEE 802.11a/b/g as the interface have been deployed to provide wireless Internet access services [15,16]. By upgrading to higher-data-rate technologies such as mobile WiMAX, WMNs have been considered a promising wireless access solution to replace or complement existing access networks in metropolitan areas. In the hybrid optical–wireless architecture to be introduced, therefore, we use WMN as the wireless segment. In the following sections we introduce WMNs in detail.

5.1.2 Wireless Mesh Networks

A wireless mesh network (WMN) consists of multihop wireless communication links to forward traffic to and from wired Internet entry points [14], as depicted in Figure 5.8. On a WMN, any node can communicate with any other through a route comprising multiple wireless links (e.g., nodes A and B via the route formed by nodes 1, 2, and 3). A WMN dynamically self-configures and automatically maintains the

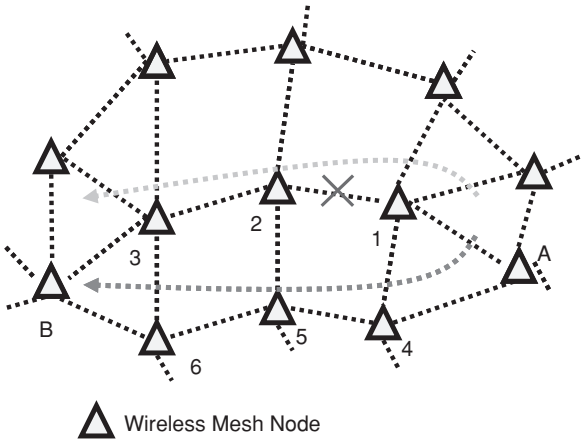


FIGURE 5.8 Generic wireless mesh network.

mesh connectivity among nodes in the network. One important feature of WMNs is the self-healing capability: for example, if the link between nodes 1 and 2 fails, the traffic between nodes A and B can go through the route formed by nodes 4, 5, and 6.

A WMN originates from mobile ad hoc networks [13], which have been developed historically for specific applications. Fueled by the advances of wireless technologies and ad hoc networking optimization, in the last decade WMNs have found more general applications, such as the extension of wireless local area networks, citywide wireless access networks, community wireless networks, and public safety networks [17]. Compared to conventional single-hop wireless networks, WMNs enhance network coverage cost-effectively and improve load balancing across the network. In addition, WMNs have several advantages, such as robustness to node failures, fast network deployment, ease of maintenance, and low initial deployment cost [14,17].

Depending on the applications and type of nodes, WMNs can be categorized as client WMNs or infrastructure WMNs.

5.1.2.1 Client Wireless Mesh Networks Client WMNs are infrastructureless, peer-to-peer wireless networks. As illustrated in Figure 5.9, a client WMN consists of only one type of mesh nodes, client devices that provide end users with an application interface. These client devices perform traffic routing and mesh configuration functionalities among themselves, and they usually have mobility. Recently, an IEEE technical task group under IEEE 802.11 has been working to develop a new client WMN standard, the IEEE 802.11s [18]. Based on the PHY interface and MAC protocol defined by IEEE 802.11a/b/g, IEEE 802.11s focuses on the network formation, maintenance, and traffic routing. For portable client devices such as laptop and PDA, the transmission power is usually limited, so IEEE 802.11s is considered as an extension of wireless local area networks (WLANs).

5.1.2.2 Infrastructure Wireless Mesh Networks An infrastructure WMN has a hierarchical network architecture consisting of two types of wireless nodes: wireless mesh routers and client devices (Figure 5.10). Wireless mesh routers are deployed as infrastructure without mobility; client devices are usually mobile and provide an application interface to end users. In this architecture there are two link

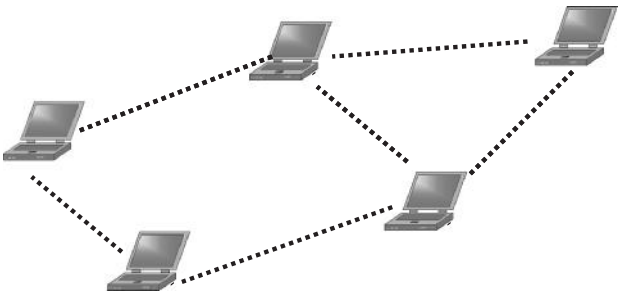


FIGURE 5.9 Client wireless mesh network.

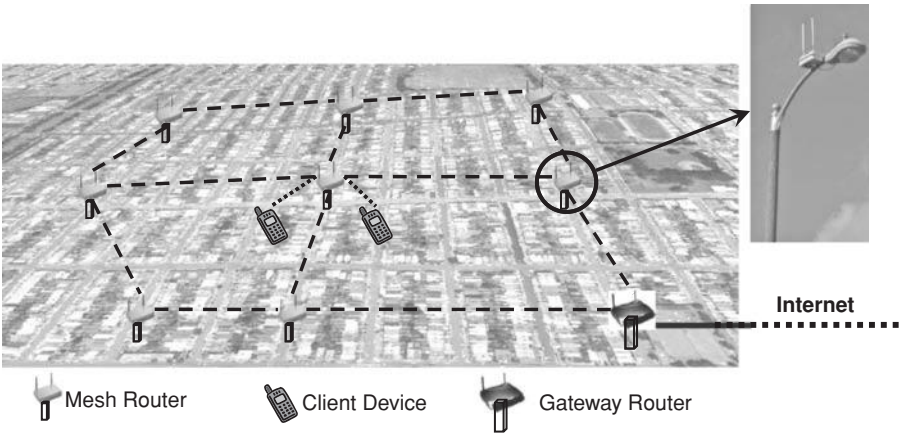


FIGURE 5.10 Infrastructure wireless mesh network.

layers: (1) the links between a mesh client and a mesh router, and (2) the links between mesh routers. As in Figure 5.10, wireless mesh routers form a self-configuring, self-healing mesh network. Among mesh routers, one or some mesh routers have wired connections to the Internet, called *gateway routers*, and traffic is distributed or aggregated from or to the routers. Considering Figure 5.8, upstream traffic from the end user is first transmitted by the mesh client and received by the nearby mesh router (i.e., router 2). Then the packets are relayed to one of the nearby gateway routers, gateway router A or B. The downstream traffic is forward from the gateway router to the end user in the same manner but in the reverse direction. IEEE802.11 technologies are widely used in the lower layer of most infrastructure WMNs due to their wide popularity among portable electronic devices and PCs and their low cost. The higher layer link can be built using various types of radio technologies, in addition to the most used IEEE 802.11 technologies. Infrastructure WMNs are emerging as a promising wireless access solution to complement or even replace wired access networks because of its cost-effectiveness, scalable architecture, and flexible network deployment.

Since this book's emphasis is on access networks, in the rest of this chapter we focus on infrastructure WMNs, and *WMN* is referred to hereafter as *infrastructure WMN*.

5.1.2.3 PHY and MAC Layers of WMN As mentioned earlier, IEEE 802.11b/a/g (Wi-Fi) technologies are widely exploited both in the links between mesh routers and in the links between mesh routers and client devices [15,16,19]. These Wi-Fi-based WMNs are going to be deployed in urban areas in North America. However, due to the multihop communications among mesh routers in an open environment, as network load increases, co-channel interference among mesh routers will increase significantly. As designed originally for local area networks, the PHY technology and CSMA-CA (carrier-sense multiple access with collision avoidance) protocol of Wi-Fi are not optimized for high-load situations. Therefore, the network

TABLE 5.1 Summary of IEEE 802.11 Standards

Parameter	802.11b	802.11g	802.11a	802.11n
Maximum bit rate (Mb/s)	11	54	54	200+
Nonoverlapping channels	3 × 20 MHz	3 × 20 MHz	15 × 20 MHz	
Maximum uplink distance (m)	100	100	50	100

efficiency at the higher layer of Wi-Fi based WMNs will degrade significantly as the network is highly loaded. The family of IEEE 802.11 standards is summarized in Table 5.1.

To improve the overall network performance of WMNs, therefore, new technologies and protocols in the PHY and MAC layers and a new routing protocol have been proposed [17]. In the PHY layer, for example, smart antenna, multi-input multi-output (MIMO), and multichannel/interface systems are being explored to enhance network capacity. MAC protocols based on distributed time- and code-division multiplexing access are expected to improve the bandwidth efficiency from CSMA/CA protocols [17]. Furthermore, in WMNs, since packets are routed among mesh routers in the presence of interference, shadowing, and fading, a cross-layer design is required to optimize routing in WMNs. To accommodate the projection of demand increase for Wi-Fi-based WMN, WMN vendor companies have proposed using WiMAX (IEEE 802.16) to enhance the higher-layer link capacity of WMNs [15]. Ultrahigh-bandwidth standards such as the emerging IEEE 802.16m, which aims to provide 1-Gb/s and 100-Mb/s shared bandwidth for residential and mobile users [20], can be employed to further enhance capacity.

5.1.2.4 Routing in Wireless Multihop Networks Routing for wireless multihop networks has been researched extensively in the context of wireless ad hoc networks. There exist several algorithms or protocols for routing with varying performance parameters. Basically, these routing protocols can be categorized into two classes: the proactive class and the reactive class, represented by the optimized link state routing (OLSR) [21] and ad hoc on-demand distance vector (AODV) [22] protocols, respectively. The proactive routing protocols compute possible paths to all destinations, regardless of the traffic demand between the communicating node pairs, and store them in routing tables. Since these tables need to be updated in a time-varying environment, the overhead may become prohibitively high for a large network. The reactive routing protocols, on the other hand, compute routing paths on an "on-demand" basis, as determined by the dynamics of the traffic categories and network topology. The overhead for maintaining a large routing table can thus be avoided. However, for reactive routing protocols the delay in establishing a new path is longer. The proactive and reactive routing protocols are summarized in Figure 5.11.

More recently, in the context of wireless community networks, several routing metrics have been suggested to incorporate underlying physical layer conditions [23–26]. For example, the expected transmission count (ETX) is proposed [25] to minimize

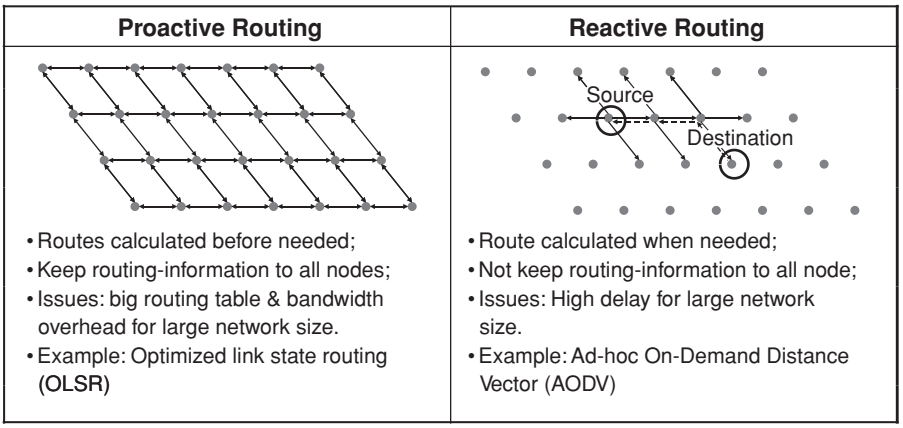


FIGURE 5.11 Routing in wireless ad hoc networks.

the expected total number of packet transmissions required to deliver a packet successfully to the destination. The metric incorporates the effects of packet loss rates in both the forward and reverse directions of a link and the interference among the successive hops of a path. In [25], ETX was adapted into two conventional shortest-path routing protocols, DSDV [27] and DSR [28], and was shown to improve the network performance via simulations and a test bed implementation consisting of 29 nodes in an office building. A similar metric, the expected transmission time (ETT) [23,24], considers the bandwidth of a link as well as the loss rate. The metric was used in a routing scheme for networks employing multiple radio interfaces and was shown to improve performance over a scheme based on ETX. In [26], another routing algorithm based on a cross-layer approach is proposed. The authors introduce physical layer metrics representing the transmission rate, interference, and packet error rate (PER), and develop a heuristic algorithm for joint power control and routing to optimize the information-theoretic capacity [29] of "dense" multihop wireless networks. Simulation results showed that the heuristic algorithm improved network throughput noticeably compared to AODV [22], DSDV [27], and DSR [28], commonly benchmarked routing protocols.

Although WMN is one type of wireless multihop network, it is very different from wireless ad hoc networks in terms of network architectures and applications. For example, wireless ad hoc networks are designed to support mobility among the nodes, to minimize power consumption, and to forward peer-to-peer traffic among the nodes. On the other hand, the wireless mesh routers in WMNs have no mobility, less power consumption constraint, and higher computational capability. In addition, the traffic pattern in WMNs is very different from the peer-to-peer traffic in wireless ad hoc networks since in WMNs, if not all, forwarded most, traffic is between a gateway router and end users [30]. The differences between WMNs and wireless ad hoc networks are summarized in Figure 5.12.

	Node	Backhaul	Traffic Type	Power	Computation Ability
Wireless Ad-Hoc Network	Mobile	No	Any node to any node	Battery	Low
Wireless Mesh Network	Stationary	Yes	Sink/source at gateway	Plugged	High

FIGURE 5.12 WMNs vs. wireless ad hoc networks.

To date, most routing protocols of WMNs have been retrofitted based on wireless ad hoc network architectures. Therefore, we envision that more research should be conducted on the routing algorithm specifically for WMNs.

5.1.2.5 Propagation Environments and Models of Wireless Mesh Networks Although there has been extensive research in communications techniques and networking protocols for wireless ad hoc networks and more recently for WMNs, most research is based on idealized propagation models that make analysis tractable but do not represent all the issues of radio propagation in a real environment. As a result, protocols and algorithms tailored to these idealized environments do not actually perform as expected in real networks. In this section we review some critical issues in a propagation model when a WMN is deployed in highly populated urban areas.

Without loss of generality, the urban area where a WMN is deployed can be assumed to have a square grid of buildings and streets. Both users on the street and users in buildings should be considered in the propagation model. For users in buildings, additional attenuation due to building penetration needs to be addressed. [31–37]. Mesh routers are usually placed on streetlight poles in urban areas. Hence, the topology of WMNs can be assumed to be grid networks. The distance between adjacent mesh routers is one of the key parameters used when deploying a WMN; it is a complicated function of many factors across various network layers, such as the SINR required for targeted error performance under physical layer (PHY) configurations, a medium access control (MAC) scheme, and a routing strategy.

Essentially three components are involved in characterizing and modeling the propagation of radio signals:

1. *Distance-dependent path loss:* describes the trend in the average signal power received, as a function of the distance between the transmitter and the receiver;
2. *Large-scale fading or shadowing:* characterizes fluctuation in the signal power received locally averaged over several wavelengths. Note that the distance-dependent path loss and large-scale fading jointly describe large-scale variations in the signal strength received.
3. *Small-scale fading or multipath fading:* characterizes the fluctuation on the order of a wavelength around the local average signal strength, which is caused by replicas of the signal transmitted, with random-phase shifts that arrive and add up at the receiver.

