

Advanced Optical Wireless Communication Systems

Optical wireless communications is a dynamic area of research and development. Combining fundamental theory with a broad overview, this book is an ideal reference for anyone working in the field, as well as a valuable guide for self-study. It begins by describing important issues in optical wireless theory, including coding and modulation techniques for optical wireless, wireless optical CDMA communication systems, equalization and Markov chains in cloud channels, and optical MIMO systems, as well as explaining key issues in information theory for optical wireless channels. The next part describes unique channels that could be found in optical wireless applications, such as NLOS UV atmospheric scattering channels, underwater communication links, and a combination of hybrid RF/optical wireless systems. The final part describes applications of optical wireless technology, such as quantum encryption, visible light communication, IR links, and sensor networks, with step-by-step guidelines to help reduce design time and cost.

Shlomi Arnon is an Associate Professor at the Department of Electrical and Computer Engineering at Ben-Gurion University (BGU), Israel, and the Principal Investigator of Israel Partnership with NASA LUNAR Science Institute. In addition to research, Professor Arnon and his students work on many challenging engineering projects with emphasis on the humanitarian dimension, such as developing a system to detect human survival after earthquakes, or an infant respiration monitoring system to prevent cardiac arrest and apnea.

John R. Barry is a Professor of Telecommunications in the School of Electrical and Computer Engineering at the Georgia Institute of Technology. He is a coauthor of *Digital Communication* (2004), and *Iterative Timing Recovery: A Per-Survivor Approach* (VDM, 2009), and he is the author of *Wireless Infrared Communications* (1994).

George K. Karagiannidis is an Associate Professor of Digital Communications Systems in the Electrical and Computer Engineering Department, and Head of the Telecommunications Systems and Networks Laboratory, at Aristotle University of Thessaloniki. He is co-recipient of the Best Paper Award of the Wireless Communications Symposium (WCS) in the IEEE International Conference on Communications (ICC'07).

Robert Schober is a Professor and Canada Research Chair in Wireless Communications at the University of British Columbia (UBC), Vancouver, Canada. He has received numerous awards, including best paper awards from the German Information Technology Society (ITG), the European Association for Signal, Speech and Image Processing (EURASIP), IEEE ICUWB 2006, the International Zurich Seminar on Broadband Communications, and European Wireless 2000.

Murat Uysal is an Associate Professor at Özyeğin University, Istanbul, where he leads the Communication Theory and Technologies (CT&T) Research Group. Dr. Uysal is the recipient of several awards including the NSERC Discovery Accelerator Supplement Award, University of Waterloo Engineering Research Excellence Award, and the TUBA Distinguished Young Scientist Award.

Advanced Optical Wireless Communication Systems

Edited by

SHLOMI ARNON

Ben-Gurion University (BGU), Israel

JOHN R. BARRY

Georgia Institute of Technology, USA

GEORGE K. KARAGIANNIDIS

Aristotle University of Thessaloniki, Greece

ROBERT SCHOBER

University of British Columbia (UBC), Canada

MURAT UYSAL

Özyeğin University, Turkey



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521197878

© Cambridge University Press 2012

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2012

Printed in the United Kingdom at the University Press, Cambridge

A catalog record for this publication is available from the British Library

ISBN 978-0-521-19787-8 hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

| | | |
|----------|---|---------------|
| | <i>List of contributors</i> | <i>page x</i> |
| | Part I Outlook | 1 |
| 1 | Introduction Shlomi Arnon, John Barry, George Karagiannidis, Robert Schober, and Murat Uysal | 3 |
| | Part II Optical wireless communication theory | 9 |
| 2 | Coded modulation techniques for optical wireless channels Ivan B. Djordjevic | 11 |
| | 2.1 Atmospheric turbulence channel modeling | 12 |
| | 2.2 Codes on graphs | 13 |
| | 2.3 Coded-MIMO free-space optical communication | 19 |
| | 2.4 Raptor codes for temporally correlated FSO channels | 26 |
| | 2.5 Adaptive modulation and coding (AMC) for FSO communications | 29 |
| | 2.6 Multidimensional coded modulation for FSO communications | 35 |
| | 2.7 Free-space optical OFDM communication | 38 |
| | 2.8 Heterogeneous optical networks (HONs) | 43 |
| | 2.9 Summary | 48 |
| | Acknowledgments | 49 |
| | References | 49 |
| 3 | Wireless optical CDMA communication systems Jawad A. Salehi, Babak M. Ghaffari, and Mehdi D. Matinfar | 54 |
| | 3.1 Introduction | 54 |
| | 3.2 OCDMA system description | 55 |
| | 3.3 Indoor wireless optical CDMA LAN | 59 |
| | 3.4 Free-space optical CDMA systems | 68 |
| | 3.5 Modulation | 75 |
| | 3.6 Experimental prototypes | 81 |

| | | |
|----------|---|------------|
| | Acknowledgment | 84 |
| | References | 84 |
| 4 | Pointing error statistics | 87 |
| | Shlomi Arnon | |
| | References | 89 |
| 5 | Equalization and Markov chains in cloud channel | 90 |
| | Mohsen Kavehrad | |
| | 5.1 Introduction | 91 |
| | 5.2 Channel propagation modeling | 92 |
| | 5.3 Modeling results and eigen analyses | 99 |
| | 5.4 Equalization related issues | 103 |
| | 5.5 Summary and conclusions | 112 |
| | Acknowledgment | 113 |
| | References | 113 |
| 6 | Multiple-input multiple-output techniques for indoor optical wireless communications | 116 |
| | Steve Hranilovic | |
| | 6.1 Indoor OW MIMO channel characteristics | 117 |
| | 6.2 MIMO for diffuse OW channels | 119 |
| | 6.3 Spot-diffusing OW MIMO systems | 123 |
| | 6.4 Point-to-Point OW MIMO communications | 127 |
| | 6.5 Future directions | 138 |
| | References | 139 |
| 7 | Channel capacity | 146 |
| | Amos Lapidoth, Stefan M. Moser, and Michèle Wigger | |
| | 7.1 Introduction and channel models | 146 |
| | 7.2 Capacity results | 150 |
| | 7.3 Proof techniques | 163 |
| | References | 172 |
| | Part III Unique channels | 175 |
| 8 | Modeling and characterization of ultraviolet scattering communication channels | 177 |
| | Haipeng Ding, Brian M. Sadler, Gang Chen, and Zhengyuan Xu | |
| | 8.1 Introduction | 177 |
| | 8.2 Single scattering models | 181 |

| | | |
|-----------|--|------------|
| 8.3 | Multiple scattering models | 183 |
| 8.4 | NLOS UV channel measurement systems | 189 |
| 8.5 | Numerical and experimental results | 192 |
| 8.6 | Summary | 198 |
| | References | 199 |
| 9 | Free-space optical communications underwater | 201 |
| | Brandon Cochenour and Linda Mullen | |
| 9.1 | Introduction: towards a link equation | 201 |
| 9.2 | Introduction to ocean optics | 202 |
| 9.3 | Channel characterization: theory | 213 |
| 9.4 | Experimental research in wireless optical communications underwater | 218 |
| 9.5 | System design for uFSO links | 228 |
| 9.6 | Summary | 236 |
| | References | 237 |
| 10 | The optical wireless channel | 240 |
| | Roger Green and Mark Leeson | |
| 10.1 | Introduction | 240 |
| 10.2 | System configurations | 241 |
| 10.3 | Optical sources | 242 |
| 10.4 | Optical detectors | 244 |
| 10.5 | Optical filters | 245 |
| 10.6 | Nature of the optical wireless channel | 247 |
| 10.7 | Interference sources | 248 |
| 10.8 | Impact of interference on BER | 251 |
| 10.9 | Channel impulse response | 253 |
| 10.10 | Hardware aspects of the receiver-amplifier in the indoor channel environment | 255 |
| 10.11 | Modulation schemes for optical wireless | 263 |
| 10.12 | Optics for optical wireless | 267 |
| 10.13 | Concluding remarks | 268 |
| | References | 269 |
| 11 | Hybrid RF/FSO communications | 273 |
| | Nick Letzepis and Albert Guillén i Fàbregas | |
| 11.1 | Introduction | 273 |
| 11.2 | Channel model | 275 |
| 11.3 | Information-theoretic preliminaries | 281 |
| 11.4 | Uniform power allocation | 287 |
| 11.5 | Power allocation | 292 |
| 11.6 | Conclusions and summary | 295 |

| | | |
|----------------|---|------------|
| | Appendix A Kullback–Leibler divergence between Poisson and Gaussian distributions | 297 |
| | Appendix B Derivative of the mutual information for discrete-input Poisson channels | 297 |
| | Acknowledgments | 299 |
| | References | 299 |
| Part IV | Applications | 303 |
| 12 | Quantum key distribution | 305 |
| | Rupert Ursin, Nathan Langford and Andreas Poppe | |
| | 12.1 Motivation | 305 |
| | 12.2 Security considerations of QKD | 306 |
| | 12.3 QKD protocols | 308 |
| | 12.4 Technical implementation of a free-space setup | 312 |
| | 12.5 QKD networks | 319 |
| | References | 326 |
| 13 | Optical modulating retro-reflectors | 328 |
| | William Rabinovich | |
| | 13.1 Introduction | 328 |
| | 13.2 Modulating retro-reflector link budgets | 330 |
| | 13.3 The optical retro-reflector | 332 |
| | 13.4 The optical modulator | 334 |
| | 13.5 Modulating retro-reflector applications and field demonstrations | 341 |
| | 13.6 Conclusion | 347 |
| | References | 347 |
| 14 | Visible-light communications | 351 |
| | Kang Tae-Gyu | |
| | 14.1 VLC principle | 351 |
| | 14.2 VLC standards | 354 |
| | 14.3 VLC research and development | 359 |
| | 14.4 VLC applications | 361 |
| | 14.5 Future work | 367 |
| | References | 367 |
| 15 | Optical wireless in sensor networks | 369 |
| | Dominic C. O'Brien and Sashigaran Sivathanan | |
| | 15.1 Introduction | 369 |
| | 15.2 Free-space optical (FSO) sensor network | 371 |

| | | |
|------|---|-----|
| 15.3 | Radio frequency/Free-space optical (RF/FSO) sensor network system | 378 |
| 15.4 | Conclusions | 383 |
| 15.5 | Acknowledgments | 384 |
| | References | 384 |
| | <i>Index</i> | 388 |

Contributors

Shlomi Arnon

Ben Gurion University of the Negev, Israel

John R. Barry

Georgia Institute of Technology

Gang Chen

University of California

Brandon Cochenour

Naval Air Systems Command (NAVAIR), USA

Haipeng Ding

University of California

Ivan Djordjevic

University of Arizona

Babak M. Ghaffari

Sharif University of Technology, Iran

Roger Green

University of Warwick

Steve Hranilovic

McMaster University, Canada

Albert G. i Fàbregas

University of Cambridge

Mohsen Kavehrad

Pennsylvania State University

George K. Karagiannidis

Aristotle University of Thessaloniki, Greece

Nathan Langford

University of Oxford

Amos Lapidath

ETH Zurich

Nick Letzepis

Defence Science and Technology Organisation, Australia

Mark Leeson

University of Warwick

Mehdi D. Matinfar

Sharif University of Technology, Iran

Stefan M. Moser

National Chiao Tung University, Taiwan

Linda Mullen

Naval Air Systems Command (NAVAIR), USA

Dominic O'Brien

University of Oxford

Andreas Poppe

AIT Austrian Institute of Technology GmbH

William Rabinovich

US Naval Research Laboratory

Brian M. Sadler

Army Research Laboratory, USA

Jawad A. Salehi

Sharif University of Technology, Iran

Robert Schober

University of British Columbia, Canada

Sashigaran Sivathasan

Curtin University of Technology, Malaysia

Rupert Ursin

Institute for Quantum Optics and Quantum Information (IQOQI),
Austrian Academy of Sciences

Kang Tae-Gyu

Electronics and Telecommunications Research Institute (ETRI), South Korea

Murat Uysal

Özyeğin University, Turkey

Michèle Wigger

Télécom ParisTech, France

Zhengyuan Xu

University of California

Part I

Outlook

1 Introduction

Shlomi Arnon, John Barry, George Karagiannidis, Robert Schober,
and Murat Uysal

Optical wireless communication is an emerging and dynamic research and development area that has generated a vast number of interesting solutions to very complicated communication challenges. For example, high data rate, high capacity and minimum interference links for short-range communication for inter-building communication, computer-to-computer communication, or sensor networks. At the opposite extreme is a long-range link in the order of millions of kilometers in the new mission to Mars and other solar system planets. It is important to mention that optical wireless communication is one of the oldest methods that humanity has used for communication. In prehistoric times humans used fire and smoke to communicate; later in history, Roman optical heliographs and Sumerians signalling towers were the communication systems of these empires. An analogous technology was used by Napoleonic Signalling Towers and “recently” by the light photo-phone of Alexander Graham Bell back in the 1880s.