The characteristics of each component depend strongly on the propagation environment and system parameters, such as channel bandwidth, carrier frequency, and antenna configurations at both transmitter and receiver. For highly populated urban areas, measurement, characterization, and modeling of radio propagation have been reported [34–41]. In particular, references 38 and 40 provide good models and parameters for large-scale variation: reference 38 suggested urban propagation models based on extensive measurements conducted in San Francisco, and reference 40 proposed models based on measurements made in central Stockholm. Next we describe the three components in detail.

5.1.2.6 Capacity and Scalability Issues of Wireless Mesh Networks

The insatiable demand for bandwidth from end users requires access network to support high and scalable capacity, which is, however, a challenging issue for WMNs because the traffic aggregation on the relay links between mesh routers tends to grow as the number of hops increases. Unlike fiberoptic access networks, where the capacity can be enhanced by adding more wavelengths, the bandwidth resource of wireless networks is so scarce and expensive that the capacity cannot be increased by increasing the frequency channels. Although the bandwidth efficiency of WMNs can be enhanced by employing advanced PHY technologies and MAC protocols, frequency reuse is still the most effective way to enhance the overall capacity. In cellular systems, the network capacity can be enhanced by deploying more base stations in an area, which results in smaller cells, a technique called *cell splitting*. In WMNs, however, straightforward cell splitting (i.e., deploying more wireless mesh routers in an area) will not lead to capacity enhancement [14,17].

To clarify this issue, let us consider the one-dimensional WMN shown in Figure 5.13(a). In this one-dimensional WMN, each mesh router (MR) is modeled as a node that connects to other nodes and the local water tank with pipes. The water tank models the service area (or cell) under each mesh router. In each water tank, the traffic load of users (DMR) is modeled as water. The thickness of the pipes signifies the capacity of the links, and we assume that the capacities of two types of links or layers (CMR-user, CMR-MR) are independent. The gateway router (GR), which has backhaul connection to the Internet, is the outlet of the entire system, where we assume that there is only upstream traffic.

As the traffic load from end users increases under each mesh router, the link capacity between the mesh router and end users may be exhausted, as indicated in Figure 5.13(b). To reduce the loading of each mesh router, we can deploy more mesh routers in a given area with lower transmission power, as in Figure 5.13(c), similar to cell splitting in cellular systems. Note that after cell splitting, the pipe between the gateway router and its adjacent mesh router needs to conduct the aggregated flows (D_{agg}) from more mesh routers; in this exemplar system the number increases from three in Figure 5.13(a) to five in Figure 5.13(c).

Also note that the pipe connecting the routers may become thinner after cell splitting because of the increased co-channel interference from other routers. Since the link connecting the gateway router and its adjacent mesh router needs to aggregate traffic on the basis of additional mesh routers, the aggregated load (D_{agg}) may soon

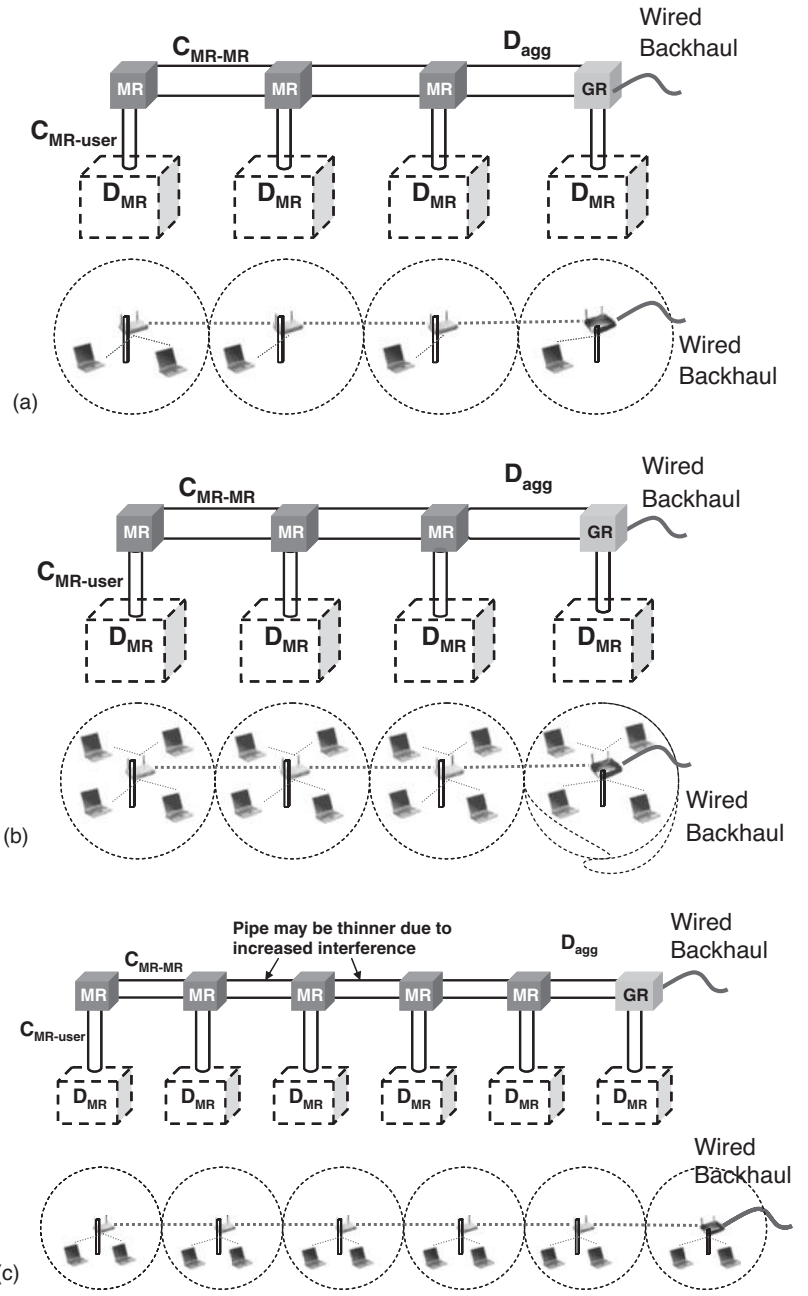


FIGURE 5.13 Scalability in a one-dimensional exemplar WMN. MR, wireless mesh router; GR, wireless gateway router; D_{agg}, aggregated data rate; C_{MR-MR}, link capacity between adjacent MRs, $C_{MR-user}$, link capacity between an MR and users; D_{MR}, total user demand within the area served by an MR.

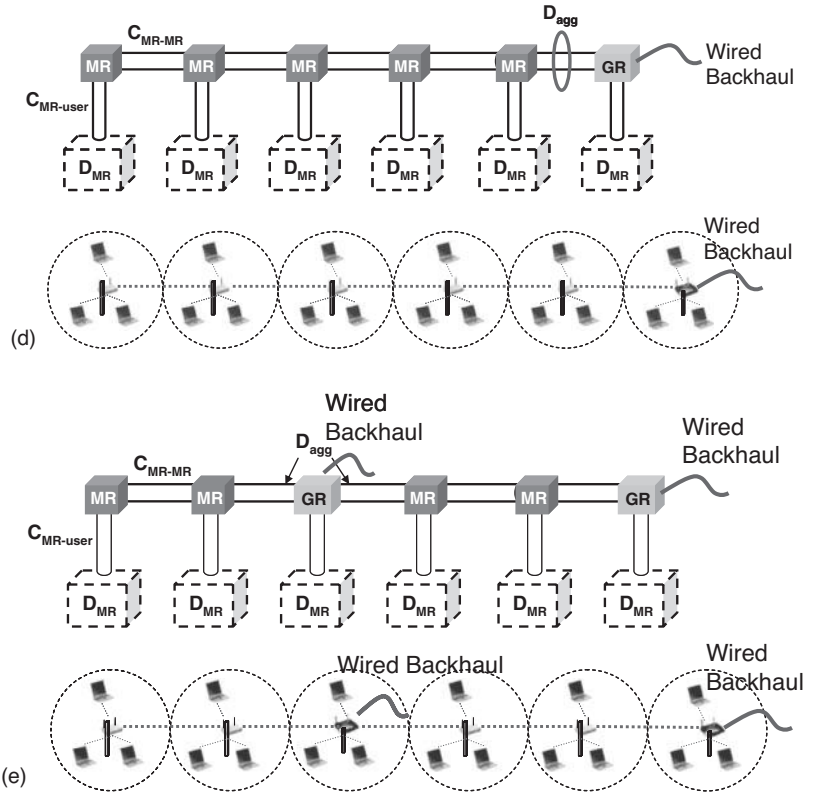


FIGURE 5.13 (*Continued*) Scalability in a one-dimensional exemplar WMN.

exhaust the pipe capacity connecting the gateway router and its adjacent mesh router, as shown in Figure 5.13(d). The bandwidth insufficiency of this link will choke each mesh router's throughput and the overall network capacity of a WMN. In other words, after cell splitting, although each router's average loading is reduced, the throughput of each router and the overall capacity of the WMN may not necessarily increase. To improve the performance, a straightforward solution is to upgrade the link capacity by exploiting more frequency channels and radio interfaces on the routers [17] (i.e., widening the pipes of the model). However, since the number of available channels in the license-free bands is few and the right to use the licensed bands is very expensive, this approach is limited either by the finite capacity or by the cost. Instead of resorting to upgrading the capacity of wireless links, the throughput per router and overall capacity can be enhanced by scaling the number of gateway routers in accord with that of the mesh routers [i.e., replacing the middle mesh router with another gateway router, as illustrated in Figure 5.13(e)]. After placing this gateway router, traffic flows from some mesh routers can be routed to the new gateway router, and thus the bottleneck at the original gateway router link is mitigated.

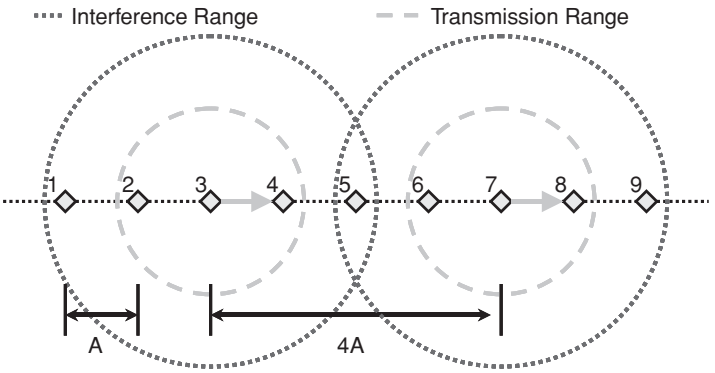
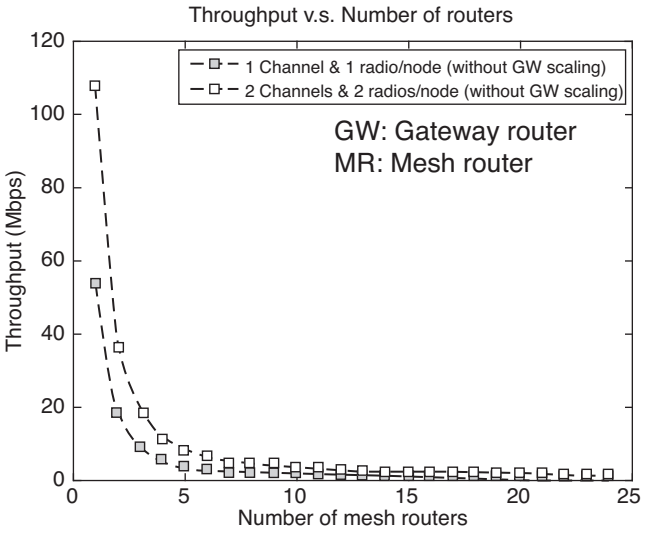


FIGURE 5.14 One-dimensional WMN.

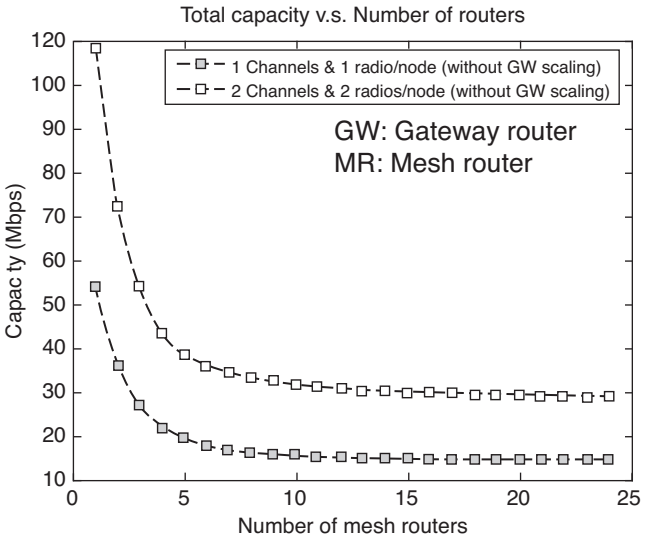
After qualitative insight into the scalability issues of WMNs and the proposed solution, let's analyze the maximum throughput per router and the overall capacity of a one-dimensional WMN. In this analysis we first assume that there is only one gateway router at the end of the one-dimensional WMN, as in Figure 5.14, and that the link speed of the higher layer of links among the mesh and gateway routers is 54 Mb/s and does not interfere with the links of the lower layer. We assume further that the distance between any two adjacent mesh routers of the WMN is a constant, A, as shown in Figure 5.14, and that all the routers are identical in transmission power, receiver sensitivity, and type of antenna (omnidirectional).

The radio channel(s) are spatially reused, and the transmission of one mesh router will introduce co-channel interference to other mesh routers. By applying a simplified range model [42,43], we assume that the transmission range (TR) is A^+ ($A < A^+ < 2A$), and the interference range (IR) is $2A^+$ ($2A < 2A^+ < 3A$). In other words, each mesh router can receive packets correctly only from its two adjacent routers and is interfered with routers that are up to two hops away. Given these assumptions, we simulate the throughput per router and the overall capacity as a function of the number of mesh routers. As shown in Figure 5.15(a) and (b), the maximum throughput drops to zero and the overall network capacity reaches a finite value as the number of routers continues to increase.

On the other hand, if we add one gateway router to every four additional mesh routers to keep the ratio of gateway routers to mesh routers constant, the average throughput will approach a finite value and the overall capacity will continue scaling up as shown in Figure 5.15(c) and (d), respectively. Another important observation revealed in Figure 5.15(c) and (d) is that given the same number of mesh routers in WMNs, increasing the ratio of gateway routers over mesh routers can improve the throughput per router and the overall network capacity. This is because given a higher gateway router/mesh router ratio, the average number of hops that a packet takes is reduced.