Obviously, the data rate, quality of service delivered, and transceiver technologies employed have improved greatly from those early optical wireless technologies. In its many applications, optical wireless communication links have already succeeded in becoming part of our everyday lives at our homes and offices. Optical wireless products are already well familiar, ranging from visible-light communication (VLC), TV remote control to IrDA ports that currently have a worldwide installed base of hundreds of million of units with tens of percent annual growth. Optical wireless is also widely available on personal computers, peripherals, embedded systems and devices of all types, terrestrial and in-building optical wireless LANs, network of sensors, and inter-satellite link applications.

The book includes three main parts: Part II Optical wireless communication theory, Part III Unique channels, and Part IV Applications.

Part II describes important issues in optical wireless theory starting with Chapter 2 about coding and modulation techniques for optical wireless channels by Ivan B. Djordjevic. The author explains that the communication over the FSO channel is achieved through the line-of-sight (LOS) between two distant transceivers. An optical wave propagating over the FSO channel experiences fluctuations in amplitude and phase due to atmospheric turbulence, which represents a fundamental problem present even under clear sky conditions. In this chapter, several coded modulation concepts

enabling communication over strong atmospheric turbulence channels are described: (i) coded-multiple-input multiple-output (MIMO), (ii) raptor coding, (iii) adaptive modulation and coding (AMC), (iv) multidimensional coded modulation, and (v) coded-orthogonal frequency division multiplexing (OFDM). Furthermore, the concept of heterogeneous optical networking is discussed. Chapter 3 titled “Wireless optical CDMA communication systems” by Jawad A. Salehi, Babak M. Ghaffari, and Mehdi D. Matinfar, describes a particular and advanced form of optical wireless communication systems, namely optical code-division multiple-access (OCDMA), in the context of wireless optical systems. As wireless optical communication systems gets more mature and become viable for multi-user communication systems, advanced multiple-access techniques become more important and attractive in such systems. Among all multiple-access techniques in optical domain, OCDMA is of utmost interest because of its flexibility, ease of implementation, no need for synchronization among many users and soft traffic handling capability. The deployment of OCDMA communication systems in both indoor and outdoor free-space optical links is also analyzed. Chapter 4 is “Pointing error statistics” by Shlomi Arnon. In this chapter the author presents a simple model that describes the effect of the statistic of the pointing error on the performance of communication systems. Chapter 5 is “Equalization and Markov chains in cloud channel” by M. Kavehrad. The focus of this chapter is on investigating the possibility of simplifying the task of calculating the performances in cloud channel by a direct extraction of state transition matrices associated with standard Markov modeling from the MCRT computer simulations programs. Chapter 6 by Steve Hranilovic considers multiple-input multiple-output (MIMO) techniques employing a number of optical sources as transmitters and a collection of photodiodes as receivers for indoor optical wireless (OW) channels. The OW MIMO systems discussed in this chapter differ fundamentally from those used in radio channels. In particular, the signalling constraints imposed by intensity-modulated/direct-detection (IM/DD) systems limit the direct application of theory from radio channels. Nonetheless, MIMO techniques can be applied to OW channels to yield improvements in reliability and to improve data rates. This chapter starts with a brief overview of the characteristics of indoor OW MIMO systems. Given that the application and available gains depend on channel architecture, the balance of the chapter considers the use of MIMO techniques in three main OW channel topologies: diffuse, spot-diffusing, and point-to-point. The last chapter in this part, Chapter 7 by Amos Lapidot, Stefan M. Moser, Michèle Wigger, describes the basics of channel capacity. In this chapter the authors focus on communication systems that employ pulse amplitude modulation (PAM), which in the case of optical communication is called pulse intensity modulation. In such systems the transmitter modulates the information bits onto continuous-time pulses of duration T , and the receiver preprocesses the incoming continuous-time signal by integrating it over nonoverlapping intervals of length T . Such continuous-time systems can be modeled as discrete-time channels where the (discrete) time k input and output correspond to the integrals of the continuous-time transmitted and received signals (i.e., optical intensities) from kT to $(k + 1)T$. Note that for such discrete-time systems, the achieved data rate is not measured in bits (or nats) per second, but in bits (or nats) per channel use. They discuss three different discrete-time, pulse

intensity modulated, optical channel models: the discrete-time Poisson channel, the free-space optical intensity channel, and the optical intensity channel with input-dependent Gaussian noise.

Part III describes unique channels that could be found in optical wireless applications. Chapter 8: “Modeling and characterization of wireless ultraviolet scattering communication channels” by Zhengyuan Xu, Brian M. Sadler, Gang Chen, and Haipeng Ding, describes the modeling and characterization issues that arise from short-range communications over non-line-of-sight (NLOS) ultraviolet (UV) atmospheric scattering channels. The chapter starts by presenting the unique channel properties and history of NLOS UV communications, and introducing outdoor NLOS UV scattering link geometries. Both single and multiple scattering effects are considered, including channel impulse response and link path loss. Analysis and Monte Carlo simulation are employed to investigate the UV channel properties. The authors also report on experimental outdoor channel measurements, and compare with theoretical predictions. Chapter 9 titled “Free space optical communications underwater” by Brandon Cochenour and Linda Mullen, serves as both an introduction to the field of light propagation underwater, as well as a survey of current literature pertaining to underwater free-space optics (uFSO) or underwater optical wireless communication. The authors begin with a simple examination of a link budget equation. Next, they present an introduction of ocean optics in order to gain an appreciation for the challenges involved with implementing free-space optical links underwater. They then discuss state-of-the-art theoretical and experimental methods for predicting beam propagation in seawater. Finally, they present some common uFSO link types, and discuss the system-level design issues associated with each. Chapter 10 by Roger Green and Mark Leeson deals with Indoor IR communication channel. Infrared (IR) indoor optical wireless (OW) potentially combines the high bandwidth availability of optical communications with the mobility found in radio frequency (RF) wireless communication systems. So although IR is currently overshadowed by a multitude of home and office RF wireless networking schemes, it has significant potential when bandwidth demand is high. Compared to an RF system, OW offers the advantageous opportunity for high-speed medium- to short-range communications operating within a virtually unlimited and unregulated bandwidth spectrum using lower-cost components. Thus, the first sections of the chapter provide a brief overview of the system configurations, sources, detectors and filters used for OW followed by consideration of bit error rate (BER) performance in typical indoor scenarios. The third chapter in this part, Chapter 11, describes the concept of hybrid RF/optical wireless systems channel. The authors of this chapter are Nick Letzepis and Albert G. i Fàbregas. The authors remind us that in free-space optical (FSO) communication an optical carrier is employed to convey information wirelessly. FSO systems have the potential to provide fiber-like data rates with the advantages of quick deployment times, high security, and no frequency regulations. Unfortunately such links are highly susceptible to atmospheric effects. Scintillation induced by atmospheric turbulence causes random fluctuations in the received irradiance of the optical laser beam. Numerous studies have shown that performance degradation caused by scintillation can be significantly reduced through the use of multiple-lasers and multiple-apertures, creating the well-known multiple-input

multiple-output (MIMO) channel. However, it is the large attenuating effects of cloud and fog that pose the most formidable challenge. Extreme low-visibility fog can cause signal attenuation on the order of hundreds of decibels per kilometre. One method to improve the reliability in these circumstances is to introduce a radio frequency (RF) link to create a hybrid FSO/RF communication system. When the FSO link is blocked by cloud or fog, the RF link maintains reliable communications, albeit at a reduced data rate. Typically a millimetre wavelength carrier is selected for the RF link to achieve data rates comparable to that of the FSO link. At these wavelengths, the RF link is also subject to atmospheric effects, including rain and scintillation, but less affected by fog. The two channels are therefore complementary: the FSO signal is severely attenuated by fog, whereas the RF signal is not; and the RF signal is severely attenuated by rain, whereas the FSO is not. Both, however, are affected by scintillation. They propose a channel model for hybrid FSO/RF communications based on the well-known parallel channel, that takes into account the differences in signaling rate, and the atmospheric fading effects present in both the FSO and RF links. These fading effects are slow compared to typical data rates and, as such, each channel is based on a block-fading channel mode.

Part IV covers applications based on optical wireless communication. It begins with Chapter 12 about quantum encryption by Rupert Ursin, Nathan Langford, and Andreas Poppe. In this chapter the authors explain that the ability to guarantee security and privacy in communication are critical factors in encouraging people to accept and trust new tools and methods for today's information-based society (e.g. eCommerce, eHealth) and for future services (e.g. eGovernment, eVoting). The trend towards faster electronics provides the ability to handle longer keys, thus providing better security, but also increases the possibility of breaking keys in state-of-the-art cryptosystems. Nevertheless modern quantum cryptography has created a new paradigm for cryptographic communication, which provides strong security and incontrovertible evidence of any attempted eavesdropping which is based on theoretically and experimentally proven laws of nature. This technique, called quantum key distribution (QKD), generates a symmetrical classical bit string using the correlations of measurements on quantum systems and has already developed into a mature technology providing products capable of everyday use. The main hurdle for quantum communication is that, with present fiber and detector technologies, terrestrial QKD links are limited to distances of just over 100 km, well within reach of how far someone could travel in a short time to simply deliver the information in person. In the future, however, it will be possible to extend the distances spanned by individual fiber-based QKD links by using repeater nodes. These individual QKD links could then be combined to create larger and more complex QKD networks which will allow many different combinations of users to be connected over the same infrastructure. The economic benefits of such an interlinking network approach to QKD will be most apparent in a typical metropolitan scenario, where many potential users are likely to be located in a relatively small area, each wanting to be able to communicate securely with many different partners. Chapter 13 covers modulating retro-reflectors by William Rabinovich. In this chapter the author make it clear that direct FSO links with active terminals on both ends have many good applications. There

are, however, other applications in which the two ends of the link have different payload and power capabilities. Some examples include: unattended sensors, small unmanned aerial vehicles (UAVs) and small, tele-operated robots. For these applications a modulating retro-reflector (MRR) may be an appropriate solution. The MRR imposes a modulation on the interrogating beam and passively retro-reflects it back to the interrogator. The passive retro-reflector will generally have a large field of view over which incident light will be reflected back to its source, thus eliminating, or greatly reducing, pointing requirements on this end of the link. Despite this, the retro-reflected beam divergence can be very small, preserving the desirable features of direct FSO such as security and non-interference. Chapter 14 by Kang Tae-Gyu describes the emerging technology of visible-light communication. Visible-light communications is the name given to a wireless communication system that conveys information by modulating light that is visible to the human eye. Communications may not be the primary purpose of the light; in many applications the light primarily serves as a source of illumination. Interest in VLC has grown rapidly with the growth of visible-light light emitting diodes (LEDs) for illumination. The motivation is clear: When a room is illuminated by LEDs, why not exploit it to provide communications as well as illumination? This sharing of resources can save electric power and raw materials.

Chapter 15 targets the area of sensor networks and is written by Dominic C. O'Brien and Sashigaran Sivathasan. In this chapter two architectures that use retro-reflectors are described, based on the use of retro-reflecting links for sensor network application. In the first optical links provide all the communications of the system, and in the second this is augmented by the use of short-range RF links. These are given as representative examples only, and are not meant to represent any "best" approach to using optical wireless in sensor networks. However, they do illustrate some of the challenges and potential advantages of using this technique.

The combination of the different chapters within the book provides a unique database and a wide base of knowledge. The aspiration is to serve as a textbook for a graduate-level course for students in electrical engineering, electro optics engineering, communication engineering, and physics. It is also intended to serve as a source for self-study and as a reference book for senior engineers involved in the design of wireless communication systems. The background required for this book includes good knowledge in the areas of generating and detection of optical signal, probability and stochastic process, and communication theory. Part of this information and additional reading could be found in the books: *Applied Aspects of Optical Communication and LIDAR* by N. Blaunstein, S. Arnon, N. Kopeika, A. Zilberman, and *Optical Communication* Second Edition, by R. Gagliardi and S. Karp; and in the optical wireless communication special issues in *OSA/JON* 2006 and *IEEE/JSAC* 2009.

Part II

Optical wireless communication theory

2 Coded modulation techniques for optical wireless channels

Ivan B. Djordjevic

The transport capabilities of optical communication systems have increased tremendously in the past two decades, primarily due to advances in optical devices and technologies, and have enabled the Internet as we know it today with all its impacts on the modern society. Future internet technologies should be able to support a wide range of services containing a large amount of multimedia over different network types at high transmission speeds. The future optical networks should allow the interoperability of radio frequency (RF), fiber-optic and free-space optical (FSO) technologies. However, the incompatibility of RF/microwave and fiber-optics technologies is an important limiting factor in efforts to further increase future transport capabilities of such hybrid networks. Because of its flexibility, FSO communication is a technology that can potentially solve the incompatibility problems between RF and optical technologies. Moreover, FSO technologies can address any type of connectivity needs in optical networks. To elaborate, in metropolitan area networks (MANs), FSO communications can be used to extend the existing MAN rings; in enterprise networks, FSO can be used to enable local area network (LAN)-to-LAN connectivity and intercampus connectivity; and, last but not the least, FSO is an excellent candidate for the last-mile connectivity. FSO links are considered as a viable solution for various applications listed above because of the following properties [1]–[15]: (i) the high-directivity of the optical beam provides high power efficiency and spatial isolation from other potential interferers, a property not inherent in RF/microwave communications, (ii) FSO transmission is unlicensed, (iii) the large fractional-bandwidth coupled with high optical gain using moderate powers permits very high data rate transmission, (iv) the state-of-the-art fiber-optics communications employ intensity modulation with direct detection (IM/DD), and the components for IM/DD are widely available, and (v) FSO links are relatively easy to install and easily accessible for repositioning when necessary.

The communication over the FSO channel is achieved through the line-of-sight (LOS) between two distant transceivers. An optical wave propagating over the FSO channel experiences fluctuations in amplitude and phase due to atmospheric turbulence, which represents a fundamental problem present even under clear sky conditions. In this chapter, we describe several coded modulation concepts enabling communication over strong atmospheric turbulence channels: (i) coded-multiple-input multiple-output

(MIMO), (ii) raptor coding, (iii) adaptive modulation and coding (AMC), (iv) multidimensional coded modulation, and (v) coded-orthogonal frequency division multiplexing (OFDM). We also discuss the concept of heterogeneous optical networking. This chapter is organized as follows. The FSO channel model is introduced in Section 2.1. The codes on graphs suitable for use in FSO communications are described in Section 2.2. The concept of coded-MIMO is introduced in Section 2.3. The raptor coding concept for temporally correlated FSO channels is described in Section 2.4. The AMC concept is introduced in Section 2.5. We consider both feed-back AMC and hybrid FSO-RF communication scenarios. The multidimensional coded modulation concept is introduced in Section 2.6. The concept of FSO-OFDM transmission is introduced in Section 2.7. The heterogeneous optical networking concept is introduced in Section 2.8. Finally, Section 2.9 summarizes this chapter.

2.1 Atmospheric turbulence channel modeling

A commonly used turbulence model assumes that the variations of the medium can be understood as individual cells of air or eddies of different diameters and refractive indices. In the context of geometrical optics, these eddies may be observed as lenses that randomly refract the optical wavefront, generating a distorted intensity profile at the receiver of a communication system. The intensity fluctuation is known as scintillation, and represents one of the most important factors that limit the performance of an atmospheric FSO communication link. The most widely accepted theory of turbulence is due to Kolmogorov [1]. This theory assumes that kinetic energy from large turbulent eddies, characterized by the parameter known as outer scale L_0 , is transferred without loss to the eddies of decreasing size down to sizes of a few millimeters characterized by the inner scale parameter l_0 . The inner scale represents the cell size at which energy is dissipated by viscosity. The refractive index varies randomly across the different turbulent eddies and causes phase and amplitude variations to the wavefront. Turbulence can also cause the random drifts of optical beams – a phenomenon usually referred to as wandering – and can induce beam focusing.

The outer scale is assumed to be infinite in this chapter. Understanding the turbulence effects under zero inner scale is important as it represents a physical bound for the optical atmospheric channel and as such it has been of interest to researchers [1]. To account for the strength of turbulence we use the unitless Rytov variance, given by [1]

$$\sigma_R^2 = 1.23 C_n^2 k^{7/6} L^{11/6}, \quad (2.1)$$

where $k = 2\pi/\lambda$ is the wave number, λ is the wavelength, L is the propagation distance, and C_n^2 denotes the refractive index structure parameter, which is constant for horizontal paths. Weak fluctuations are associated with $\sigma_R^2 < 1$, strong fluctuations with $\sigma_R^2 > 1$, and the saturation regime is defined by $\sigma_R^2 \rightarrow \infty$ [1].

To characterize the FSO channel from a communication theory point of view, it is useful to give a statistical representation of scintillation. The reliability of a communication link can be determined if we use a good probabilistic model for the turbulence. Several

probability density functions (PDFs) have been proposed for the intensity variations at the receiver of an optical link [6]–[11]. Al-Habash *et al.* [12] proposed a statistical model that factorizes the irradiance as the product of two independent random processes each with a Gamma probability density function (PDF). The PDF of the intensity fluctuation is therefore

$$f(I) = \frac{2(\alpha\beta)^{(\alpha+\beta)/2}}{\Gamma(\alpha)\Gamma(\beta)} I^{(\alpha+\beta)/2-1} K_{\alpha-\beta} \left(2\sqrt{\alpha\beta I} \right), \quad I > 0, \quad (2.2)$$

where I is the signal intensity, α and β are parameters of the PDF, $\Gamma(\cdot)$ is the Gamma function, and $K_{\alpha-\beta}(\cdot)$ is the modified Bessel function of the second kind of order $\alpha - \beta$.

The parameters α and β of the PDF that predicts the scintillation experienced by plane waves in the case of $l_0 = 0$, are given by the expressions [4],[5]

$$\alpha = \left(\exp \left[\frac{0.49\sigma_R^2}{(1 + 1.11\sigma_R^{12/5})^{7/6}} \right] - 1 \right)^{-1} \quad \text{and} \\ \beta = \left(\exp \left[\frac{0.51\sigma_R^2}{(1 + 0.69\sigma_R^{12/5})^{5/6}} \right] - 1 \right)^{-1}, \quad (2.3)$$

where σ_R^2 is the Rytov variance as given in Eq. (2.1). This is a very interesting expression, because the PDF of the intensity fluctuations at the receiver can be predicted from the physical turbulence conditions. The predicted distribution matches very well the distributions obtained from numerical propagation simulations and experiments [1],[6].

2.2 Codes on graphs

The codes on graphs of interest in optical communications include turbo codes, turbo-product codes, and LDPC codes. Turbo codes [25],[26] can be considered as the generalization of the concatenated codes, where during iterative decoding the decoders exchange the soft messages for a certain number of times. Turbo codes can approach channel capacity closely in the region of interest for wireless communications. However, they exhibit strong error floors in the region of interest for optical communications; therefore, alternative iterative soft decoding approaches are to be sought. As recently shown in [16]–[24], turbo-product codes and LDPC codes can provide excellent coding gains and, when properly designed, do not exhibit an error floor in the region of interest for optical communications.

A turbo-product code (TPC) is an $(n_1 n_2, k_1 k_2, d_1 d_2)$ code in which codewords form an $n_1 \times n_2$ array such that each row is a codeword from an (n_1, k_1, d_1) code C_1 , and each column is a codeword from an (n_2, k_2, d_2) code C_2 . With n_i , k_i and d_i ($i = 1, 2$) we denoted the codeword length, dimension, and minimum distance, respectively, of the i th component code. The soft bit reliabilities are iterated between decoders for C_1 and C_2 . In optical communications, TPCs based on BCH component codes have been intensively studied, e.g. [25],[26].

If the parity-check matrix has a low density of 1's and the number of 1's per row and per column are both constant, the code is said to be a *regular LDPC* code. To facilitate the implementation at high speed, we prefer the use of regular rather than irregular LDPC codes. The graphical representation of LDPC codes, known as bipartite (Tanner) graph representation, is helpful for the efficient description of LDPC decoding algorithms. A *bipartite (Tanner) graph* is a graph whose nodes may be separated into two classes (*variable* and *check* nodes), and where *undirected edges* may only connect two nodes not residing in the same class. The Tanner graph of a code is drawn according to the following rule: check (function) node c is connected to variable (bit) node v whenever element h_{cv} in a parity-check matrix \mathbf{H} is a 1. In an $m \times n$ parity-check matrix, there are $m = n - k$ check nodes and n variable nodes.

Example 2.1: As an illustrative example, consider the \mathbf{H} -matrix of the following code

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}.$$

For any valid codeword $x = [x_0 x_1 \dots x_{n-1}]$, the checks used to decode the codeword are written as,

- Equation (c_0): $x_0 + x_2 + x_4 = 0 \pmod{2}$
- Equation (c_1): $x_0 + x_3 + x_5 = 0 \pmod{2}$
- Equation (c_2): $x_1 + x_2 + x_5 = 0 \pmod{2}$
- Equation (c_3): $x_1 + x_3 + x_4 = 0 \pmod{2}$

The bipartite graph (Tanner graph) representation of this code is given in Figure 2.1(a).

The circles represent the bit (variable) nodes while squares represent the check (function) nodes. For example, the variable nodes x_0 , x_2 , and x_4 are involved in Eq. (c_0), and therefore connected to check node c_0 . A closed path in a bipartite graph comprising l edges that closes back on itself is called a *cycle* of length l . The shortest cycle in the bipartite graph is called the *girth*. The girth influences the minimum distance of LDPC codes, correlates the extrinsic log-likelihood ratios (LLRs), and therefore affects the decoding performance. The use of large girth LDPC codes is preferable because the large girth increases the minimum distance and de-correlates the extrinsic information in the decoding process. To improve the iterative decoding performance, we have to avoid cycles of length 4, and preferably 6 as well. To check for the existence of short cycles, one has to search over the \mathbf{H} -matrix for the patterns shown in Figure 2.1(b,c).

2.2.1 Quasi-cyclic (QC) binary LDPC codes

In this section, we describe a method for designing large-girth QC LDPC codes. Based on Tanner's bound for the minimum distance of an LDPC code [27]

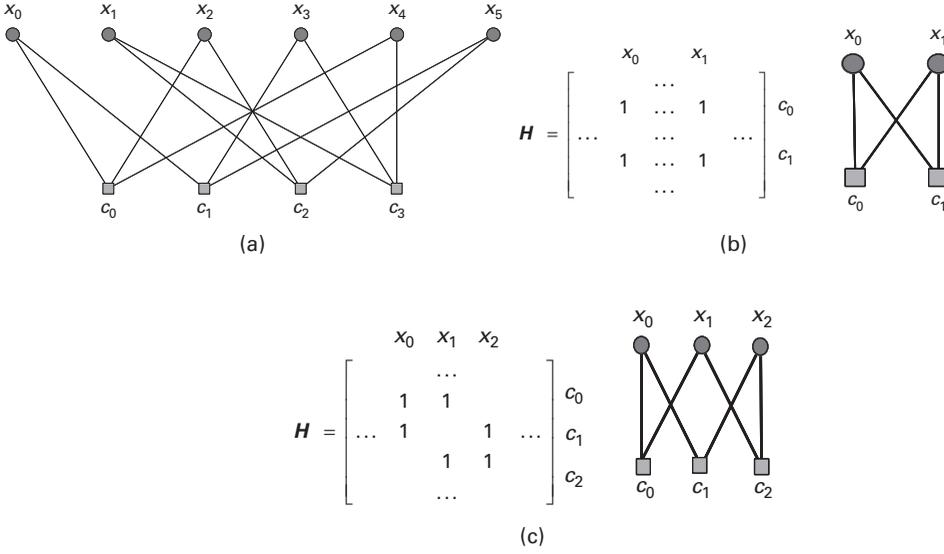


Figure 2.1 (a) Bipartite graph of (6, 2) code described by \mathbf{H} matrix above. Cycles in a Tanner graph: (b) cycle of length 4, and (c) cycle of length 6.