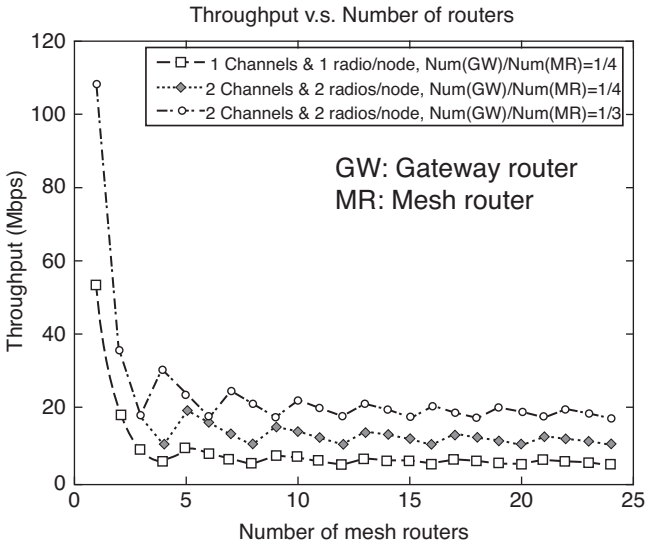


(a) Throughput vs. number of routers without increase of gateway (link rate equals to 54Mbps)

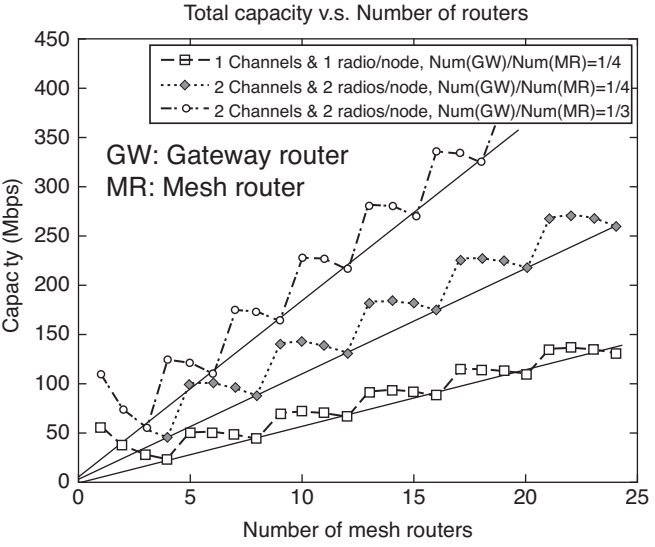


(b) Capacity vs. number of routers without increase of gateway (link rate equals to 54Mbps)

FIGURE 5.15 Throughput per router and overall capacity of a one-dimensional WMN.



(c) Throughput vs. number of routers with increase of gateway (link rate equals to 54Mbps)



(d) Capacity vs. number of routers with increase of gateway (link rate equals to 54Mbps)

FIGURE 5.15 (*Continued*) Throughput per router and overall capacity of a one-dimensional WMN.

As a result, to enable flexible increase of wireless gateway routers in a WMN, a broadband and scalable backhaul network is required. The backhaul network needs to connect cost-effectively to numerous wireless gateway routers scattered geographically throughout a large area, and to accommodate the increase of wireless gateway routers in the future.

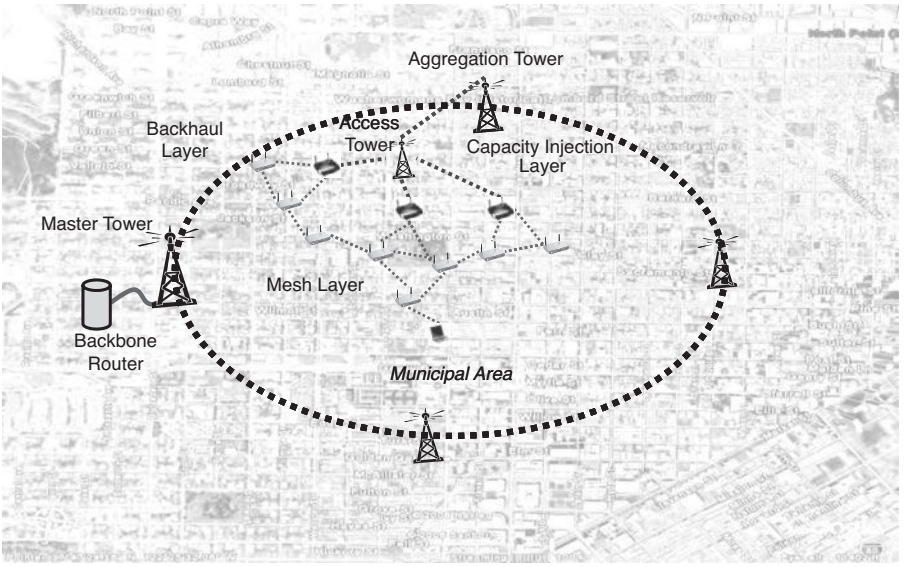
The backhaul network for WMNs is built on either wireless or wired infrastructure. In the following sections we review these backhaul technologies.

5.1.2.7 Backhaul Network Example for WMNs To reduce deployment time and cost, wireless backhaul is an attractive solution for network operators. Taking the proposal of a wireless access network for San Francisco [44] as an example, wireless point-to-multipoint and ring links are deployed to collect and distribute WMN traffic across the city. The network architecture is shown in Figure 5.16(a), which forms a hierarchical wireless access network. There are five types of nodes in the hierarchical architecture: wireless mesh routers, wireless gateway routers, access towers, aggregation towers, and master towers. As shown in Figure 5.16(b), there are three layers, consisting of wireless links between different types of nodes.

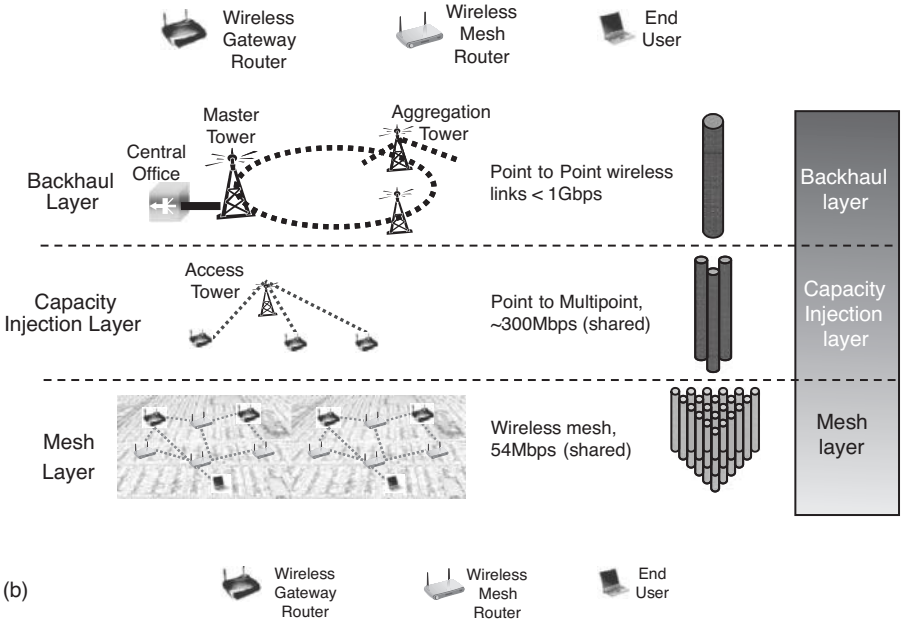
- *First layer (mesh layer):* consists of wireless links between wireless mesh routers and gateway routers of a WMN, which are deployed throughout the urban area and penetrate a user's premises to provide ubiquitous, blanket-coverage connection.
- *Second layer (capacity injection layer):* consists of high-speed point-to-multipoint wireless links between the access tower and wireless gateway routers of a WMN and aggregates the traffic from the mesh layer. For each access tower, approximately 300 Mb/s of bandwidth [45] is shared among multiple wireless gateway routers.
- *Third layer (backhaul layer):* consists of (a) point-to-point wireless links between the access towers and aggregation towers, and (b) point-to-point wireless links among the aggregation towers and master tower.

The wireless links among the aggregation towers and master tower form a ring network, and a pair of radios is used to facilitate redundancy. Note that in this layer each point-to-point link has a dedicated (unshared) bandwidth that is less than 1 Gb/s.

Note that the second and third layers jointly constitute the backhaul for the lower WMN layer. Proprietary wireless technologies have been employed to realize the backhaul network [44]. Although this proposed wireless backhaul is readily deployed to bring the service up quickly, its capacity is very limited for supporting the aggregated traffic of the entire city. The capacity in the second and third layers will therefore be exhausted soon after the network is deployed. The bandwidth insufficiency in the backhaul layers will be further exacerbated as the bandwidth of the lower WMN layer is enhanced by employing advanced wireless technologies.



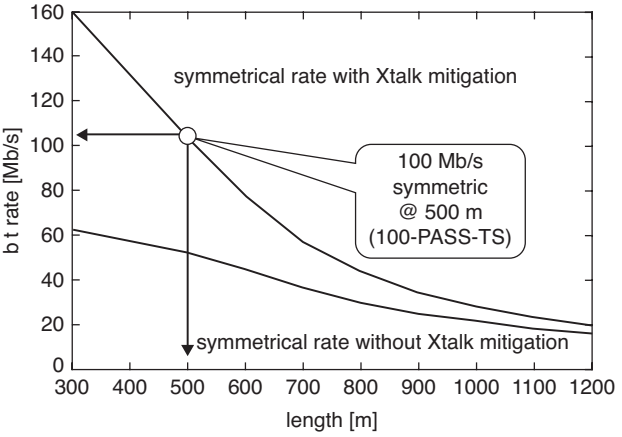
(a)



(b)

FIGURE 5.16 Hierarchical wireless access network (proposed by Google-Earthlink to the city of San Francisco).

5.1.2.8 Broadband Backhaul Technologies To aggregate the huge volume of WMN traffic from an urban area, a broadband backhaul network is imperative. Besides providing high capacity, the backhaul network should also support long-distance transmission to collect traffic from numerous gateway routers that are densely deployed in an urban area. To fulfill these two requirements simultaneously, copper wire technologies may not be an appropriate solution because of their limited bandwidth–distance product. The bit rate and transmission distance of the advanced digital subscriber loop technology, VDSL2, and the coaxial cable are summarized in Figure 5.17(a) and (b). As a result, we imagine that ultrabroadband wireless



(a) Bit rate versus distance of VDSL2

Coax Cable Signal Loss (Attenuation) in dB per 100ft*

Loss*	RG-174	RG-58	RG-8X	RG-213	RG-6	RG-11	RF-9914	RF-9913
1MHz	1.9dB	0.4dB	0.5dB	0.2dB	0.2dB	0.2dB	0.3dB	0.2dB
10MHz	3.3dB	1.4dB	1.0dB	0.6dB	0.6dB	0.4dB	0.5dB	0.4dB
50MHz	6.6dB	3.3dB	2.5dB	1.6dB	1.4dB	1.0dB	1.1dB	0.9dB
100MHz	8.9dB	4.9dB	3.6dB	2.2dB	2.0dB	1.6dB	1.5dB	1.4dB
200MHz	11.9dB	7.3dB	5.4dB	3.3dB	2.8dB	2.3dB	2.0dB	1.8dB
400MHz	17.3 B	11.2dB	7.9dB	4.8dB	4.3dB	3.5dB	2.9dB	2.6dB
700MHz	26.0dB	16.9dB	11.0dB	6.6dB	5.6dB	4.7dB	3.8dB	3.6dB
900MHz	27.9 B	20.1dB	12.6dB	7.7dB	6.0dB	5.4dB	4.9dB	4.2dB
1GHz	32.0dB	21.5dB	13.5dB	8.3dB	6.1dB	5.6dB	5.3dB	4.5dB
Imped	50ohm	50ohm	50ohm	50ohm	75ohm	75ohm	50ohm	50ohm

(b) Attenuation of coaxial cable

FIGURE 5.17 Fundamental limit of the transmission distance of copper wire technologies.

TABLE 5.2 Comparison of Broadband Backhaul Technologies for WMN

Parameter	Millimeter-Wave (60 GHz)	Free-Space Optics	Optical Fiber Access
Maximum rate	1.25 Gb/s per TX/RX	1.25 Gb/s per TX/RX	2.5/1.25 Gb/s per λ
Maximum distance (km)	1.2	2	20
Link type	Point-to-point	Point-to-point	Point-to-point
Reliability	Low (line of sight)	Low (line of sight)	Medium
Capacity scalability	Medium	High (WDM)	High (WDM)
Dominant cost	Tx/Rx	Tx/Rx	Infrastructure deployment

technologies and fiber optics are appropriate candidates. Let's review and compare these two types of solutions.

For wireless backhaul approaches, millimeter-wave [46] and free-space optical [47] technologies are two candidates because of their high capacity. However, due to their high carrier frequencies, these broadband wireless technologies are strongly subject to line-of-sight conditions and thus reliability becomes an issue: Links may degrade or be broken due to obstruction or even to harsh weather conditions. The degraded or broken backhaul link will affect numerous downstream WMN users. In addition, these wireless technologies can operate at full rate only within certain distance limitations: The maximum distance between millimeter-wave transceivers is 1.2 km [46], and the maximum distance between free-space optical transceivers is 2 km [47]. This distance limitation will increase the deployment cost and constrain the flexibility as a backhaul network is deployed. The data rate, cost, and transmission distance of commercial millimeter-wave and free-space optical products are summarized in Table 5.2.

In contrast to these broadband wireless backhaul technologies, fiber optical backhaul has an ultrahigh capacity, long transmission distance, and high reliability, as also summarized in Table 5.2. Some may argue that the cost of infrastructure deployment of optical networks requires expensive investment. Thanks to technological advances, the cost of optical devices has dropped significantly and is now less than that of broadband wireless equipment. In the next section we investigate the deployment cost of the fiber-optical infrastructure for access networks.

5.1.2.9 Deployment Cost of Backhaul Network Based on TDM-PON

Historically, fiber optical telecommunications technology was costly, so it was employed mainly for long-haul systems. Nonetheless, to realize the last-mile optical connection to end users, new optical fiber deployment techniques and strategies have been developed for passive optical networks to reduce the deployment cost. Currently, there are three deployment methods: (1) aerial: stringing fibers above the ground, sometimes using existing utility poles; (2) buried: digging trenches to install fiber; and (3) conduit: placing fiber in existing conduit underground.

TABLE 5.3 Cost of Fiber Deployment for PON

Deployment Method	Deployment Cost ($/km)
Aerial[a]	900
Trenching[a]	1200
Existing conduit[a]	700
New conduit[b]	4000

[a] Ronald Heron, Alcatel, Inc., NFOEC 2007 Keynote.
[b] M. Hajduczenia et al., Optimized passive optical network deployment, *J. Opt. Networking*, Sept. 2007.

The typical costs of the three deployment methods are summarized in Table 5.3. By comparing the dominant costs of wireless technologies and optical access networks, the use of fiber optic networks as an economically feasible solution for backhaul connection to a WMN is justified. Note that besides deploying new optical fibers, in urban areas dark fibers or dim fibers (parts of the fibers have been used for communications) are deployed along the streets. These fibers can readily be employed as part of the backhaul network at even lower cost.