$$d \geq \begin{cases} 1 + \frac{w_c}{w_c - 2} ((w_c - 1)^{\lfloor (g-2)/4 \rfloor} - 1), & g/2 = 2m + 1 \\ 1 + \frac{w_c}{w_c - 2} ((w_c - 1)^{\lfloor (g-2)/4 \rfloor} - 1) + (w_c - 1)^{\lfloor (g-2)/4 \rfloor}, & g/2 = 2m \end{cases} \quad (2.4)$$

where g and w_c denote the girth of the code graph and the column weight, respectively, and where d stands for the minimum distance of the code; it follows that a large girth leads to an exponential increase in the minimum distance, provided that the column weight is at least 3. ($\lfloor \cdot \rfloor$ denotes the largest integer less than or equal to the enclosed quantity.) For example, the minimum distance of girth-10 codes with column weight $r = 3$ is at least 10. The parity-check matrix of regular QC LDPC codes [17],[28] can be represented by

$$\mathbf{H} = \begin{bmatrix} I & I & I & \dots & I \\ I & P^{S[1]} & P^{S[2]} & \dots & P^{S[c-1]} \\ I & P^{2S[1]} & P^{2S[2]} & \dots & P^{2S[c-1]} \\ \dots & \dots & \dots & \dots & \dots \\ I & P^{(r-1)S[1]} & P^{(r-1)S[2]} & \dots & P^{(r-1)S[c-1]} \end{bmatrix}, \quad (2.5)$$

where I is $B \times B$ (B is a prime number) identity matrix, P is $B \times B$ permutation matrix given by $P = (p_{ij})_{B \times B}$, $p_{i,i+1} = p_{B,1} = 1$ (zero otherwise), and where r and c represent the number of block-rows and block-columns in (2.5), respectively. The set of integers S are to be carefully chosen from the set $\{0, 1, \dots, B-1\}$ so that cycles of short length, in the corresponding Tanner (bipartite) graph representation of (2.5), are

avoided. According to Theorem 2.1 in [28], we have to avoid the cycles of length $2k$ ($k = 3$ or 4) defined by the following equation

$$S[i_1]j_1 + S[i_2]j_2 + \cdots S[i_k]j_k = S[i_1]j_2 + S[i_2]j_3 + \cdots S[i_k]j_1 \pmod{B}, \quad (2.6)$$

where the closed path is defined by $(i_1, j_1), (i_1, j_2), (i_2, j_2), (i_2, j_3), \dots, (i_k, j_k), (i_k, j_1)$ with the pair of indices denoting row-column indices of permutation-blocks in (2.5) such that $l_m \neq l_{m+1}, l_k \neq l_1$ ($m = 1, 2, \dots, k; l \in \{i, j\}$). Therefore, we have to identify the sequence of integers $S[i] \in \{0, 1, \dots, B-1\}$ ($i = 0, 1, \dots, r-1; r < B$) not satisfying Equation (2.6), which can be done either by computer search or in a combinatorial fashion. For example, to design the QC LDPC codes in [23], we introduced the concept of the cyclic-invariant difference set (CIDS). The CIDS-based codes come naturally as girth-6 codes, and to increase the girth we had to selectively remove certain elements from a CIDS. The design of LDPC codes of rate above 0.8, column weight 3 and girth $g \geq 10$ using the CIDS approach is very challenging and is still an open problem. Instead, in our recent paper [29] (see also [17]), we solved this problem by developing an efficient computer search algorithm. We add an integer at a time from the set $\{0, 1, \dots, B-1\}$ (not used before) to the initial set S and check if the Equation (2.6) is satisfied. If Equation (2.6) is satisfied, we remove that integer from the set S and continue our search with another integer from set $\{0, 1, \dots, B-1\}$ until we exploit all the elements from $\{0, 1, \dots, B-1\}$. The code rate of these QC codes, R , is lower-bounded by

$$R \geq \frac{|S|B - rB}{|S|B} = 1 - r/|S|, \quad (2.7)$$

and the codeword length is $|S|B$, where $|S|$ denotes the cardinality of set S . For a given code rate R_0 , the number of elements from S to be used is $\lfloor r/(1 - R_0) \rfloor$. With this algorithm, LDPC codes of arbitrary code rate can be designed.

Example 2.2: By setting $B = 2311$, the set of integers to be used in (2.5) is obtained as $S = \{1, 2, 7, 14, 30, 51, 78, 104, 129, 212, 223, 318, 427, 600, 808\}$. The corresponding LDPC code has rate $R_0 = 1 - 3/15 = 0.8$, column weight 3, girth-10 and length $|S|B = 15 \cdot 2311 = 34665$. In the example above, the initial set of integers was $S = \{1, 2, 7\}$, and the set of rows to be used in (2.5) is $\{1, 3, 6\}$. The use of a different initial set will result in a different set from that obtained above.

Example 2.3: By setting $B = 269$, the set S is obtained as $S = \{0, 2, 3, 5, 9, 11, 12, 14, 27, 29, 30, 32, 36, 38, 39, 41, 81, 83, 84, 86, 90, 92, 93, 95, 108, 110, 111, 113, 117, 119, 120, 122\}$. If 30 integers are used, the corresponding LDPC code has rate $R_0 = 1 - 3/30 = 0.9$, column weight 3, girth-8 and length $30 \cdot 269 = 8070$.

2.2.2 Decoding of LDPC codes

In this section, we describe the min-sum-with-correction-term algorithm [30] (see also [17]). It is a simplified version of the original algorithm proposed by Gallager [31]. Gallager proposed a near optimal iterative decoding algorithm for LDPC codes that computes the distributions of the variables in order to calculate the *a posteriori probability* (APP) of a bit v_i of a codeword $v = [v_0 v_1 \dots v_{n-1}]$ to be equal to 1 given a received vector $y = [y_0 y_1 \dots y_{n-1}]$. This iterative decoding scheme involves passing the extrinsic information back and forth among the c -nodes and the v -nodes over the edges to update the distribution estimation. Each iteration in this scheme is composed of two half-iterations. In Figure 2.2, we illustrate both the first and the second halves of an iteration of the algorithm. As an example, in Figure 2.2(a), we show the message sent from v -node v_i to c -node c_j . The v_i -node collects the information from the channel (y_i sample), in addition to extrinsic information from other c -nodes connected to the v_i -node, processes them and sends the extrinsic information (not already available information) to c_j . This extrinsic information contains the information about the probability $\Pr(c_j = b|y_i)$, where $b \in \{0, 1\}$. This is performed in all c -nodes connected to the v_i -node. On the other hand, Figure 2.2(b) shows the extrinsic information sent from c -node c_j to the v -node v_j , which contains the information about $\Pr(c_j$ equation is satisfied $|y)$. This is done repeatedly for all the c -nodes connected to the v_i -node. After this intuitive description, we describe the min-sum-with-correction-term algorithm in more detail because of its simplicity and suitability for high-speed implementation. Generally, we can either compute the APP $\Pr(v_i|y)$ or the APP ratio $l(v_i) = \Pr(v_i = 0|y)/\Pr(v_i = 1|y)$, which is also referred to as the likelihood ratio. In the log-domain version of the sum-product algorithm, we replace these likelihood ratios with log-likelihood ratios (LLRs) due to the fact that the probability domain includes many multiplications which leads to numerical instabilities, whereas the computation using LLRs involves addition only. Moreover, the log-domain representation is more suitable for finite precision representation. Thus, we compute the LLRs by $L(v_i) = \log[\Pr(v_i = 0|y)/\Pr(v_i = 1|y)]$. For the final decision, if $L(v_i) > 0$, we decide in favor of 0 and if $L(v_i) < 0$, we decide in favor of 1. To further explain the algorithm, we introduce the following notations due to MacKay [32]:

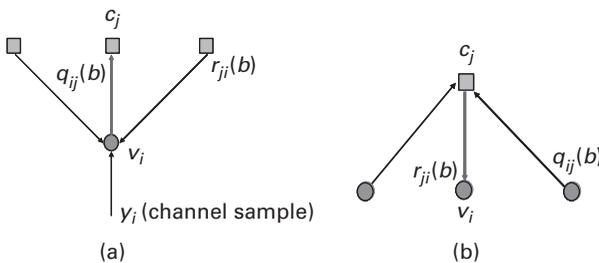


Figure 2.2 Illustration of the half-iterations of the sum-product algorithm: (a) first half-iteration: extrinsic information sent from v -nodes to c -nodes, and (b) second half-iteration: extrinsic information sent from c -nodes to v -nodes.

$$\begin{aligned}
V_j &= \{v\text{-nodes connected to } c\text{-node } c_j\} \\
V_j \setminus i &= \{v\text{-nodes connected to } c\text{-node } c_j\} \setminus \{v\text{-node } v_i\} \\
C_i &= \{c\text{-nodes connected to } v\text{-node } v_i\} \\
C_i \setminus j &= \{c\text{-nodes connected to } v\text{-node } v_i\} \setminus \{c\text{-node } c_j\} \\
M_v(\sim i) &= \{\text{messages from all } v\text{-nodes except node } v_i\} \\
M_c(\sim j) &= \{\text{messages from all } c\text{-nodes except node } c_j\} \\
P_i &= \Pr(v_i = 1 | y_i) \\
S_i &= \text{event that the parity-check equations involving } c_i \text{ are satisfied} \\
q_{ij}(b) &= \Pr(v_i = b | S_i, y_i, M_c(\sim j)) \\
r_{ji}(b) &= \Pr(\text{check equation } c_j \text{ is satisfied} | v_i = b, M_v(\sim i))
\end{aligned}$$

In the log-domain version of the sum-product algorithm, all the calculations are performed in the log-domain as follows:

$$L(v_i) = \log \left[\frac{\Pr(v_i = 0 | y_i)}{\Pr(v_i = 1 | y_i)} \right], L(r_{ji}) = \log \left[\frac{r_{ji}(0)}{r_{ji}(1)} \right], L(q_{ji}) = \log \left[\frac{q_{ji}(0)}{q_{ji}(1)} \right]. \quad (2.8)$$

The algorithm starts with the initialization step where we set $L(v_i)$ as follows:

$$\begin{aligned}
L(v_i) &= (-1)^{y_i} \log \left(\frac{1 - \varepsilon}{\varepsilon} \right), \quad \text{for BSC} \\
L(v_i) &= 2 \frac{y_i}{\sigma^2}, \quad \text{for binary input AWGN} \\
L(v_i) &= \log \left(\frac{\sigma_1}{\sigma_0} \right) - \frac{(y_i - \mu_0)^2}{2\sigma_0^2} + \frac{(y_i - \mu_1)^2}{2\sigma_1^2}, \quad \text{for BA-AWGN} \\
L(v_i) &= \log \left(\frac{\Pr(v_i = 0 | y_i)}{\Pr(v_i = 1 | y_i)} \right), \quad \text{for arbitrary channel}
\end{aligned} \quad (2.9)$$

where ε is the probability of error in the binary symmetric channel (BSC), σ^2 is the variance of the Gaussian distribution of the additive white Gaussian noise (AWGN), and μ_j and σ_j^2 ($j = 0, 1$) represent the mean and the variance of Gaussian process corresponding to bits $j = 0, 1$ of a binary asymmetric (BA)-AWGN channel. After initialization of $L(q_{ij})$, we calculate $L(r_{ji})$ as follows:

$$\begin{aligned}
L(r_{ji}) &= L \left(\sum_{i' \in V_j \setminus i} b_{j'} \right) = L(\dots \oplus b_k \oplus b_l \oplus b_m \oplus b_n \dots) \\
&= \dots L_k \boxed{+} L_l \boxed{+} L_m \boxed{+} L_n \boxed{+} \dots
\end{aligned} \quad (2.10)$$

where \oplus denotes the modulo-2 addition, and $\boxed{+}$ denotes a pairwise computation defined by

$$\begin{aligned}
L_a \boxed{+} L_b &= \text{sign}(L_a) \text{sign}(L_b) \min(|L_a|, |L_b|) + s(L_a, L_b) \\
s(L_a, L_b) &= \log \left(1 + e^{-|L_a + L_b|} \right) - \log \left(1 + e^{-|L_a - L_b|} \right).
\end{aligned} \quad (2.11)$$

The term $s(L_a, L_b)$ is the correction term and can be implemented as a lookup table (LUT). Upon calculation of $L(r_{ji})$, we update

$$L(q_{ij}) = L(v_i) + \sum_{j' \in C_i \setminus j} L(r_{j'i}), \quad L(Q_i) = L(v_i) + \sum_{j \in C_i} L(r_{ji}). \quad (2.12)$$

Finally, the decision step is as follows:

$$\hat{v}_i = \begin{cases} 1, & L(Q_i) < 0 \\ 0, & \text{otherwise.} \end{cases} \quad (2.13)$$

If the syndrome equation $\hat{v}H^T = \mathbf{0}$ is satisfied or the maximum number of iterations is reached, we stop, otherwise, we recalculate $L(r_{ji})$ and update $L(q_{ij})$ and $L(Q_i)$ and check again. It is important to set the number of iterations high enough to ensure that most of the codewords are decoded correctly and low enough not to affect the processing time.

2.3 Coded-MIMO free-space optical communication

The performance of FSO communication systems can be improved by using MIMO communication techniques [14],[15],[33]–[38]. In the case of FSO communications, the MIMO concept is realized by employing *multiple optical sources at the transmitter side and multiple detectors at the receiver side* (this scheme can also be called multiple lasers – multiple detectors [MLMD]). In this section we describe LDPC-coded repetition MIMO [38],[39] as an efficient way to deal with strong atmospheric turbulence. Although this concept is analogous to the wireless MIMO concept in RF case, the underlying physics is different, and new optimal and sub-optimal configurations for this channel are needed. Notice that space-time coding can also be used, but since we are concerned with direct detection, the signal constellation points cannot be negative or complex, which results in suboptimal performance. For more details on space-time coding concept for the FSO channel an interested reader is referred to [38].

2.3.1 LDPC-coded repetition MIMO concept

A block diagram of the LDPC-coded repetition MIMO scheme is shown in Figure 2.3. M optical sources are all pointed toward the distant array of N photodetectors using an expanding telescope. We assume that the beam spots on the receiver side are sufficiently wide to illuminate a whole photodetector array. This approach might help in simplifying the transmitter–receiver pointing problem. We further assume that the receiver's implementation is based on a p.i.n. photodetector in a Trans-impedance Amplifier (TA) configuration.

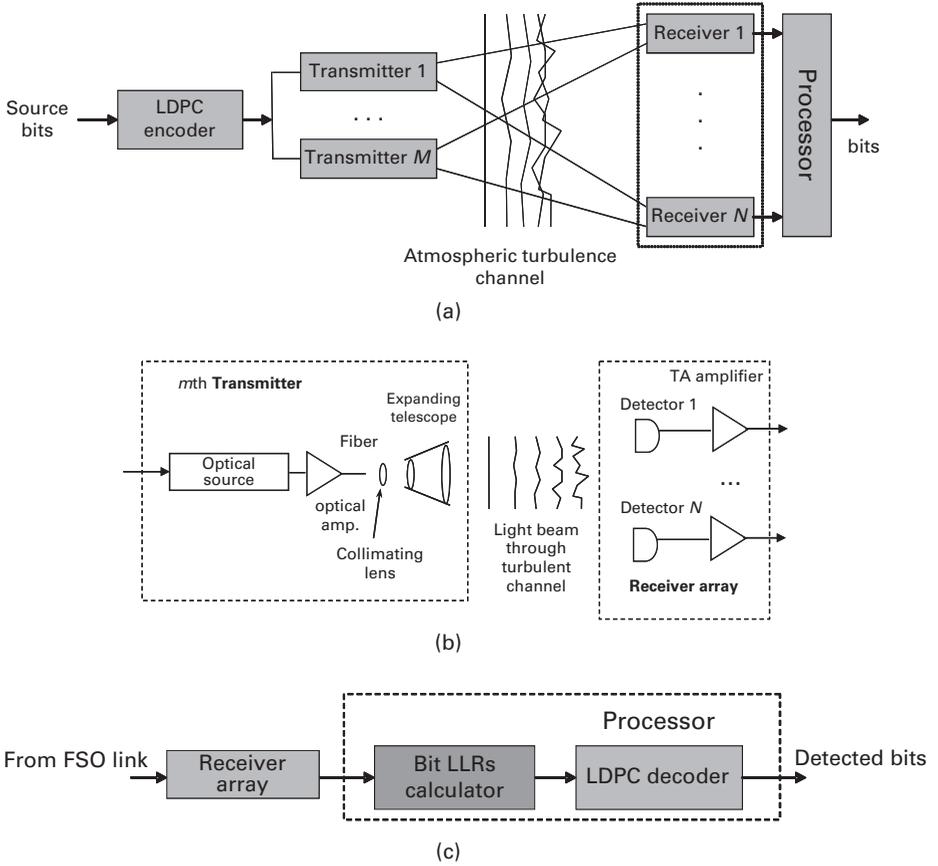


Figure 2.3 (a) Atmospheric optical LDPC-coded repetition MIMO system, (b) m th transmitter and receiver array configurations, and (c) processor configuration.

The n th photodetector output can be represented by

$$y_n(l) = x(l) \sum_{m=1}^M I'_{nm} + z_n(l); \quad n = 1, \dots, N; x(l) \in \{0, A\} \quad (2.14)$$

where A denotes the intensity of the pulse in the absence of scintillation, and $x(l)$ denotes data symbol at the l th time-slot; I'_{nm} represents the intensity channel coefficient between the n th photodetector ($n = 1, 2, \dots, N$) and the m th ($m = 1, 2, \dots, M$) optical source, which is described by the Gamma-Gamma PDF given above. The optical sources and photodetectors are positioned in such a way that different transmitted symbols experience independent atmospheric turbulence conditions; z_n denotes the n th receiver TA thermal noise that is modeled as a zero-mean Gaussian process with double-side power spectral density $N_0/2$. We assume nonnegative real signalling over the atmospheric turbulence channel with direct detection on a receiver side. Assuming that

the receiver TA thermal noise is white-Gaussian with a double-side-power spectral density $N_0/2$, the LLR of a symbol $x(l)$ (at the l th time-slot) for a binary repetition MIMO transmission is determined by

$$L(x(l)) = \log \left\{ \prod_{n=1}^N \frac{\frac{1}{\sqrt{2\pi\sqrt{N_0/2}}} \exp\left[-\frac{y_n(l)^2}{2N_0/2}\right]}{\frac{1}{\sqrt{2\pi\sqrt{N_0/2}}} \exp\left[-\frac{\left(y_n(l) - \sum_{m=1}^M I_{nm}\right)^2}{N_0}\right]} \right\} = \sum_{n=1}^N \left\{ -\frac{y_n(l)^2}{N_0} + \frac{\left(y_n(l) - \sum_{m=1}^M I_{nm}\right)^2}{N_0} \right\}. \quad (2.15)$$

2.3.2 Bit-interleaved LDPC-coded pulse amplitude modulation (PAM)

The block scheme of the bit-interleaved (BI) LDPC-coded PAM technique, for a repetition MIMO transmission, is shown in Figure 2.4. The source bit stream is encoded using an (n, k) LDPC code of the code rate $r = k/n$ (k the number of information bits, n the codeword length). The $l \times L$ block-interleaver (L is an integer multiple of the codeword length n), collects l code words written row-wise. The mapper accepts l bits at a time from the interleaver column-wise and determines the corresponding symbol for the Q -ary ($Q = 2^l$) PAM signalling using a *Gray mapping* rule. The number of columns in block-interleaver L is determined by a data rate, and temporal correlation of the channel. The basis function is given by

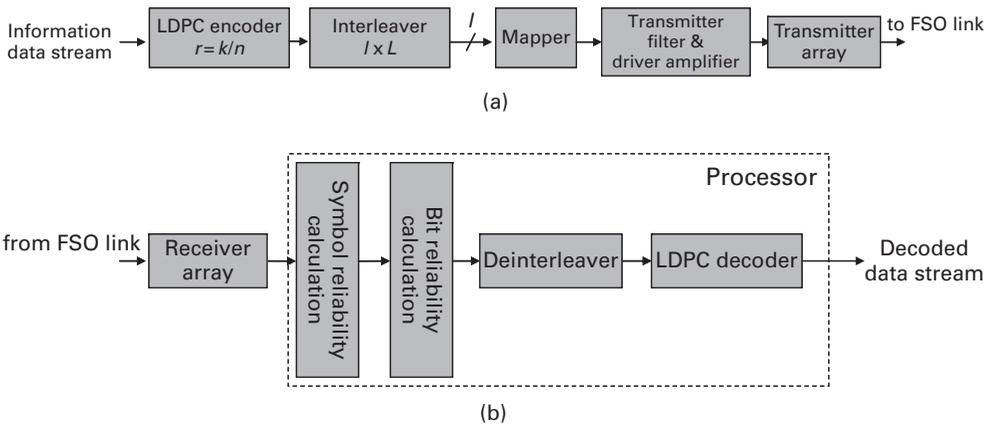


Figure 2.4 Repetition MIMO BI-LDPC-coded PAM configurations of: (a) transmitter and (b) receiver.

$$\phi_{\text{PAM}}(t) = \frac{1}{\sqrt{T}} \text{rect}(t/T), \quad \text{rect}(t) = \begin{cases} 1, & 0 \leq t < 1 \\ 0, & \text{otherwise} \end{cases} \quad (2.16)$$

while the signal constellation points by $A_q = qd$ ($q = 0, 1, \dots, Q-1$), where d is the separation between two neighboring points. The average symbol energy is given by

$$E_s = \frac{(Q-1)(2Q-1)}{6} d^2, \quad (2.17)$$

and it is related to the bit-energy E_b by $E_s = E_b \log_2 Q$. With this BI LDPC-coded modulation scheme, the neighboring information bits from the same source are allocated into different PAM symbols.

The outputs of the N receivers in the repetition MIMO, denoted as y_n ($n = 1, 2, \dots, N$), are processed to determine the symbol reliabilities (LLRs) by

$$\begin{aligned} \lambda(q) &= \log \left\{ \frac{1}{\sqrt{2\pi} \sqrt{N_0/2}} \exp \left[-\frac{\left(y_n - \text{map}(q) d \sum_{m=1}^M I_{nm} \right)^2}{N_0} \right] \right\} \\ &= -\sum_{n=1}^N \frac{\left(y_n - \text{map}(q) d \sum_{m=1}^M I_{nm} \right)^2}{N_0} + \log \left(\frac{1}{\sqrt{2\pi} \sqrt{N_0/2}} \right), \end{aligned} \quad (2.18)$$

where $\text{map}(q)$ denotes a corresponding mapping rule. The last term in Eq. (2.18) can be neglected in the practical implementation because it is constant for all symbols. Further, denote by c_j the j th bit in an observed symbol q 's binary representation $c = (c_1, c_2, \dots, c_l)$. The bit reliabilities $L(c_j)$ are determined from symbol reliabilities by

$$L(c_i) = \log \frac{\sum_{c:c_i=0} \exp[\lambda(q)] \exp\left(\sum_{c:c_j=0, j \neq i} L_a(c_j)\right)}{\sum_{c:c_i=1} \exp[\lambda(q)] \exp\left(\sum_{c:c_j=0, j \neq i} L_a(c_j)\right)}, \quad (2.19)$$

and forwarded to the LDPC decoder. Therefore, the i th bit reliability is calculated as the logarithm of the ratio of a probability that $c_i = 0$ and probability that $c_i = 1$. In the nominator, the summation is done over all symbols q having 0 at position i , while in the denominator it is done over all symbols q having 1 at the position i . With $L_a(c_j)$ we denoted a-priori information determined from the LDPC decoder extrinsic LLRs. The inner summation in (2.19) is done over all bits of symbol q , selected in the outer summation, for which $c_j = 0, j \neq i$. By iterating the extrinsic reliabilities between the APP demapper and the LDPC decoder, the overall bit error rate (BER) performance can be improved. The hard decisions from the LDPC decoder are delivered to the end-user.

Other multilevel schemes, such as those based on Quadrature Amplitude-Modulation (QAM) are also applicable. However, the use of an additional DC bias is required because negative signals cannot be transmitted over an IM/DD system, and the power efficiency of such schemes is low.

To improve the BER performance, we perform the iteration of the *extrinsic* information between the a-posteriori probability PAM demapper and LDPC decoder. For a convergence behavior analysis, we perform the EXIT chart analysis [40]. To determine the Mutual Information (MI) transfer characteristics of the demapper, we model a-priori input LLR, $L_{M,a}$, as a conditional Gaussian random variable. The MI between bit c in a codeword and the corresponding input LLR ($L_{M,a}$) is determined numerically. Similarly, the MI $I_{LM,e}$ between c and $L_{M,e}$ is calculated numerically, but with the PDF of c and $L_{M,e}$ determined from the histogram obtained by Monte Carlo simulation. By observing the $I_{LM,e}$ as a function of the MI of $I_{LM,a}$ and receiver signal-to-noise ratio, E/N_0 , in dB, the demapper EXIT characteristic (denoted as T_M) is given by

$$I_{LM,e} = T_M(I_{LM,a}, E/N_0). \quad (2.20)$$

The EXIT characteristic of the LDPC decoder (denoted by T_D) is defined in a similar fashion as

$$I_{LD,e} = T_D(I_{LD,a}). \quad (2.21)$$

The “turbo” demapping based receiver operates by passing extrinsic LLRs between the demapper and LDPC decoder. The iterative process starts with an initial demapping in which $L_{M,a}$ is set to zero yielding to $I_{LM,a} = 0$. The demapper output LLRs, described by

$$I_{LM,e} = I_{LD,a}$$

are fed to the LDPC decoder. The LDPC decoder output LLRs, described by

$$I_{LD,e} = I_{LM,a}$$

are fed to the APP demapper. The iterative procedure is repeated until convergence or the maximum number of iterations has been reached. This procedure is illustrated in Figure 2.5, where the APP demapper and LDPC decoder EXIT charts are shown together on the same graph; 4-PAM, 8-PAM, and 16-PAM are observed, as well as the natural and Gray mapping. The EXIT curves have different slopes for different mappings. The existence of a “tunnel” between corresponding demapping and decoder curves indicates that the iteration between demapper and decoder will be successful. The smallest signal-to-noise ratio at which the iterative scheme starts to converge is known as the threshold (pinch-off) limit [40]. The threshold limit in the case of 16-PAM (Figure 2.5(b)) is about 8 dB worse as compared to 4-PAM (Figure 2.5(a)). The BER vs. electrical SNR in the presence of scintillation (per photodetector), for a strong turbulence regime ($\sigma_R = 3.0$, $\alpha = 5.485$, $\beta = 1.1156$), are shown in Figure 2.6. The BER is shown for a different number of optical sources, and photodetectors, by employing an (6419,4794) irregular girth-6 LDPC code of a rate 0.747 designed using the concept of the pairwise-balanced-design (PBD) [42]. The LDPC-coded repetition MIMO with $M = 4$, $N = 1$ provides more than 14 dB improvement over LDPC-coded OOK with single optical source and single photodetector.