As optical technology is used to provide backhauling to WMNs, the integration of optical fiber backhaul and WMNs leads to a hybrid access network architecture. In the following sections we introduce hybrid optical–wireless access networks.

5.2 HYBRID OPTICAL–WIRELESS ACCESS NETWORK ARCHITECTURE

A generic hybrid optical–wireless access network is illustrated in Figure 5.18, where wireless links provide ubiquitous links in end-users' premises, and optical backhaul aggregates the wireless traffic. A cellular system and its backhauling network comprise perhaps one of the first hybrid optical–wireless networks. In this type of hybrid network, the wireless base stations are connected to the central office/network server through dedicated optical point-to-point links. Compared to the aggregated voice/data traffic of a wireless base station, however, the dedicated high-speed optical link overprovisions bandwidth, and the infrastructure topology prohibits bandwidth and device sharing among multiple base stations.

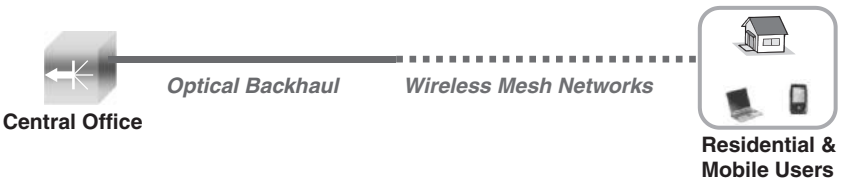


FIGURE 5.18 Hybrid optical–wireless access network.

Another type of hybrid optical and wireless network is the radio over fiber (ROF) technology [48]. ROF technologies allow signals of wireless communication systems, such as cellular systems and WLAN, to be transmitted transparently over optical fiber. Since the optical fiber system is transparent to the wireless signal, all the processing, such as modulation/demodulation, coding/decoing, handover, and so on, is carried out in a central station. The two ends of the optical systems translate wireless signal to and from wireless radio signal only in the physical (PHY) layer. Since our focus is on the hybrid network architecture, which includes the physical, data link, and networking layers, ROF is outside the scope of our interest.

In metro WMNs, numerous wireless routers need to be deployed in users' premises to provide blanket coverage. The traffic of these wireless routers that are geographically scattered across an urban area is forwarded to a few central offices or network servers. To connect to widely scattered wireless routers and accommodate the ever-increasing traffic growth from end users, a cost-effective, scalable, and multiplexed optical backhaul technology is desirable.

To fulfill these requirements, TDM-PON technology is a promising solution because of its technological maturity and cost justification. The benefits of using TDM-PON to backhaul WMNs include (1) cost-effectiveness: TDM-PON allows a large portion of infrastructure and devices in the central office to be shared among end users, the ONUs/wireless nodes; (2) bandwidth efficiency: the inherent statistical multiplexing of TDM-PON optimizes bandwidth usage; (3) topology flexibility: TDM-PON supports different types of network topologies, such as tree, bus, and ring [49], in connecting to widely scattered WMN routers.

5.2.1 Leveraging TDM-PON for Smooth Upgrade of Hierarchical Wireless Access Networks

A hybrid optical wireless architecture that uses TDM-PON technology to connect to the WMNs deployed in a metropolitan area has been proposed [50,51]. The generic hybrid optical and wireless architecture is illustrated in Figure 5.19, where the backhaul network consists of a fiber ring and tree networks. ONUs are deployed at the ends of tree networks and connected to wireless nodes. Since TDM-PON streams are multiplexed at different wavelengths, a large number of TDM-PON streams will require dense wavelength-division multiplexing (DWDM) for both downstream and upstream traffic. At the joint of the ring and a tree network, a low-loss wavelength add/drop filter is installed to add or drop a wavelength set, consisting of multiple wavelengths, to separate the downstream and upstream traffic. The wavelength add/drop filters segregate the physical network into numerous point-to-multipoint networks, on which dedicated wavelengths carry traffic between the central office and wireless gateway routers. Hence, TDM-PON technology can readily be employed. Note that although not shown, optical amplifiers may be required to compensate for the component loss along the optical ring network and at the joint nodes if a large number of ONUs are connected at the ends of the tree network.

Under this hybrid optical–wireless architecture, the upstream traffic is first received by a nearby mesh router via the links of the lower layer and then relayed to a nearby

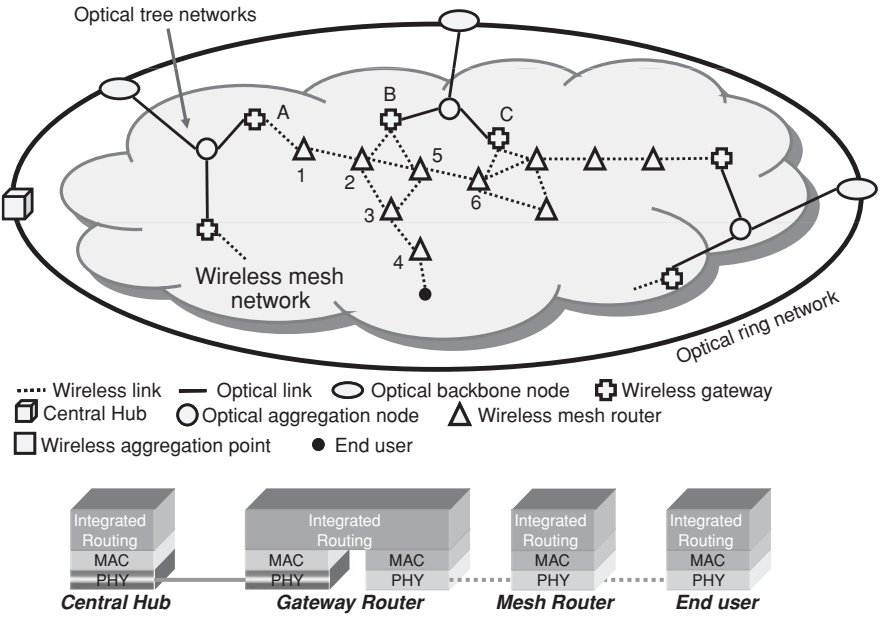


FIGURE 5.19 Proposed hybrid optical–wireless access network [50,51].

gateway router through multihop communications. In Figure 5.19, for example, router 4 aggregates traffic of nearby end users and relays it over routers 3, 2, and 1 to reach gateway router A. Once traffic reaches the gateway router, it is forwarded to the central office over the optical backhaul. For the downstream traffic, packets are first routed to one of the gateway routers, such as B in Figure 5.19, then forwarded on a specific route (e.g., through routers 5, 3, and 4) to the end user.

Since the optical backhaul and WMNs are implemented with different technologies, such as Ethernet PON (EPON) and Wi-Fi, interoperability is needed at the interface between the ONU and wireless gateway router. To address this issue, the optical backhaul and WMN can be either fused at the networking layer using an IP router, or employ an application-specific integrated circuit (ASIC) designed to translate the packet formats.

Integration of the point-to-multipoint optical backhaul and WMNs paves multiple routes between the central hub and an end user. An integrated routing paradigm that can dynamically choose the optimum route is therefore essential for a hybrid optical–wireless network. Although routing in WMNs by itself is a challenging issue, as described earlier, we assume that the optical backhaul can help to collect some network conditions in WMNs, such as link status, traffic loading, and interference and help to determine the optimum route. An integrated routing algorithm to achieve load balancing as congestion occurs in the wireless mesh network has been proposed [51]. This integrated routing algorithm is discussed in the following sections.

Hybrid architecture can be used to smoothly upgrade a hierarchical wireless access network, as discussed earlier. In fact, a wireless backhaul in a hierarchical wireless access network [Figure 5.16(b)] enables quick and cost-effective deployment, so it is strategic to deploy a hierarchical wireless access network and then upgrade incrementally to a hybrid network. In the following section we describe the smooth upgrading path.

5.2.2 Upgrading Path

The proposed hybrid architecture is designed to smoothly upgrade a hierarchical wireless access network [Figure 5.16(a) and (b)]. The upgrading path proposed [51,52] begins with the replacement of the optical link at the backhaul layer in Figure 5.16(b). Because the top layer aggregates traffic from the lower layers, this layer will become the first bandwidth bottleneck.

Figure 5.20(a) shows the upstream wireless segments next to the central office. These wireless links aggregate traffic from the downstream segments along the ring, so they need to be upgraded first. Figure 5.20(b) shows the upgrade of the first wireless link; a fiber is deployed to replace the wireless link. A TDM PON stream at a pair of wavelengths (a $\lambda 1$ pair) is allotted to facilitate communications between the central office and the first aggregation tower. The $\lambda 1$ pair consists of two different wavelengths for separating the downstream and upstream traffic. At the first aggregation tower, a low-loss optical wavelength add/drop is installed to add or drop the wavelength pair $\lambda 1$. The wireless links under the first aggregation tower (i.e., the capacity injection layer in Figure 5.16) and the wireless link between the first and second aggregation towers (i.e., the backhaul layer in Figure 5.16) remain intact. Note that the first aggregation tower enjoys the full bandwidth provisioned by the TDM-PON stream, and if needed, more wavelength pairs that carry TDM-PON streams can be allotted in addition to the $\lambda 1$ pair.

Figure 5.20(c) illustrates that the upgrade proceeds to the second wireless link in the backhaul layer. In Figure 5.20(c), a $\lambda 2$ pair is allotted to carry a new TDM-PON stream, which bypasses the first aggregation tower and drops at the second aggregation tower. The replacement can be applied to the subsequent wireless links until the entire backhaul layer is upgraded.

As the wireless links in a ring network can be upgraded with optical links, the wireless links between the aggregation and access towers can also be upgraded with optical links. Note that in reference 44, a point-to-point wireless link is dedicated to facilitate communication between the aggregation tower and a nearby access tower. So point-to-multipoint topology of the aggregation and access towers is based on many expensive point-to-point links. To upgrade these links, we can simply install an ONU by the access tower to replace the wireless link. Figure 5.20(d) shows that as the wireless links are gradually upgraded, the point-to-multipoint MAC protocol of TDM-PON will automatically manage the point-to-multipoint network and its bandwidth allocation.

Once the bottleneck in the backhaul layer is resolved by upgrading to optical links, the capacity injection layer will become the next bandwidth bottleneck. The upgrade

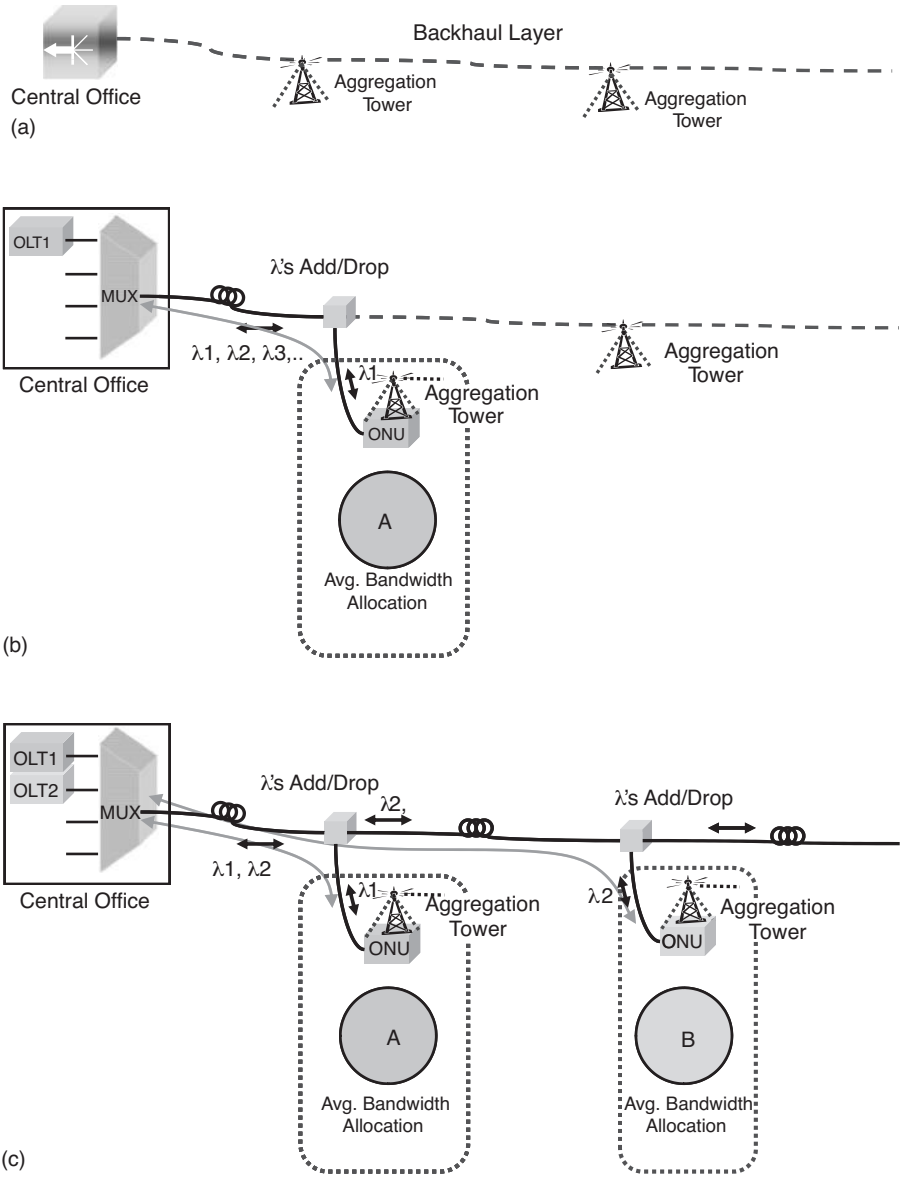
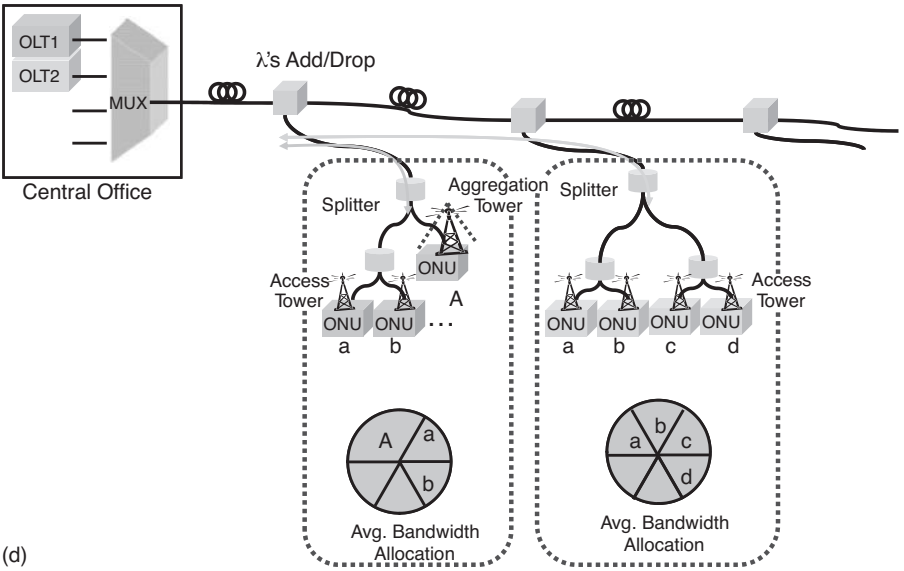
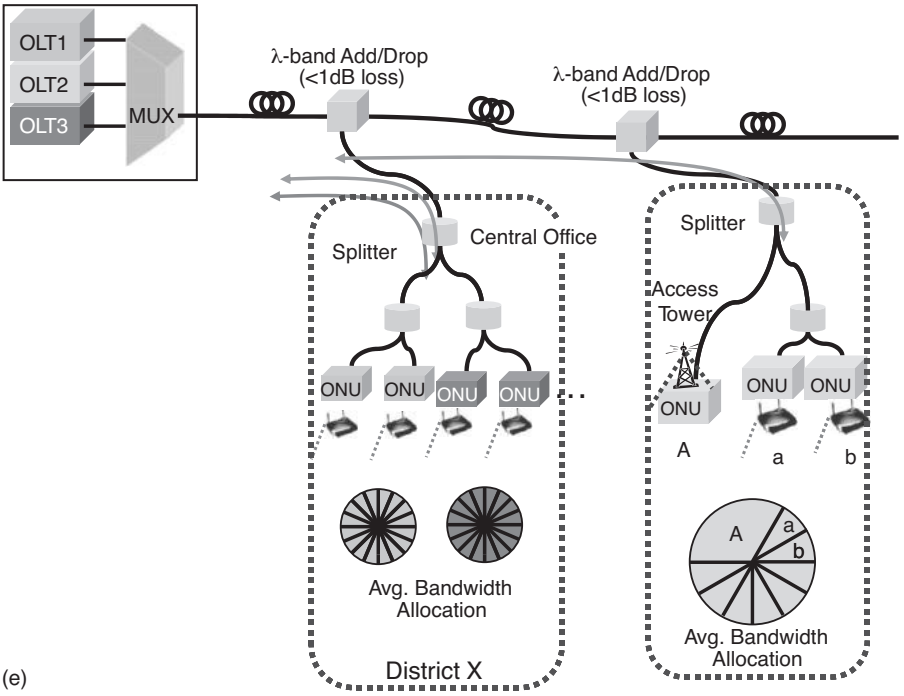