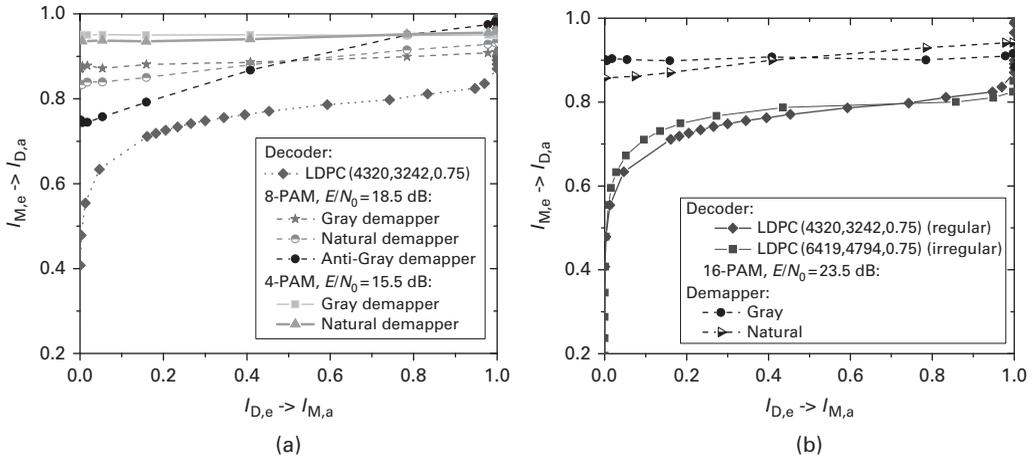


Figure 2.5 EXIT chart for different PAM constellations: (a) 4-PAM and 8-PAM, and (b) 16-PAM. (After ref. [38]; ©IEEE 2008; reprinted with permission.)

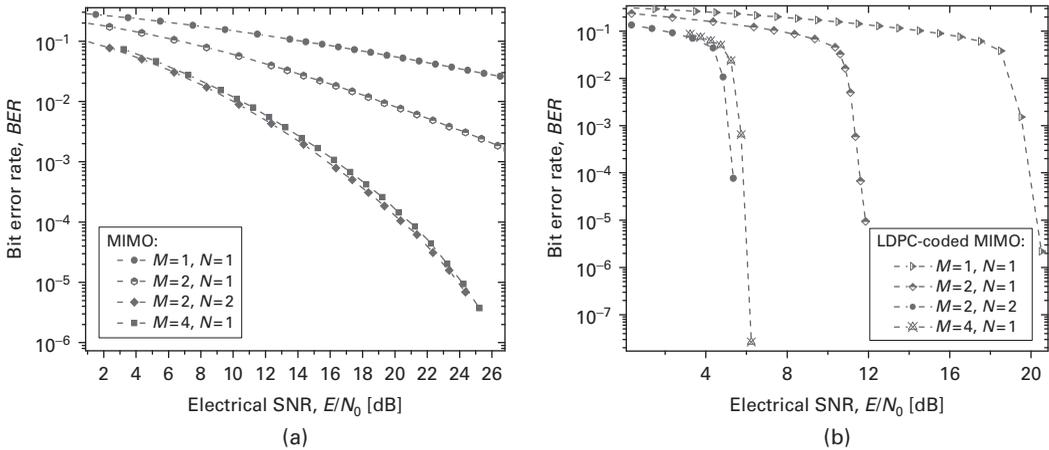


Figure 2.6 BERs of binary LDPC(6419,4794)-coded repetition MIMO: (a) uncoded case, and (b) coded case. (After ref. [38]; ©IEEE 2008; reprinted with permission.)

The results of simulations for bit-interleaved LDPC(6419,4794)-coded PAM are shown in Figure 2.7 for different MIMO configurations and different number of signal constellation points employing the Gray mapping rule. Once more, excellent BER performance improvement is obtained, about 23 dB for $M = N = 4$, $Q = 4$ over $M = N = 1$, $Q = 4$. The comparison for different component LDPC codes is given in Figure 2.7(a). The scheme employing a girth-6 (g -6) irregular PBD-based LDPC code of rate 0.75 performs comparable to a girth-8 regular QC-LDPC code of the same rate. The scheme based on a girth-8 regular balanced incomplete block design (BIBD) [42] based LDPC code of rate 0.81 performs worse than 0.75 codes. However, the SNR difference is becoming smaller as the constellation size grows.

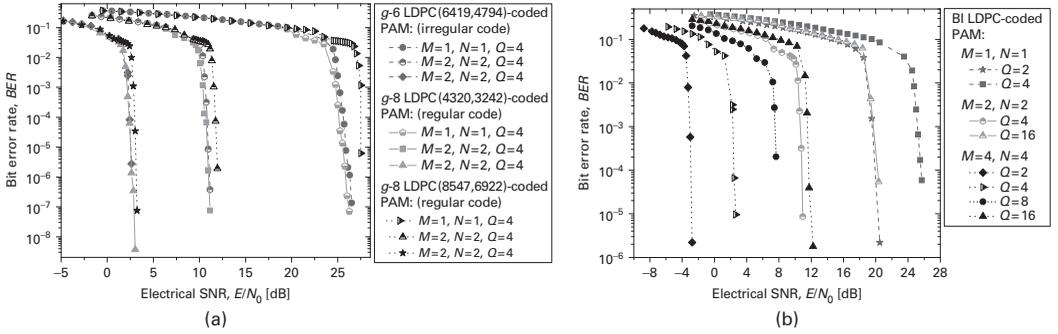


Figure 2.7 (a) BER performance of BI-LDPC-coded PAM with repetition MIMO for different LDPC component codes. (b) BER performance of BI-LDPC(6419,4794)-coded PAM with repetition MIMO. (After ref. [38]; ©IEEE 2008; reprinted with permission.)

2.3.3 Bit-interleaved LDPC-coded pulse position modulation (PPM)

The LDPC-coded PAM scheme described in the previous section is not power efficient. In this section we describe a power-efficient scheme based on Q -ary PPM. The source bit streams are encoded using an (n, k) LDPC code whose code rate is $r = k/n$, where k denotes the number of information bits, and n denotes the codeword length. The output of the LDPC encoder is forwarded to an $m \times n$ block-interleaver, which collects m code words in row-wise fashion. The mapper takes m bits at a time from the interleaver column-wise and determines the corresponding slot for Q -ary ($Q = 2^m$) PPM signalling based on a *Gray mapping* rule. Therefore, in this scheme, the neighboring information bits originating from the same source are allocated to different PPM symbols, which might improve the tolerance against atmospheric turbulence. In Q -ary PPM, at each signalling interval T_s a pulse of light of duration $T = T_s/Q$ is transmitted by a laser. Notice that the signalling interval T_s is subdivided into Q slots of duration T . The total transmitted power, denoted by P_{tot} , is the same regardless of the number of lasers, and the power per laser is P_{tot}/M . The i th ($i = 1, 2, \dots, M$) laser modulated beam is projected toward the j th ($j = 1, 2, \dots, N$) receiver using an expanding telescope.

The outputs of N receivers when the symbol q is transmitted, denoted as $Z_{n,q}$ ($n = 1, 2, \dots, N; q = 1, 2, \dots, Q$), are processed to determine the symbol reliabilities $\lambda(q)$ ($q = 1, 2, \dots, Q$) by

$$\lambda(q) = -\frac{\sum_{n=1}^N \left(Z_{n,q} - \frac{\sqrt{E_s}}{M} \sum_{l=1}^M I_{n,l} \right)^2}{\sigma^2} - \frac{\sum_{n=1}^N \sum_{l=1, l \neq q}^Q Z_{n,l}}{\sigma^2}. \quad (2.22)$$

In (2.22), E_s denotes the symbol energy of an uncoded symbol in electrical domain (in the absence of scintillation), which is related to the bit energy E_b by $E_s = E_b \log_2 Q$; σ^2 denotes the variance of TA thermal noise (modeled as AWGN), and it is related to the double-side power spectral density N_0 by $\sigma^2 = N_0/2$; I_{nl} denotes the intensity of the light incident to the n th photodetector ($n = 1, 2, \dots, N$), originating from the

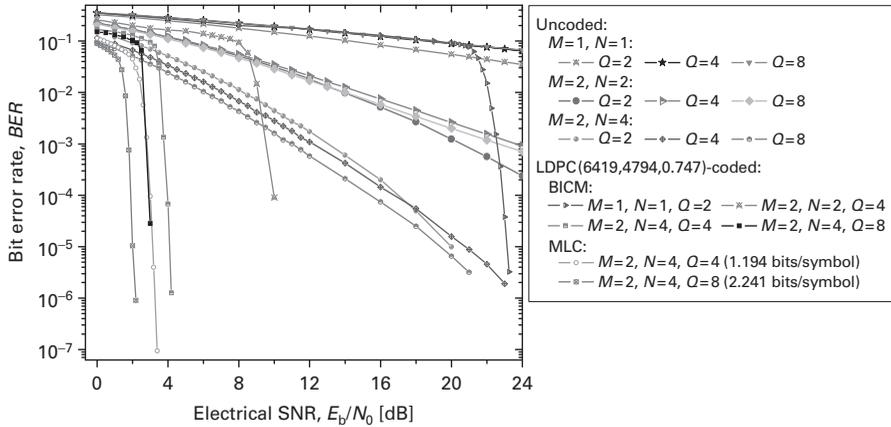


Figure 2.8 BER performance of bit-interleaved LDPC-coded modulation against MLC for different MIMO (MLMD) configurations.

l th ($l = 1, 2, \dots, M$) laser source. The results of simulations assuming the operation under the strong turbulence regime ($\sigma_R = 3.0$) are shown in Figure 2.8, for different number of lasers, photodetectors, and number of slots. The (6419,4794) irregular girth-6 LDPC code of rate 0.747 designed using the concept of PBD [42], is employed. The multilevel coding (MLC) scheme with spectral efficiency of 2.241 bits/symbol combined with MIMO scheme employing two lasers and four photodetectors provides 21dB improvement over LDPC-coded binary PPM employing one laser and one photodetector. The corresponding BICM scheme of higher spectral efficiency (3bits/symbol) performs about 1 dB worse. The number of inner iterations in the sum-product LDPC decoder is set to 25 in both schemes, while the number of outer iterations in the BICM scheme is set to 10. MLC employs parallel independent LDPC decoding. Therefore, the bit-interleaved LDPC-coded modulation scheme, although simpler to implement than the MLC scheme has a slightly worse BER performance.

2.4 Raptor codes for temporally correlated FSO channels

A *rate-less code* is an error-correcting code whose code-rate can be changed according to the time-varying channel conditions [64]–[67]. The well-known rate-less codes are: (i) *punctured codes*, in which the rate is varied by puncturing the parity bits so that the effective code rate is increased; and (ii) *fountain codes*, in particular *Raptor codes*, in which the rate is varied by changing the codeword length. A *Raptor code* is obtained by concatenating an inner error-correcting code (the pre-code), with an outer Luby-transform (LT) code [65]. An LT Code is a sparse random linear code, with a very simple decoding algorithm. The *LT encoding* can be described by the following algorithm [65],[67]. Each encoded symbol x_n is generated from the message symbols $s_1, \dots, s_{K,LT}$ as follows [67]: (i) randomly choose the degree d_n from a degree

distribution $\Omega(x)$, and (ii) choose uniformly at random d_n distinct input symbols, and set x_n equal to the bitwise sum, modulo 2, of those d_n symbols. *LT decoding* can be described by the following algorithm. The decoder's task is to recover s from $x = sG$, where G is the matrix associated with the graph (of the pseudorandom matrix) by using the sum-product algorithm [67]:

1. Find a check node x_n that is connected to only one source symbol s_k
 - (a) Set $s_k = x_n$.
 - (b) Add s_k to all checks $x_{n'}$ that are connected to s_k :
 $x_{n'} := x_{n'} + s_k$ for all n' such that $G_{nk} = 1$.
 - (c) Remove all the edges connected to the source symbol s_k .
2. Repeat (1) until all s_k are determined.

LT codes, although simple to encode and decode, have the following two drawbacks: d_n can take any integer value up to the size of the input word, leading to a decoding complexity $O(K_{LT} \log K_{LT})$, and an error floor is observed at high SNRs. To avoid these problems, Raptor codes are advocated in [64],[66]. A Raptor code is formed by concatenating an inner error correcting code (the pre-code) with an outer LT code, as illustrated below, in Figure 2.9(a). In this example, the inner code is a systematic code, so that we can collapse the Tanner graph as illustrated in Figure 2.9(b). The joint Tanner graph is obtained by joining the inner code parity-check nodes with LT check nodes. We can use this graph to perform joint iterative decoding.