FIGURE 5.20 Smooth upgrading path of the wireless backhaul proposed by Google and Earthlink for a wireless access network in San Francisco.



(d)



(e)

FIGURE 5.20 (*Continued*) Smooth upgrading path of the wireless backhaul proposed by Google and Earthlink for a wireless access network in San Francisco.

can be realized with the same principle by further deploying ONU at a wireless gateway router to gradually replace wireless links, as shown in Figure 5.20(e). If a certain district needs more bandwidth [e.g., district X in Figure 5.20(e)], another TDM-PON stream on a different wavelength pair can be allotted. Passive optical filters will be required at the ONU to separate different wavelengths. Note that the upgrade will eventually result in the hybrid optical–wireless architecture illustrated in Figure 5.19.

5.2.3 Reconfigurable Optical Backhaul Architecture

In a hybrid optical–wireless network, the bandwidth demand from different districts can vary drastically within a certain time period. For example, a business district will require a large bandwidth during daytime but little during nighttime, and the residential district the opposite. Instead of overprovisioning based on the peak demand or underprovisioning to optimize the network investment, it is desirable to have a reconfigurable network on which the bandwidth can be reallocated among districts serving multiple TDM-PON streams. Given the TDM/WDM backhaul structure, the reconfigurability can be realized by using tunable optical transceiver at the ONU. For example, in Figure 5.21 districts X and Y are both connected to the same physical tree network and are served by PON1 and 2, respectively. Assume that tunable transceivers are used at ONUs and that each ONU will only transmit and receive a certain wavelength pair. If PON1 is highly loaded and PON2 is not at a certain moment, by tuning the transmitting and receiving wavelengths of ONU3 from 1D, 1U to 2D, 2U, ONU3 is enabled to join PON2 to reduce PON1's loading. In this way the bandwidth is reallocated based on the dynamic demand.

To optimize the bandwidth utilization with the reconfigurability proposed, the central hub needs to monitor the bandwidth demand and allocate the bandwidth based on TDM-PON technologies. Figure 5.22 shows an example that can be implemented

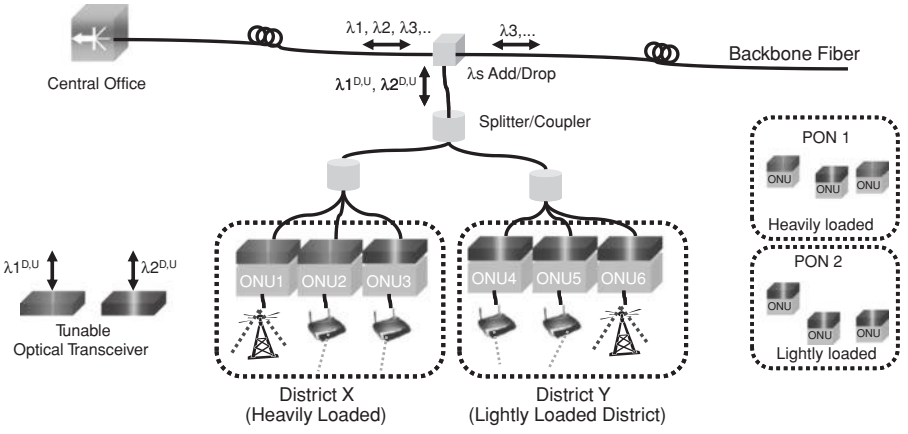


FIGURE 5.21 Reconfigurable optical backhaul.

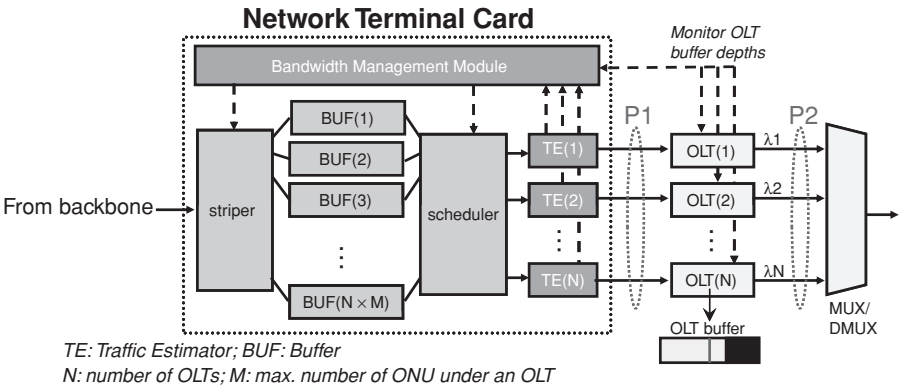


FIGURE 5.22 System architecture in the central office.

in the central office to fulfill this purpose. Behind the OLTs, a network terminal (NT) is devised to manage multiple TDM-PONs. In the NT, the system bandwidth management module monitors the buffer depth of each OLT continually for downstream traffic. When any PON is heavily loaded, reflected by the OLT buffer depth exceeding a certain threshold, the system bandwidth management module will instruct the heavily loaded PON to deregister some ONUs and reregister them to lightly loaded PON(s). The number of ONUs to be moved depends on the average loading among PONs, which are monitored continuously by the traffic estimators (TEs). Before an ONU is deregistered, its packets queued in the OLT will be emptied. After an ONU is deregistered, the incoming traffic to that ONU is stored temporarily in the queue of NT until reregistration is complete. Note that the aforementioned ONU deregistration and registration can readily be realized by MPCPDU messages (deregistration and registration processes) and PLOAM messages (deactivation and ONU activation procedures) for EPON and GPON, respectively.

As applied to optical access networks, tunable transceivers have been proposed for different purposes, such as bandwidth efficiency improvement [52], network scalability [53], and inventory simplification [54]. Due to the rapid advance of optical technology, various tunable laser technologies have been developed [54]. To optimize the cost issue, one of the promising solutions is tunable long-wavelength vertical-cavity surface-emitting lasers (VCSELs) [55]. Lower cost results from integrated manufacturing, easy packaging, and testing. Recently, an integrated fast wavelength-selective photodetection was developed [56]. Its tuning time is in the nanosecond range, and the monolithic design facilitates cost reduction.

With slowly tunable transceivers, the tuning time can be a major overhead drain during reconfiguration, leading to network performance degrading. For example, as the traffic loading of a PON oscillates, reconfiguration will be triggered frequently and the network efficiency decreases. To minimize performance degradation, the buffer size can be increased with the higher threshold of a reconfiguration trigger. As reconfiguration is triggered, the tuning time will take a relatively small fraction of overhead

compared to the time needed to empty the packets stored in the OLT buffer, and the network efficiency could be improved. With this approach, although the efficiency is improved, packets will suffer more delay from a longer queuing time in the buffer, and the QoS degrades. To relieve QoS degradation, traffic differentiation (i.e., first forwarding high-priority traffic) is a solution to enhance the QoS during reconfiguration.

5.2.3.1 Performance Simulation of the Reconfigurable Optical Backhaul The performance of the proposed reconfigurable architecture is compared with fixed architectures through simulations. Assume that both the reconfigurable and fixed architectures consist of two EPONs and that the aggregated traffic is Poisson traffic. First we simulate the average packet delay of the two PONs in both architectures. In the simulation, (1) the average load of PON2 is fixed at 0.2 and that of PON1 is changed from 0 to 1.2 at $t = 1$ s; (2) the buffer size in the OLT is 128 MB [57]; (3) the network reconfiguration is triggered as the OLT buffer depth exceeds 50% (buffer depth refers to the total queued bits/bytes in a buffer); and (4) total reconfiguration overhead is 250 ms (including ONU deregistration, transceiver tuning time, ONU registration, and propagation delays). The traffic estimator is realized by a low-pass filter (LPF) with an averaging window of 10 ms. Figure 5.23 shows the dynamic response to a variation in traffic load. The traffic allocation between two PONs is measured at P1 in Figure 5.22, the outputs of the NT card. As shown, the reconfigurable architecture is able to balance the overall traffic load between two PONs after T_{P1}, which is approximately

$$T_{\mathrm{P1}} \approx \frac{\mathrm{TH}}{D_{\mathrm{in}} - D_{\mathrm{out}}} + \mathrm{OV} \tag{5.1}$$

where TH is the buffer depth threshold triggering reconfiguration, D_{out} the OLT data rate, D_{in} the traffic rate allocated to the OLT, and OV the reconfiguration overhead (i.e., tunable devices).

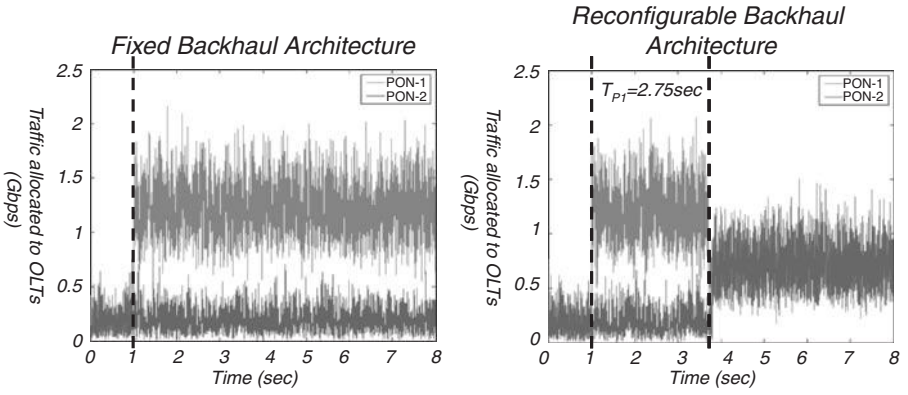


FIGURE 5.23 Traffic throughput at P1 in Figure 5.23.

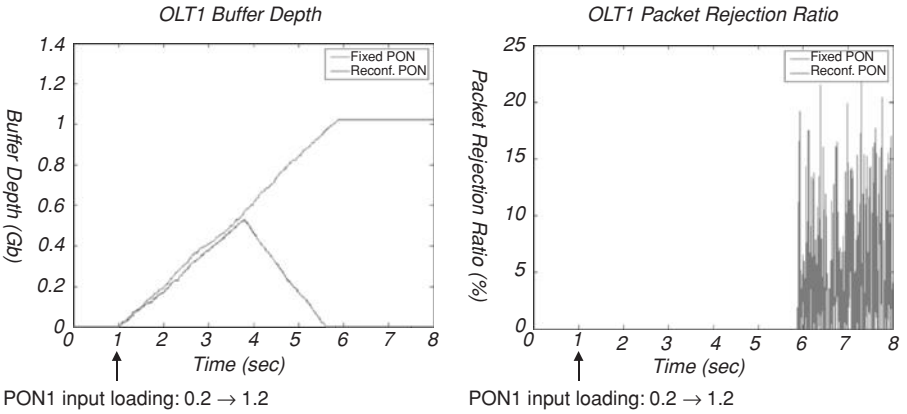


FIGURE 5.24 Buffer depth and packet rejection ratio of PON1 in both architectures.

Figure 5.24 shows the buffer depth and packet loss of PON1 in both architectures. After a reconfiguration period, the buffer depth decreases in PON1 of the reconfigurable architecture due to load balancing, while that of the fixed architecture keeps increasing, which eventually results in packet loss. The estimated maximum buffer depth of PON1,

$$\mathrm{BD}_{\mathrm{MAX}} = T_{\mathrm{P1}} \times (D_{\mathrm{in}} - D_{\mathrm{out}}) \tag{5.2}$$

Figure 5.25 shows the throughput of each PON and the combined throughput of both architectures, measured at P2 in Figure 5.22. The results show that the combined

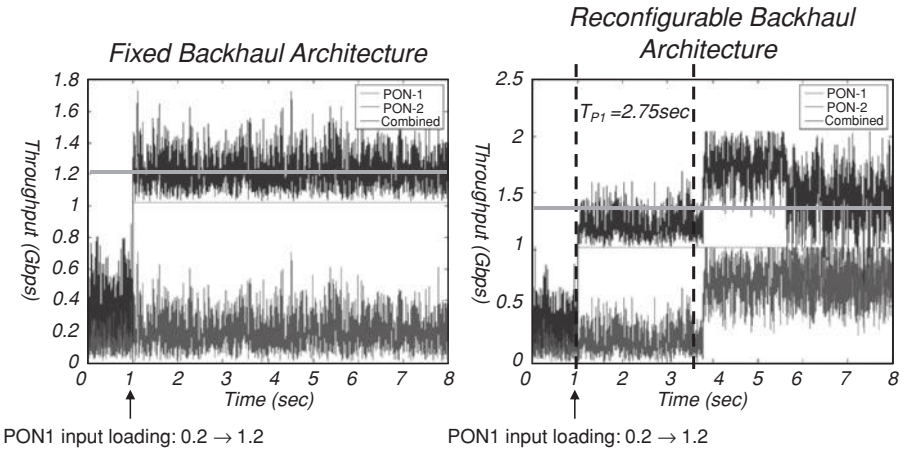


FIGURE 5.25 Throughput of each PON and combined throughput of both architectures; measured at P2 in Figure 3.4.