The LLR to be passed from check-node c to variable-node v can be calculated as follows:

$$L(\mu_{c,v}^{(t)}) = \begin{cases} 2 \tanh^{-1} \left[\tanh \left(\frac{L(y_c)}{2} \right) \prod_{w \in V_c, w \neq v} \tanh \left(\frac{L(\eta_{w,c}^{(t-1)})}{2} \right) \right], & \forall c \in C_{LT} \\ 2 \tanh^{-1} \left[\prod_{w \in V_c, w \neq v} \tanh \left(\frac{L(\eta_{w,c}^{(t-1)})}{2} \right) \right], & \forall c \in C_{LDPC} \end{cases} \quad (2.23)$$

where V_c denotes the set of variable-nodes connected to check-node c , C_{LT} denotes the set of LT check-nodes, and C_{LDPC} denotes the set of inner code check-nodes (in the

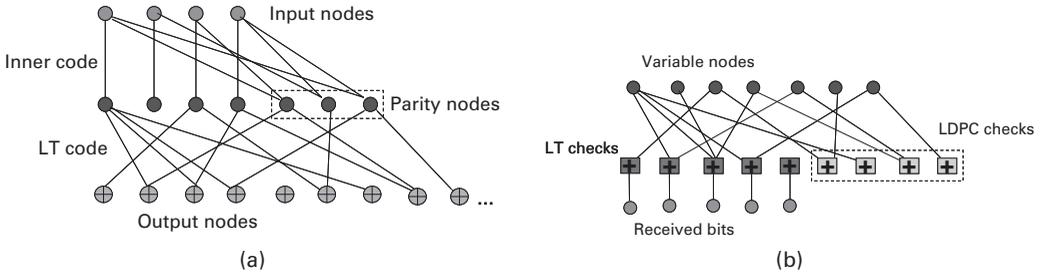


Figure 2.9 (a) Tanner graph representation of a Raptor code. (b) Joint Tanner graph representation of a raptor code.

example above the inner code is an LDPC code). In (2.23), we use the character μ to denote the passage of messages from c -nodes to v -nodes, and the character η to denote the passage of messages in the opposite direction. The superscript t is used to denote the current iteration, and $(t-1)$ denotes the previous one; y_c is the sample that corresponds to transmitted codeword bit x_v . The LLR to be passed from variable-node v to check-node c can be calculated as follows:

$$L(\eta_{v,c}^{(t)}) = \begin{cases} \sum_{d \in C_v, d \neq c} L(\mu_{d,v}^{(t-1)}), & \forall c \in C_{LT} \\ L(x_v^{(t)}) + \sum_{d \in C_v, d \neq c} L(\mu_{d,v}^{(t-1)}), & \forall c \in C_{LDPC} \end{cases}. \quad (2.24)$$

The variable-node v LLR-update rule is given by

$$L(x_v^{(t)}) = \sum_{c \in C_v} L(\mu_{c,v}^{(t)}). \quad (2.25)$$

Finally, the decision step is as follows:

$$\hat{x}_v = \begin{cases} 1, & L(x_v^{(t)}) < 0 \\ 0, & \text{otherwise} \end{cases}. \quad (2.26)$$

If the syndrome equation $\hat{x}H^T = \mathbf{0}$ (\mathbf{x} denotes the codeword and \mathbf{H} is the joint parity-check matrix) is satisfied or the maximum number of iterations is reached, we stop, otherwise, we recalculate (2.23)–(2.26) and check again.

The system model for study of the suitability of raptor codes for use in temporally correlated FSO channels is shown in Figure 2.10. The training sequence is used to determine the channel state information (CSI), which is feedback to the transmitter that can be code-rate adapted according to the FSO channel conditions.

For raptor codes performance evaluation, the experimental setup shown in Figure 2.11 is used. A CW laser was used to estimate the channel conditions, the variations due to scintillation are recorded and used to determine the histogram and autocovariance functions. The raptor code is created by concatenating a regular LDPC(495,433) (MacKay) code of rate 0.87 and an LT code. The user bit sequence is fed into a buffer that stores 40 words of $k_C = 433$ bits each, which are encoded using the LDPC code above. The LT number of information bits is therefore determined by $k_{LT} = 19800 (= 40 \times 495)$.

The results of Monte Carlo integration, used to determine the information rates as described in [55], are shown in Figure 2.12. The comparison is performed with respect

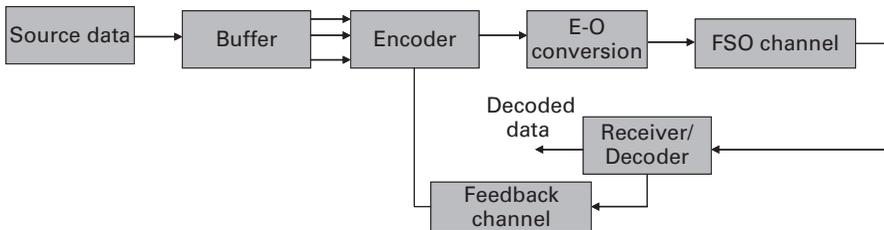


Figure 2.10 FSO system with feedback channel.

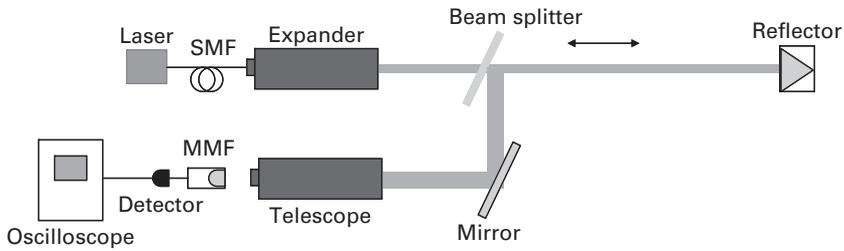


Figure 2.11 Experimental setup. Total transmission link is 600 m.

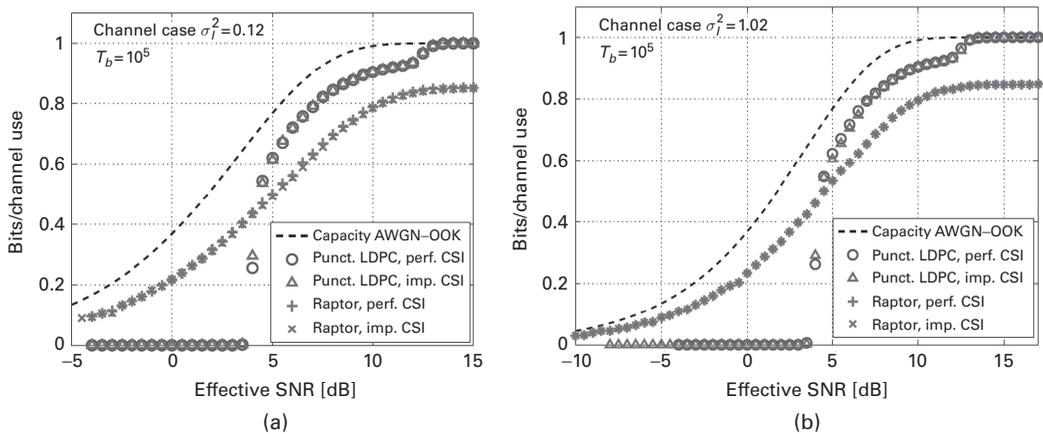


Figure 2.12 Information rates at 1 Gb/s: (a) in the weak turbulence regime, and (b) in the strong turbulence regime. (After ref. [64]; ©IEEE 2010; reprinted with permission.)

to the corresponding punctured LDPC code. We see that at high SNRs, the punctured LDPC code outperforms the raptor code described above. However, in the regime of low SNRs the punctured LDPC code cannot be used at all. Another interesting point to notice is that the raptor code faces insignificant information rate degradation when the CSI is imperfect. The degradation of information rates in the strong turbulence regime is small compared to the weak turbulence regime. Therefore, the raptor codes can be used to enable communication under the strong turbulence regime. Another interesting approach to enable communication under the strong turbulence regime is to use the adaptive modulation and coding [68]–[71], which is described in the next section.

2.5 Adaptive modulation and coding (AMC) for FSO communications

Adaptive modulation and coding, already in use in wireless channels, enables robust and spectrally efficient transmission over the time-varying channels [68]. The key idea behind AMC is to estimate the channel conditions at the receiver side and feed the channel estimates back to the transmitter using an RF feedback channel, so that the

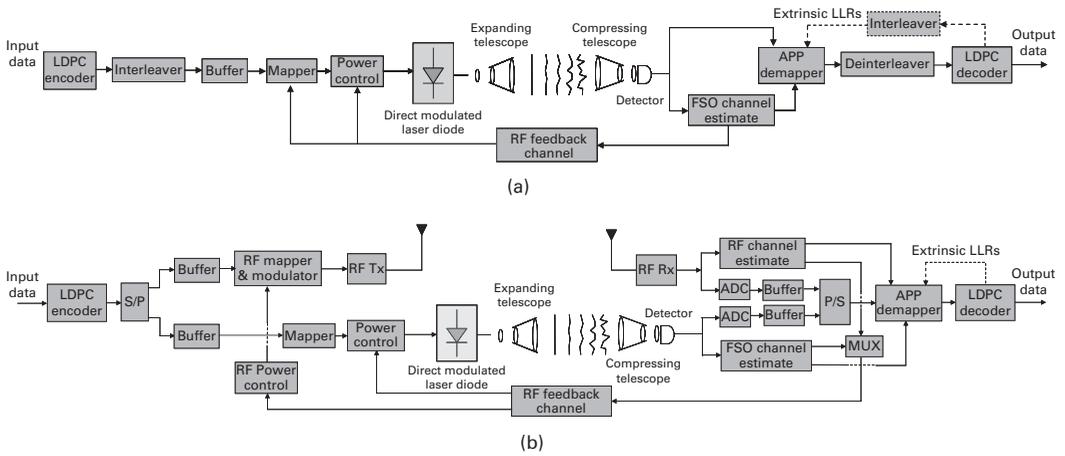


Figure 2.13 The AMC schemes for communication over FSO channel: (a) AMC FSO system with RF feedback, (b) hybrid FSO-RF communication system.

transmitter can be adapted relative to the channel conditions [69]–[71]. Because the FSO communication systems already use RF for the backup channel, this approach does not increase complexity that much. We describe several different AMC scenarios including: (i) channel inversion, (ii) truncated channel inversion, and (iii) variable-rate variable-power adaptation. We also study the improvements through *hybrid FSO-RF communication* in which RF and FSO subsystem cooperate to maximize the throughput. The adaptive FSO communication system with RF feedback, shown in Figure 2.13(a), consists of a transmitter, propagation path through the atmosphere, and a receiver. The optical transmitter includes a semiconductor laser of high launch power, an adaptive modulation and coding block, and a power control block. To reduce the system cost, the direct modulation of a laser diode is used. The modulated beam is projected toward the distant receiver by using an expanding telescope assembly. Along the propagation path through the atmosphere, the light beam experiences absorption, scattering, and atmospheric turbulence, which cause attenuation, random variations in amplitude and phase, and beam wandering. At the receiver side, an optical system collects the incoming light and focuses it onto a detector, which generates an electrical current proportional to the incoming power. The intensity channel estimate is transmitted back to the transmitter by using an RF feedback channel. Because the atmospheric turbulence changes slowly, with a correlation time ranging from $10\ \mu\text{s}$ to $10\ \text{ms}$, this is a plausible scenario for FSO channels with data rates in the order of Gb/s. Given the fact that systems integrating RF and FSO technologies have recently been reported [83] and data rates needed for RF feedback are low, this system is a promising candidate for future FSO systems.

There are many parameters that can be varied at the transmitter side relative to the FSO channel intensity gain; including data rate, power, coding rate, and combinations of different adaptation parameters. Transmitter power adaptation, similar to wireless

communications, can be used to compensate for SNR variation due to atmospheric turbulence, with the aim to maintain a desired BER. The power adaptation therefore “inverts” the FSO channel scintillation so that the FSO channel behaves similarly as an AWGN channel to the receiver. The power adaptation policy for *FSO channel inversion* is similar to that for wireless links [68], and is defined by

$$P(i_t) = C/i_t^2, \quad (2.27)$$

where $C = P/E[1/i_t^2]$ and i_t is the irradiance at time instance t ; $P(i_t)$ is the instantaneous power, and P is the average power. The FSO channel, upon channel inversion, appears to the transmitter as a standard AWGN channel with $SNR = \Gamma_0/E[1/i_t^2]$, where $\Gamma_0 = E_s/N_0$ is the signal-to-noise ratio in the absence of scintillation, with E_s being the symbol energy and $N_0/2$ being the double-side power spectral density of AWGN related to variance by $\sigma^2 = N_0/2$.

The channel inversion can be detrimental when $i_t \rightarrow 0$ because the factor $E[1/i_t^2]$ tends to infinity. To avoid this problem the *truncated channel inversion* policy from wireless communications [68] can be adopted, in which the channel inversion is performed only when the irradiance is above the certain threshold $i_{tsh} > 0$; that is

$$P(i_t) = \begin{cases} C/i_t^2, & i_t \geq i_{tsh} \\ 0, & i_t < i_{tsh} \end{cases}. \quad (2.28)$$

The threshold irradiance is to be chosen in such a way to maximize channel capacity. In variable data-rate adaptation, we change the signal constellation size for the fixed symbol rate depending on FSO channel conditions. When the FSO channel conditions are favorable we increase the constellation size, decrease it when channel conditions are not favorable, and do not transmit at all when the intensity channel coefficients are below the irradiance threshold. In the optimum power-variable data-rate adaptive scheme, we adapt the launch power and data rate so that the spectral efficiency is maximized. With this adaptation policy, more power and higher data rates are transmitted when FSO channel conditions are good, less power and lower data rates are transmitted when FSO channel is bad, and nothing is transmitted when the FSO irradiance falls below the threshold. For more details on optimum adaptation scheme an interested reader is referred to [71].