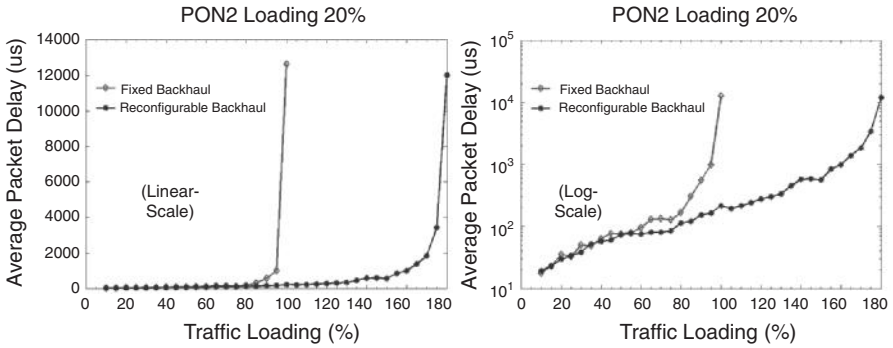


FIGURE 5.26 Long-term average packet delay of both PONs (varying load for PON1 and fixed load for PON2).

throughput of both PONs will reach the maximum rate to dequeue the stored packets in OLT1, and the required time, T_{P2}, is approximately

$$T_{P2} \approx T_{P2} + \frac{\mathrm{BD}_{\mathrm{MAX}}}{D_{\mathrm{out}} - D_{\mathrm{in,new}}} \tag{5.3}$$

where $D_{\mathrm{in,new}}$ is the new traffic rate allotted to OLT1.

Figure 5.26 shows the long-term average packet delay of two PONs in log and linear scales with different loads for PON1 and a fixed load for PON2. According to the linear scale, the fixed architecture's performance reaches its limit as the load reaches 90%, whereas the reconfigurable system does not reach its limit until the overall load reaches 180%. On the log scale, after passing 90%, the reconfigurable backhaul is insensitive to load increasing due to reconfiguration.

5.2.3.2 Experimental Testbed of Reconfigurable Optical Backhaul

After reviewing the performance improvement via dynamic load balancing realized by reconfigurable backhaul, it is desirable to investigate its hardware feasibility, performance, and compatibility to TDM-PON technologies. For these purposes, an experimental testbed (Figure 5.27) was built [51]. It consists of two OLTs equipped with fixed optical transceivers and an ONU equipped with a tunable transceiver. The functionality performed by an OLT and ONUs are programmed in a 1.25-Gb/s field-programmable gate array (FPGA). To facilitate the reconfiguration process, reconfiguration control interfaces (RCI) are implemented in the FPGA at the ONU and the central office as shown in Figure 5.27. After reconfiguration is triggered, the RCI in the central office will direct that some ONUs be moved from a heavily loaded PON (e.g., PON1) to a lightly loaded PON (e.g., PON2). The process includes the following steps.

1. The RCI at the central office sends new wavelength information [e.g., $\lambda_{2d,u}$ of the lightly loaded OLT (i.e., OLT2) to the heavily loaded OLT (i.e., OLT1)],

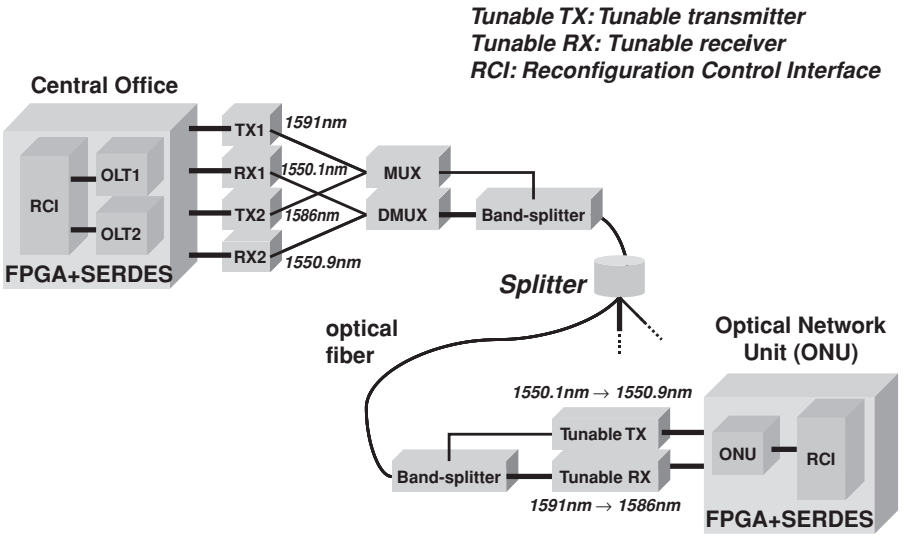
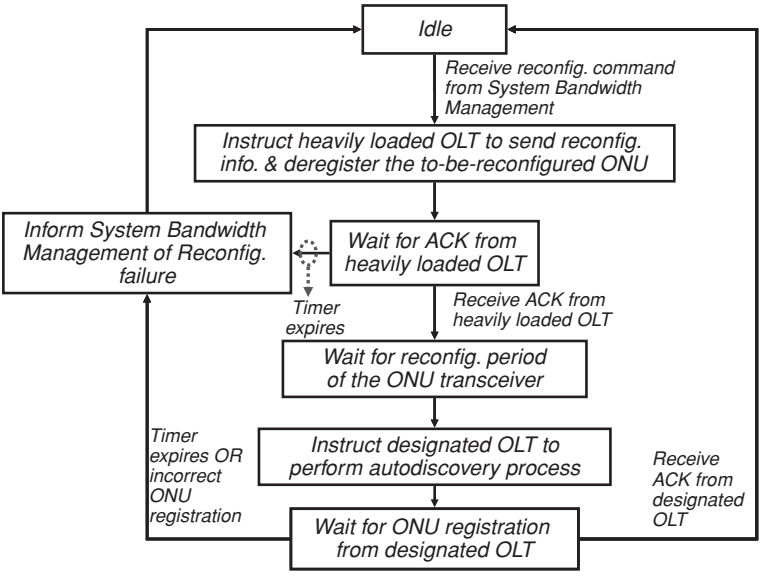


FIGURE 5.27 Reconfiguration experimental testbed.

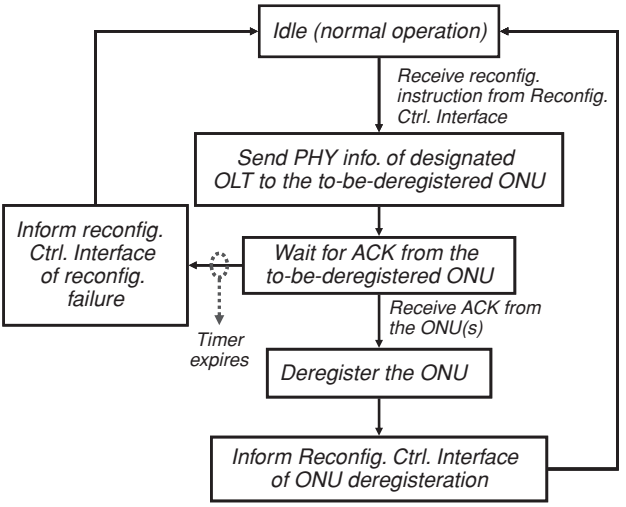
and the OLT will arrange the next available data packet to deliver the new wavelength information to the ONU to be deregistered. (Note that this information can be piggybacked on data packets as specified by the TDM PON standards.)

2. The RCI at the central hub instructs the heavily loaded OLT (e.g., OLT1) to deregister the to-be-reconfigured ONU.
3. The RCI behind the ONU stores the new wavelength information, generates an acknowledgment (ACK) signal, and passes it on to the ONU for transmission. The ONU will then arrange that the next available data packet delivers the ACK.
4. Upon reception of the ACK signal from upstream packets, the heavily loaded OLT passes it to the RCI at the central office. The RCI will proceed to instruct the heavily loaded OLT to start the standard deregistration process specified in TDM-PON standards [1,2].
5. After the deregistration ACK is delivered by the ONU, the RCI at the ONU will tune the wavelengths of the tunable transceiver to the new wavelengths of the designated OLT (e.g., from $\lambda_{1d,u}$ to $\lambda_{2d,u}$).
6. After receiving the deregistration ACK from the ONU, the RCI at the central office will wait for a certain amount of time (the waiting time is preset according to the tuning time of the optical tunable transceiver at the ONU) and instruct the designated OLT (e.g., OLT2) to start the standard discovery process defined in TDM-PON standards [1,2] to discover and register the new ONU.

To realize and examine the handshaking protocol, Figure 5.28 summarizes the state diagrams of the RCIs at the central hub, the heavily and lightly loaded OLTs, the

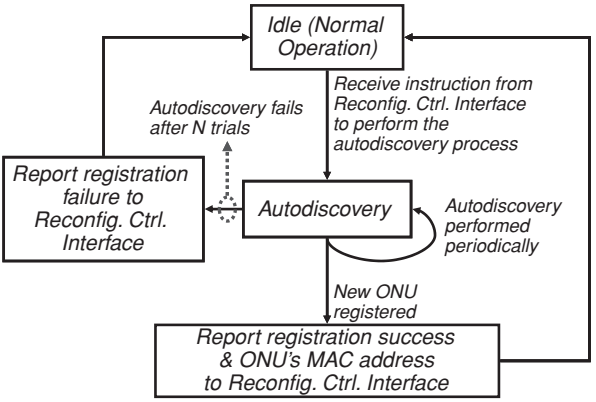


(a) State diagram of the reconfiguration control interface at the central office

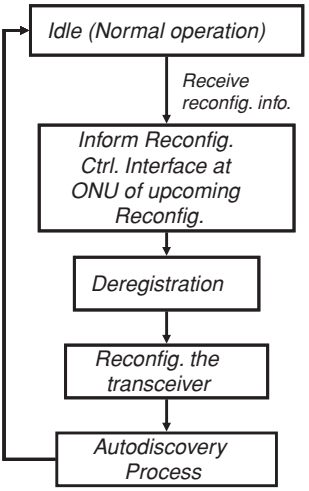


(b) State diagram of the heavily loaded OLT

FIGURE 5.28 State diagrams of the RCIs in the central office, heavily and lightly loaded OLTs, an ONU, and RCI behind the ONU.



(c) State diagram of the designated lightly loaded OLT

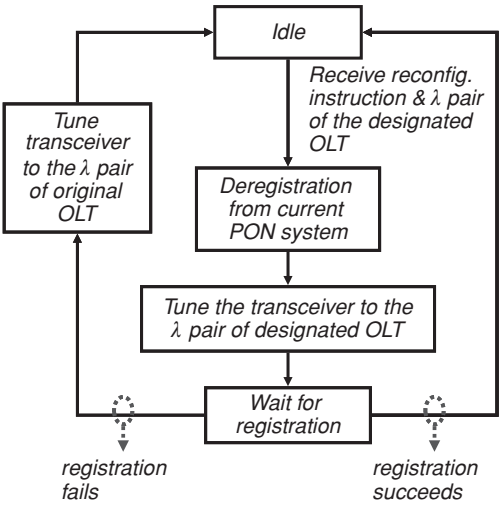


(d) State diagram of the reconfigured ONU

FIGURE 5.28 (*Continued*) State diagrams of the RCIs in the central office, heavily and lightly loaded OLTs, an ONU, and RCI behind the ONU.

reconfigurable ONU, and the RCI at the reconfigurable ONU, which are implemented in commercial FPGA chips. Note that the discovery, deregistration, and registration processes are programmed according to the IEEE 802.3ah EPON standards.

The tunable transmitter at ONU is the same as the one used in reference 58, which tunes from 1552.52 to 1550.92 nm within 80 ns. The tunable receiver at the ONU is implemented using an MEMS tunable filter, which has a wide tuning range: from 1591 to 1525 nm by changing the control voltage from 0 to 35 V. Throughout the tuning range, the insertion loss is less than 1 dB. To examine the transient response of the tunable filter, a light at 1586 nm is transmitted continuously to the tunable



(e) State diagram of the reconfiguration control interface at the reconfigured ONU

FIGURE 5.28 (*Continued*) State diagrams of the RCIs in the central office, heavily and lightly loaded OLTs, an ONU, and RCI behind the ONU.

filter, and the tunable filter tunes to receive between 1591 and 1586 nm by changing its control voltage. As shown in Figure 5.29, after the control voltage is changed, it takes 33.6 μs for the filter to stabilize to receive 1596-nm light. Since the tuning time is much longer than that of the tunable transmitter, the reconfiguration overhead incurred in the PHY layer is dominated by the tunable receiver. To accommodate the tuning time, the period between deregistration (step 2: deregister ONU from heavily loaded PON) and reregistration (step 4: reregister the ONU to lightly loaded PON) issued by an RCI at the central office is thus programmed as 50 μs in FPGA.

5.2.3.3 Control Packet Format To facilitate the communication of an RCI on the ONU and central office sides, a control packet is used on the testbed and its format is shown in Figure 5.30. It consists of the following parts:

1. Preamble (60 bytes): for clock and level recovery; specifically to facilitate upstream reception
2. Sync bytes (4 bytes): delimiter for bit synchronization
3. Frame ID: for statistics and error-checking purposes
4. Tx: sender ID
5. Rx: receiver ID
6. Frame type: indicates the type of MAC frame
7. Payload: carries RCI commands such as deregistration, ACK, new wavelength information, etc.

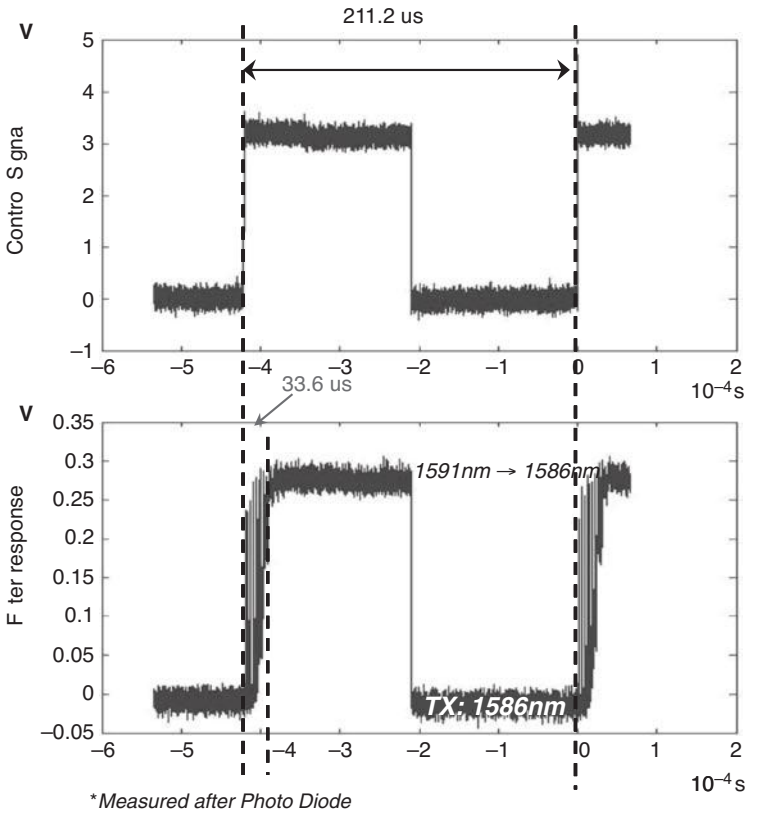


FIGURE 5.29 Dynamic response of the tunable receiver.

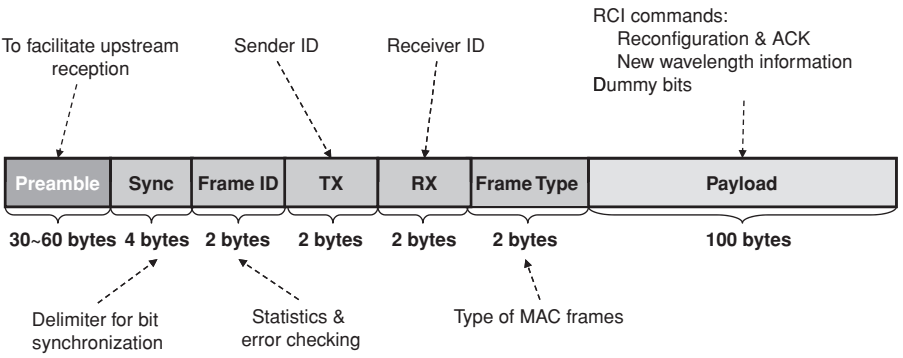


FIGURE 5.30 Control frame format.

5.2.3.4 Experimental Results The experimental result is shown in Figure 5.31, where the ONU deregistration from the OLT1, reconfiguration period, and the ONU discovery and registration performed by the OLT2 are demonstrated. Note in Figure 5.31 that the propagation delay is insignificant compared to the reconfiguration period.

The signals marked with numbers in Figure 5.31 are outputs measured on the FPGA boards, which indicate the following events:

1. OLT1 sends a deregistration message.
2. ONU receives the deregistration message.
3. ONU sends an ACK.
4. OLT1 receives the ACK.
5. After a 50-μs reconfiguration period, OLT2 sends a discovery gate message.
6. ONU receives the discovery gate message.
7. ONU sends a registration request.
8. OLT2 receives the registration request.
9. OLT2 sends a registration message.
10. ONU receives the registration message.
11. ONU sends a registration ACK.
12. OLT2 receives the registration ACK.

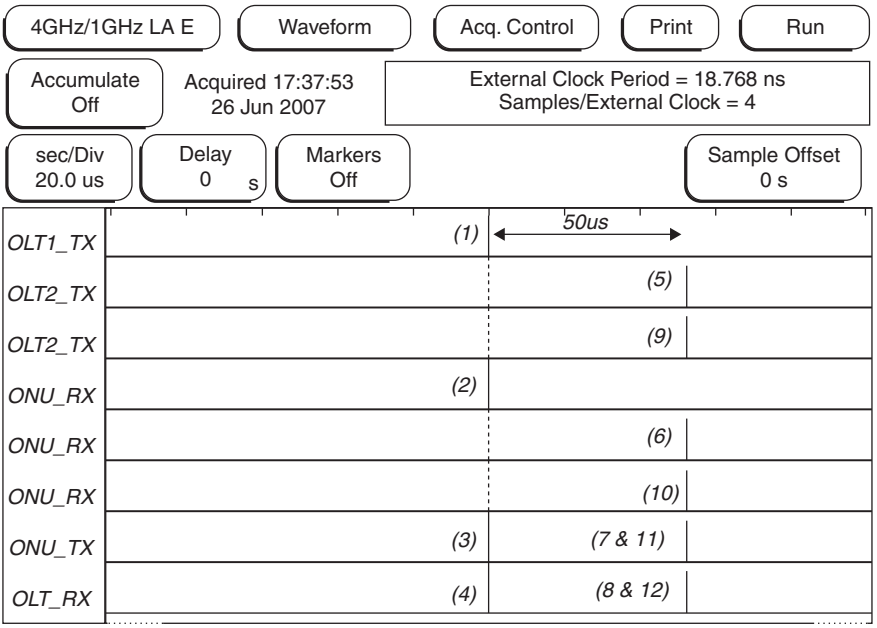


FIGURE 5.31 Experimental results of a reconfiguration testbed.

5.3 INTEGRATED ROUTING ALGORITHM FOR HYBRID ACCESS NETWORKS

In Section 5.1.2.4 we mentioned that most routing protocols for WMNs are derived from wireless ad hoc networks, which is different from WMNs in terms of node mobility, power constraint, and computation capability. In addition, the traffic pattern in a WMN is very different from that of wireless ad hoc networks because most, if not all, traffic in a WMN is forwarded between the gateway routers and end users, as in other access networks [30]. Hence, conventional proactive and reactive routing protocols developed for wireless ad hoc networks need to be modified if they were to be applied to WMNs for optimized performance.

On the hybrid optical wireless access network, there usually exist multiple routing paths between the central hub and an end user because of the point-to-multipoint topologies of the optical backhaul and WMNs. Different routing paths consist of different optical and wireless links. In determining an end-to-end routing path, the wireless link generally has a higher routing cost than that of its optical counterpart, because wireless links interfere with each other and share limited bandwidth resource in an open space. Although optical backhaul does not play a critical role in routing cost, it can help to improve the overall routing performance. Assuming that a wireless gateway can derive network conditions in its neighborhood, such as link status, interference, and traffic loading of its neighboring wireless mesh routers, all the information from different gateways can readily be collected by the central office through the optical backhaul. Then the central office can use the information from gateways to optimize packet routing on a WMN. In other words, with the help of the optical backhaul, a centralized system management is realized and results in a centralized routing paradigm.

Based on the characteristics of hybrid optical wireless network, an integrated routing algorithm has been proposed [51]. This algorithm computes the optimum route based on the up-to-date wireless link state and average traffic rate. For any wireless mesh router, this algorithm will select the optimum route, composed of wireless links, wireless gateway/ONU, and optical links, under given network conditions. Specifically, this algorithm will adapt to each router's traffic and perform dynamic load balancing among WMNs within a district. The integrated routing paradigm proposed is described by the following steps, and the corresponding operations in the system are illustrated in Figure 5.32.

1. *Wireless link state update*. Each wireless mesh router probes the link states periodically using neighboring routers. The probing can be done by measuring the retransmission ratio [25] or transmission loss rate [24] in both directions. The measurement verifies connectivity and reflects the quality and capacity of wireless links that vary over time. Each wireless mesh router then broadcasts the link states in a bounded flooding manner; that is, the packet can propagate on a WMN for only a limited number of hops, as illustrated in Figure 5.33. The hop-count limit, $H_{\max}$, is assigned according to the (wireless gateway router)/(wireless mesh router) ratio: A small ratio will require a large

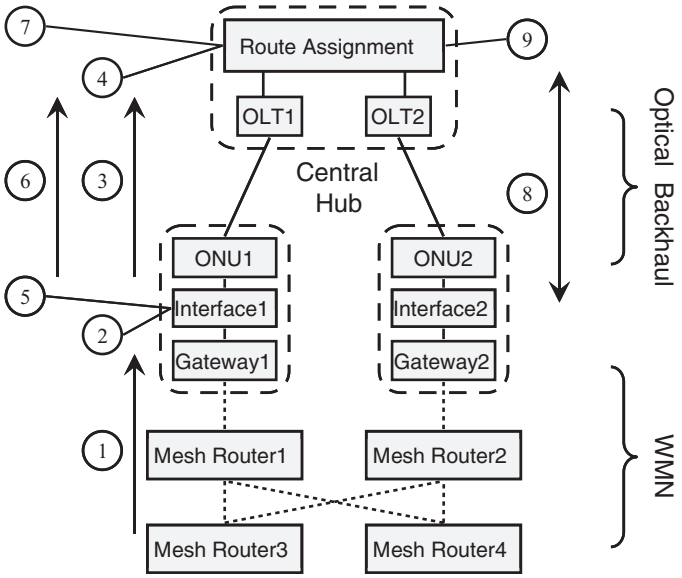


FIGURE 5.32 Integrated routing paradigm in an exemplar system.

hop-count limit. As such, only the nearby wireless gateway routers will receive the broadcast from each mesh router. Compared to wireless ad hoc networks, the overhead of the link-state update under the hybrid architecture is significantly lower because the fixed infrastructure and network engineering lead to less channel variation. A high gateway/router ratio will further reduce the overhead by limiting the propagation of update broadcast on the WMN.

2. *Local WMN route calculation*. Based on the link-state updates, the interface between the optical terminal and wireless gateway router calculates the optimum route for each mesh router within the hop-count limit in both the downstream

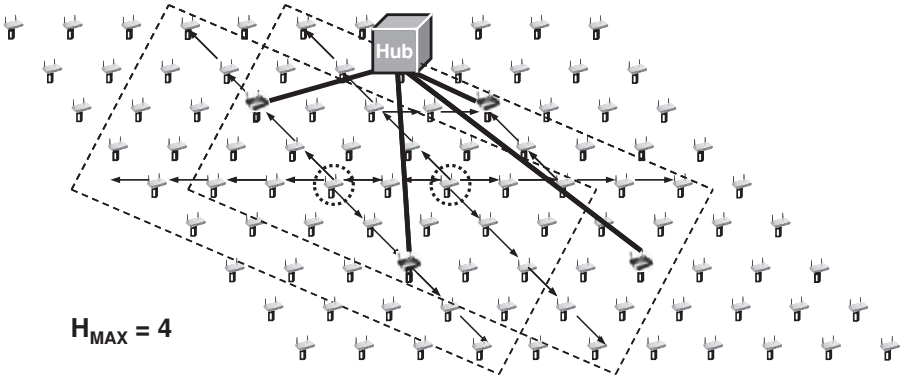


FIGURE 5.33 Bounded flooding of link-state information.

and upstream directions using a shortest-path algorithm with the link states as cost [24,25].

3. *Route cost report*. Each optical terminal reports to the central hub the calculated route cost in both directions for each wireless mesh router within $H_{\max}$.
4. *Gateway association*. After receiving the reports from all optical terminals, the route assignment module will associate every wireless mesh router with the gateway router that has the lowest cost. The comparison result between a wireless mesh router and different wireless gateway routers is stored for future reference to enable load balancing.
5. *Congestion monitoring*. At the interface between an optical terminal and a wireless gateway router, a low-pass filter is dedicated to each associated wireless mesh router to measure the average flow rates in both directions. Based on the measurement, a capacity table is continuously updated to monitor the overall loading of the local WMN. To calculate the capacity, the interference in the local WMN can be measured using a technique proposed by Padhye et al. [59]. If loading at a certain area exceeds a congestion threshold, the interface will locate the k furthermost router(s) $\mathrm{RT}\{1,2,\ldots,k\}$, of which the flows pass the hot zone and can reduce the congestion sufficiently if the flows are removed. If some routers among $\mathrm{RT}\{1,2,\ldots,k\}$ exceed a minimum hop-count threshold $H_{\min}$, these routers will be categorized to $\mathrm{RT}_{\text{flow control}}\{1,2,\ldots,i\}$ and the rest of the routers to $\mathrm{RT}_{\text{load balancing}}\{1,2,\ldots,j\}$; note that $i + j = k$.
6. *Congestion report*. The interface that detects congestion (e.g., interface 1 in Figure 4.21) sends a congestion report, including $\mathrm{RT}_{\text{load balancing}}\{1,2,\ldots,j\}$ and $\mathrm{RT}_{\text{flow control}}\{1,2,\ldots,i\}$, to the route assignment module.
7. *Alternative gateway router look-ups*. For each router in $\mathrm{RT}_{\text{load balancing}}\{1,2,\ldots,j\}$, the route assignment module checks the gateways that received a link-state update (e.g., interface 2 in Figure 4.21) based on the stored comparison result described in step 4.
8. *Wireless gateway router reassociation*. The reassociation begins with gateways with the lowest cost. The interfaces of the gateway will then check whether the loading threshold will be exceeded if the flows are added. If the threshold will be exceeded, the route assignment module will negotiate with the gateway with the second lowest cost, and so on. If routers among $\mathrm{RT}_{\text{load balancing}}\{1,2,\ldots,j\}$ fail to be associated with a new gateway, they will be included in $\mathrm{RT}_{\text{flow control}}$.
9. *Flow control*. After $\mathrm{RT}_{\text{load balancing}}\{1,2,\ldots,j\}$ are reassociated, flow control will be executed in $\mathrm{RT}_{\text{flow control}}$.

5.3.1 Simulation Results and Performance Analysis

The performance improvement achieved by the proposed algorithm implements simulation with NS-2 [60]. The simulation scenario consists of four gateways and 192 mesh routers as shown in Figure 5.34, where the distance between any two adjacent nodes is 100 m. An IEEE 802.11b module implemented in NS-2 is used, and the data

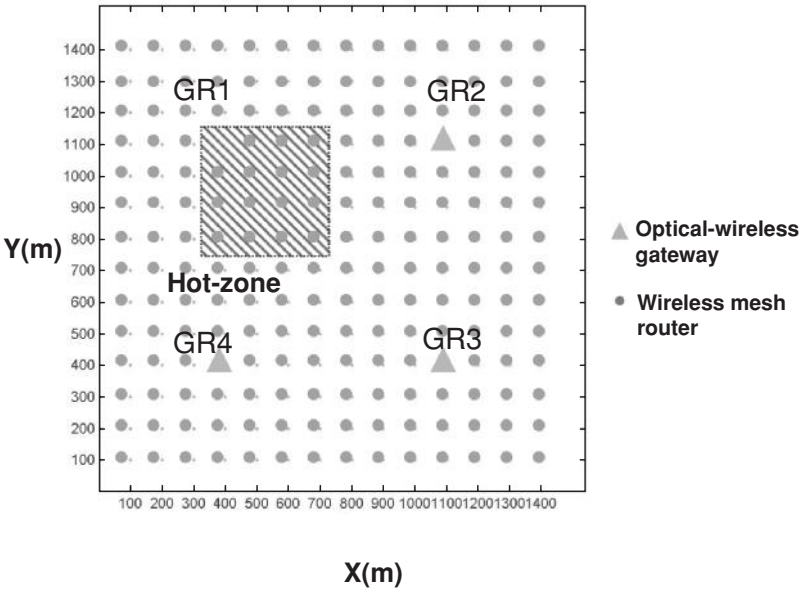


FIGURE 5.34 Bounded flooding of link-state information.

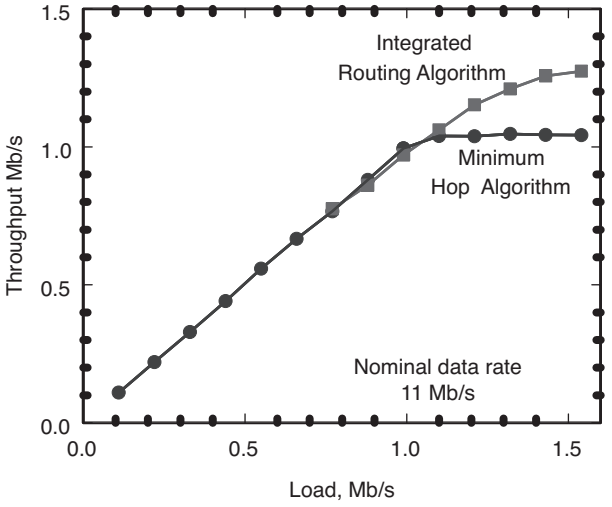


FIGURE 5.35 (a) Average throughput and packet delay comparison.

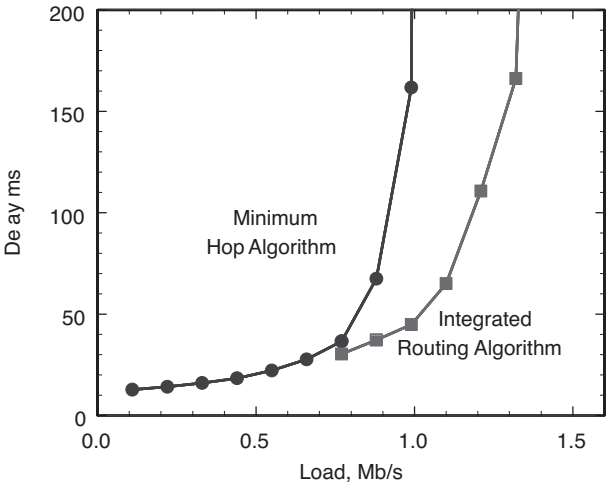


FIGURE 5.35 (*Continued*) (b) Average throughput and packet delay comparison.

rate is up to 11 Mb/s. Poisson random traffic and a two-way radio propagation model are used throughout the simulation. It is assumed that every mesh router has uniform transmission power and receiver sensitivity. By applying a range model described in the literature [42,43], we assume that the transmission range and the interference ranges equal 120 and 180 m, respectively. Each node has only one radio interface and is equipped with an omnidirectional antenna. Note that in this simulation scenario, the shortest-path routing with the link state as cost is reduced to minimum-hop routing. We inject uniform background traffic to each mesh router, and the average flow rate out of each gateway to the nearby 48 mesh routers is 1% of the data rate. Assume that there is a hot zone, as shown Figure 5.34. Additional loadings are injected on routers in the hot zone and increase the overall loading from 0.1 Mb/s to 1.5 Mb/s. The congestion threshold is set to be 8% (0.88 Mb/s) of the data rate (i.e., after the flow rate of a certain router exceeds 0.88 Mb/s, the interface will detect congestion and send a congestion report to the central hub). The throughput and packet delay performance of mesh routers in the hot zone is investigated. The simulation results in Figure 5.35(a) and (b) show that after the additional loading in the hot zone exceeds 0.77 Mb/s (in the presence of 0.11 Mb/s background traffic), flows to the boundary routers are shifted to other three gateways to balance the loading based on the algorithm proposed. As a result, the throughput and delay are improved by about 25%.

5.4 SUMMARY

In this chapter we introduce a hybrid optical–wireless access network. To address different challenges simultaneously, a hybrid network should be broadband, ubiquitous, scalable, and cost-effective. Both optical and wireless segments complement

its counterpart to fulfill these requirements. We review WiMAX technology to date and a wireless mesh network that enables the wireless segment of the hybrid architecture. We also review a hybrid optical–wireless network architecture that consists of reconfigurable optical backhaul and wireless mesh network. The reconfigurable TDM/WDM optical backhaul aims to address the scalability issue in WMN and upgrade for hierarchical wireless access networks. TDM-PON technology is leveraged due to its scalable MAC, flexible architecture, cost-effectiveness, and technological maturity. Reconfigurability is implemented in the optical backhaul for bandwidth reallocation to improve resource utilization. The experimental testbed demonstrates its feasibility. An integrated routing algorithm is developed for this hybrid architecture. This algorithm takes advantage of the optical backhaul to collect a real-time loading situation and network conditions and then locate the optimum ONU/gateway and route in a WMN. Simulation results show that throughput and delay improvement result from load balancing among multiple gateways. Although hybrid optical–wireless access is a green research field, it is a practical solution to addressing the challenges encountered by different solutions in the last-mile connection. We envision that more research effort will need to be directed to the network architecture, enabling device, integrated routing algorithm, QoS, resilience, and multicast in a hybrid optical–wireless access network in the near future.

REFERENCES

1. ITU-T G.984.
2. IEEE 802.3ah.
3. IEEE Standard 802.16-2004, *Air Interface for Fixed Broadband Wireless Access Systems*.
4. IEEE Standard 802.16e, *Air Interface for Fixed and Mobile Broadband Wireless Access Systems: Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands*.
5. A. Bahai and B. Saltzbert, *Multi-carrier Digital Communication: Theory and Applications of OFDM*, Kluwer Academic, Norwell, MA, 1999.
6. D. Love and R. Heath, OFDM power loading using limited feedback, *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, Sept. 2005.
7. H. Moon, Efficient power allocation for coded OFDM systems, Ph.D. dissertation, Stanford University, Aug. 2004.
8. O. Awoniyi, O. Oteri, and F. Tobagi, Adaptive power loading in OFDM-based WLANs and the resulting performance improvement in voice and data applications, *Proceedings of the of IEEE Vehicular Technology Conference*, Sept. 2005, vol. 2, pp. 789–794.
9. Richard Van Nee and Ramjee Prasad, *OFDM for Wireless Multimedia Communications*, Artech House, Norwood, MA, 2000.
10. Arunabha Ghosh, David R. Wolter, Jeffrey G. Andrews, and Runhua Chen, Broadband wireless access with WIMAX current performance benchmarks and future potential, *IEEE Commun.*, vol. 43, no. 2, Feb. 2005, pp. 129–136.
11. G. Nair, J. Chou, T. Madejski, K. Perycz, P. Putzolu, and J. Sydir, IEEE 802.16 medium access control and service provisioning, *Intel Technol. J.*, vol. 8, Aug. 2004.

12. http://wirelessman.org/relay/index.html.
13. http://w3.antd.nist.gov/wctg/manet.
14. R. Bruno, M. Conti, and E. Gregori, Mesh networks: commodity multihop ad hoc networks, *IEEE Commun.*, Mar. 2005, pp. 123–131.
15. http://www.tropos.com/.
16. http://www.belairnetworks.com/.
17. I. Akyildiz, A survey on wireless mesh networks, *IEEE Radio Commun.*, Sept. 2005, pp. S23–S30.
18. H. Aoki, S. Takeda, K. Yaggu, and A. Yamada, *IEEE 802.11a Wireless LAN Mesh Network Technology*, NTT DoCoMo Technical Journal, vol. 8, no. 2, Sept. 2006, pp. 13–21.
19. J. Bicket, D. Aguayo, S. Biswas, and R. Morris, Architecture and evaluation of an unplanned 802.11b mesh network, presented at MOBICOM 2005, Aug. 2005.
20. http://www.ieee802.org/16/tgm/.
21. T. Clausen and P. Jaquet, *Optimized Link State Routing Protocol* (*OLSR*), RFC 3626, Oct. 2003.
22. E. Perkins, E. Belding-Royer, and S. Das, *Ad Hoc on Demand Distance Vector* (*AODV*) *Routing*, RFC 3561, July 2003.
23. R. Draves, J. Padhye, and B. Zill, Comparisons of routing metrics for static multi-hop wireless networks, *ACM Special Interest Group on Data Communications* (*SIGCOMM*), Aug. 2004, pp. 133–144.
24. R. Draves, J. Padhye, and B. Zill, Routing in multi-radio, multi-hop wireless mesh networks, *ACM International Conference on Mobile Computing and Networking* (*MOBICOM*), 2004, pp. 114–128.
25. D. Couto, D. Aguayo, J. Bicket, and R. Morris, A high-throughput path metric for multihop wireless routing, *Wireless Networks*, vol. 11, no. 4, July 2005.
26. L. Iannone and S. Fdida, MRS: a simple cross-layer heuristic to improve throughput capacity in wireless mesh networks, presented at the Conference on Future Networking Technologies (CoNEXT), Oct. 2005.
27. C. Perkins and P. Bhagwat, Highly dynamic destination-sequenced distance-vector (DSDV) routing for mobile computers, *Comput. Commun. Rev.*, Oct. 1994, pp. 234–244.
28. D. Johnson and D. Maltz, Dynamic source routing in ad hoc wireless networks, in *Mobile Computing*, vol. 353, pp. 153–181, Kluwer Academic, Norwell, MA, 1996.
29. P. Gupta and P. Kumar, The capacity of wireless networks, *IEEE Trans. Inf. Theory*, vol. 46, no. 2, Mar. 2002, pp. 388–404.
30. J. Jun and M. L. Sichitiu, The nominal capacity of wireless mesh networks, *IEEE Wireless Commun.*, Oct. 2003, pp. 8–14.
31. H. Hoffman and D. Cox, Attenuation of 900 MHz radio waves propagating into a metal building, *IEEE Trans. Antennas Propag.*, vol. 30, no. 4, pp. 808–811, July 1982.
32. J. Durante, Building penetration loss at 900 MHz, *Proceedings of the IEEE Vehicular Technology Group*, 1973.
33. P. Wells, The attenuation of UHF radio signals by houses, *IEEE Trans. Veh. Technol.*, vol. VT-26, no. 5, Nov. 1977, p. 358.
34. D. Cox, R. Murray, and A. Norris, 800 MHz attenuation measured in and around suburban houses, *Bell Lab. Tech. J.*, vol. 63, July–Aug. 1984, pp. 921–954.

35. D. Cox, R. Murray, and A. Norris, Measurements of 800 MHz radio transmission into buildings with metallic walls, *Bell Lab. Tech. J.*, vol. 62, Nov. 1983, pp. 2695–2717.
36. D. Cox, R. Murray, and A. Norris, Antenna height dependence of 800 MHz attenuation measured in houses, *IEEE Trans. Veh. Technol.*, vol. VT-34, May, 1985, pp. 108–115.
37. H. Arnold, D. Cox, and R. Murray, Macroscopic diversity performance measured in the 800-MHz portable radio communications environment, *IEEE Trans. Antennas Propag.*, vol. 36, no. 2, Feb. 1988.
38. W. Smith, Urban propagation modeling for wireless systems, Ph.D. dissertation, Stanford University, Feb. 2004.
39. M. Feuerstein et al., Path loss, delay spread, and outage models as functions of antenna height for microcellular system design, *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, Aug. 1994, pp. 487–498.
40. J. Berg, R. Bownds, and F. Lotse, Path loss and fading models for microcells at 900 MHz, *Veh. Tech. Conf. Rec.*, May 1992, pp. 666–671.
41. A. Goldsmith and L. Greenstein, A measurement-based model for predicting coverage areas of urban microcells, *IEEE J. Sel. Areas Commun.*, vol. 11, no. 7, Sept. 1993, pp. 1013–1023.
42. J. Broch et al., A performance comparison of multi-hop wireless ad-hoc network routing protocols, *Proceedings of ACM/IEEE MOBICOM'98*, Dallas, TX, 1998.
43. J. Li, C. Blake, D. S. J. De Couto, H. I. Lee, and R. Morris, Capacity of ad hoc wireless networks, *Proceedings of ACM SIGMOBILE*, Rome, Italy, July 2001.
44. http://www.sfgov.org/site/tech_connect_page.asp.
45. http://motorola.canopywireless.com/.
46. http://www.connectronics.com/proxim/GigaLink.htm.
47. http://www.mrv.com/products/line/terescope.php.
48. A.M.J. Koonen, M. García Larrodé, A. Ng'oma, K. Wang, H. Yang, Y. Zheng, and E. Tangdiongga, *Perspectives of Radio over Fiber Technologies*, OThP3 OFC/NFOEC, 2008.
49. G. Kramer and G. Pesavento, Ethernet passive optical network (EPON): building a next-generation optical access network, *IEEE Commun.*, vol. 40, no. 2, Feb. 2002, pp. 66–73.
50. W.-T. Shaw, S.-W. Wong, N. Cheng, and L. G. Kazovsky, *MARIN Hybrid Optical-Wireless Access Network*, OFC 2006 OThM3.
51. W.-T. Shaw, S.-W. Wong, N. Cheng, K. Balasubramanian, X. Zhu, M. Maier, and L. G. Kazovsky, Hybrid architecture and integrated routing in a scalable optical-wireless access network, *J. Lightwave Technol.*, vol. 25, no. 11, Nov. 2007, pp. 3443–3451.
52. T. Chen, H. Woesner, Y. Ye, and I. Chlamtac, Wireless gigabit ethernet extension, *Broadband Networks*, vol. 1, 2005, pp. 425–433.
53. Y.-L. Hsueh, W.-T. Shaw, L. G. Kazovsky, A. Agata, and S. Yamamoto, SUCCESS PON demonstrator: experimental exploration of next-generation optical access networks, *IEEE Opt. Commun.*, Aug. 2005, pp. S26–S33.
54. Jens Buus, Tunable lasers in optical networks, *J. Lightwave Technol.*, vol. 24, no. 1, Jan. 2006.
55. C. J. Chang-Hasnain, Tunable VCSEL, *IEEE J. Sel. Top. Quantum Electron.*, vol. 6, no. 6, Nov. 2006.

56. R. Chen, D. A. B. Miller, K. Ma, and J. S. Harris, Jr., Novel electrically controlled rapidly wavelength selective photodetection using MSMs, *IEEE J. Sel. Top. Quantum Electron.*, vol. 11, no. 1, Jan.–Feb. 2005, pp. 184–189.
57. http://www.teknovus.com/files/TK3722%20PB1.pdf.
58. W.-T. Shaw et al., MARIN: metro-access ring integrated network, presented at GLOBECOM 2006, San Francisco, Nov. 2006.
59. J. Padhye, S. Agarwal, V. N. Padmanabhan, L. Qiu, A. Rao, and B. Zill, Estimation of link interference in static multi-hop wireless networks, *Proceedings of the 2005 Internet Measurement Conference*, pp. 305–310.
60. http://www.isi.edu/nsnam/ns/.

INDEX

Broadband Optical Access Networks, First Edition. Leonid G. Kazovsky, Ning Cheng, Wei-Tao Shaw, David Gutierrez, and Shing-Wa Wong.