The second AMC scheme, the adaptive hybrid FSO-RF communication system, shown in Figure 2.13(b), consists of two parallel FSO and RF channels. The LDPC encoded data stream is partially transmitted over FSO and partially over RF channel. The operating symbol rate of the FSO channel is commonly many times higher than that of the RF channel. The FSO channel comprises an FSO transmitter, propagation path through the atmosphere, and an FSO receiver. The optical transmitter includes a semiconductor laser of high launch power, an adaptive mapper, and a power control block. To reduce the system cost, direct modulation of laser diode is used. The modulated beam is projected toward the distant receiver by using an expanding telescope assembly. Along the propagation path through the atmosphere, the light beam experiences absorption, scattering, and atmospheric turbulence, which cause attenuation, and random variations in amplitude and phase. The RF channel comprises adaptive RF mapper, RF power

control, RF transmitter (Tx), transmitting antenna, wireless propagation path, receiver antenna, and RF receiver (Rx). The RF channel estimates and FSO irradiance estimates are transmitted back to transmitters using the same RF feedback channel. The data rates and powers in both channels are varied in accordance with the channel conditions. The symbol rates on both channels are kept fixed while the signal constellation sizes are varied based on channel conditions. When the FSO (RF) channel condition is favorable larger constellation size is used, when the FSO (RF) channel condition is poor, smaller constellation sizes are used, and when the FSO (RF) channel SNR falls below threshold the signal is not transmitted at all. Both the subsystems (FSO and RF) are designed to achieve the same target bit error probability (P_b). The optimum power adaptation policy for hybrid RF-FSO channel is the water-filling policy [70]:

$$\frac{K^{\text{FSO}} P^{\text{FSO}}(i_t)}{P} = \begin{cases} \frac{1}{\Gamma_{\text{tsh}}} - \frac{1}{\Gamma^{\text{FSO}}}, & \Gamma^{\text{FSO}} \geq \Gamma_{\text{tsh}}, \\ 0, & \Gamma^{\text{FSO}} < \Gamma_{\text{tsh}} \end{cases}, \quad \Gamma^{\text{FSO}} = \Gamma_0^{\text{FSO}} i_t^2$$

$$\frac{K^{\text{RF}} P^{\text{RF}}(h)}{P} = \begin{cases} \frac{1}{\Gamma_{\text{tsh}}} - \frac{1}{\Gamma^{\text{RF}}}, & \Gamma^{\text{RF}} \geq \Gamma_{\text{tsh}}, \\ 0, & \Gamma^{\text{RF}} < \Gamma_{\text{tsh}} \end{cases}, \quad \Gamma^{\text{RF}} = \Gamma_0^{\text{RF}} h^2$$
(2.29)

where Γ_{tsh} is the threshold SNR, which is common to both channels; $K^{\text{FSO}} = -1.85 \ln(5P_b)$, $K^{\text{RF}} = -1.5 \ln(5P_b)$ (with P_b being the target BER) and h is the RF fading coefficient; P^{FSO} (P^{RF}) denotes the portion of the power allocated to the FSO (RF) channel. Equation (2.29) allows us optimally to allocate the powers between FSO and RF channels depending on channel conditions.

In Figure 2.14, which is related to the FSO with RF feedback system, we show spectral efficiencies that can be achieved using optimum power and rate adaptation and M -ary pulse-amplitude modulation (MPAM) for different target bit error probabilities, and both: (a) weak turbulence regime ($\sigma_R = 0.2$, $\alpha = 51.913$, $\beta = 49.113$), and (b) strong turbulence regime ($\sigma_R = 2$, $\alpha = 4.3407$, $\beta = 1.3088$). For example, the spectral efficiency R/B of 2 bits/s/Hz at $P_b = 10^{-9}$ is achieved for symbol SNR of 23.3 dB in weak turbulence regime, and 26.2 dB in strong turbulence regime. In the same figure, we also report spectral efficiencies that can be achieved by employing the simple channel inversion power adaptation scheme. Interestingly, in the weak turbulence regime (Figure 2.14(a)) this performs comparable to the optimum adaptive-power adaptive-rate scheme. However, in the strong turbulence regime (see Figure 2.14(b)) this scheme faces significant performance degradation. Therefore, in the strong turbulence regime we have to employ the truncated channel inversion scheme, instead. In Figure 2.14, we also report spectral efficiencies that can be achieved by employing the truncated channel inversion. This scheme performs comparable to the optimum adaptation scheme in the weak turbulence regime, while at $P_b = 10^{-9}$ and for spectral efficiency of 2 bits/s/Hz it performs 3.7 dB worse, but significantly better than the simple channel inversion scheme.

By using trellis coded modulation (TCM), introduced by Ungerboeck [72], in combination with convolutional codes or by using coset codes, introduced by Forney [73], in combination with block codes, we can separate the encoding and modulation process (see [68] for more details). However, to keep the complexity of this approach reasonably

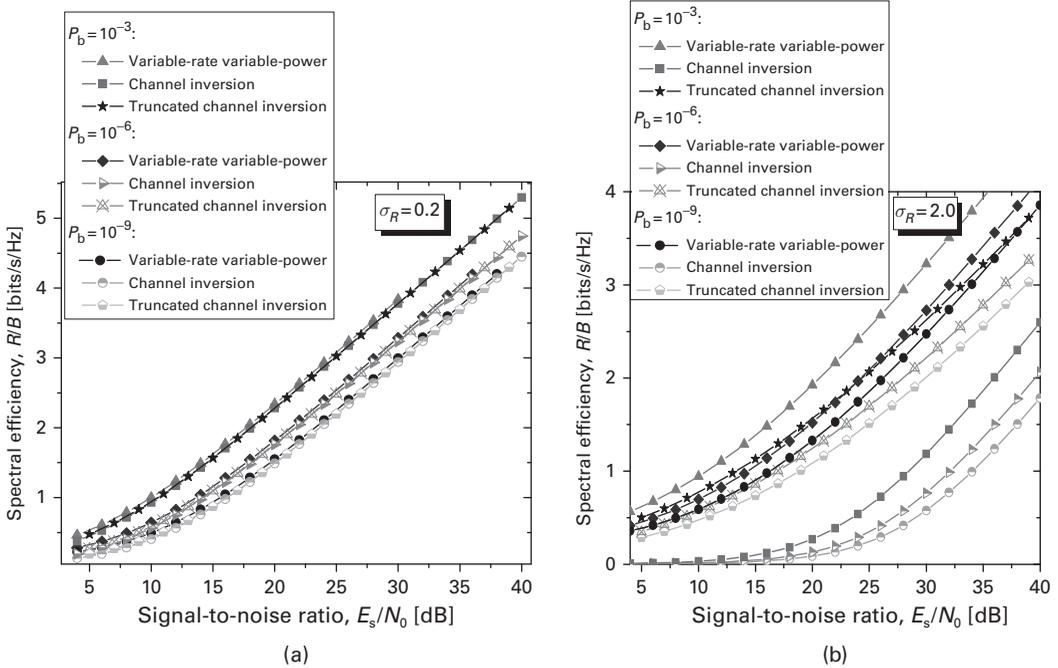


Figure 2.14 Feedback FSO system spectral efficiencies against symbol SNR for different target bit probabilities of error: (a) adaptive modulation in weak turbulence regime, and (b) adaptive modulation in strong turbulence regime. (After ref. [71]; ©IEEE 2010; reprinted with permission.)

low the convolutional or block codes should be simple and short. Those codes are in principle of low rate and weak so that coding gains are moderate. For example, the adaptive coding scheme based on TCM proposed in [68] is about 5 dB away from channel capacity. Instead, we have recently proposed [70],[71] to implement adaptive coding based on LDPC-coded modulation. The input data are LDPC encoded and written to a buffer. Based on FSO channel irradiance, i_t , $m(i_t)$ bits are taken at a time from a *buffer* and used to select the corresponding point from PAM signal constellation. The number of bits to be taken from the buffer is determined by one of three methods described above.

In Figure 2.15(a), which is also related to the FSO with RF feedback system, we show the spectral efficiency performance of adaptive LDPC(16935,13550)-coded MPAM for different adaptation scenarios. Given the fact that the channel capacity of the FSO channel under atmospheric turbulence is an open problem, we show in the same figure an upper bound in the absence of atmospheric turbulence (AWGN channel capacity) from [74]. The coding gain over adaptive modulation at $P_b = 10^{-6}$ for $R/B = 4$ bits/s/Hz is 7.2 dB in both (weak and strong) turbulence regimes. Larger coding gains are expected at lower BERs, and for higher spectral efficiencies. Further improvements can be obtained by increasing the girth of LDPC codes, and employing better modulation formats. The increase in codeword length to 100,515 does not improve R/B performance that

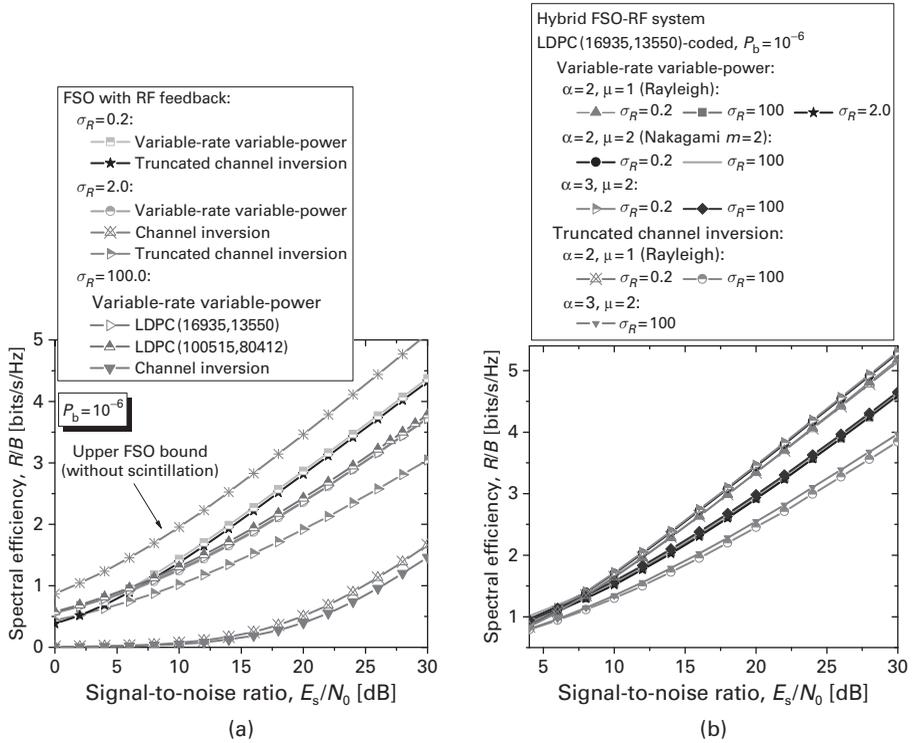


Figure 2.15 Spectral efficiencies against symbol SNR for adaptive LDPC-coded modulation: (a) FSO with RF feedback only, and (b) hybrid FSO-RF system.

much as shown in Figure 2.15(a). It is interesting to notice that by employing adaptive coding, the communication under saturation regime ($\sigma_R^2 \gg 1$) is possible, as shown in Figure 2.15(a). Moreover, for the variable-rate variable-power scheme there is no degradation in saturation regime compared to strong turbulence regime, while small degradation was found for channel inversion scheme. Overall improvement from adaptive modulation and coding for $R/B = 4$ bits/s/Hz at $P_b = 10^{-6}$ over non-adaptive uncoded modulation ranges from 10.5 dB (3.3 dB from adaptive modulation and 7.2 dB from coding) in weak turbulence regime and 38.9 dB in strong turbulence regime (31.7 dB from adaptive modulation and 7.2 dB from coding).

In Figure 2.15(b), we show the R/B performance of a hybrid FSO-RF system with adaptive LDPC(16935,13550)-coded modulation (PAM is used in FSO subsystem and QAM in RF subsystem) for different adaptation scenarios. For a spectral efficiency of 4 bits/s/Hz at a BER of 10^{-6} , the improvement of the hybrid FSO-RF system over the FSO system is 5.25 dB in Rayleigh fading ($\alpha = 2, \mu = 1$), 5.51 dB in Nakagami $m = 2$ fading ($\alpha = 2, \mu = 2$) and 5.63 dB in $\alpha = 3, \mu = 2$ fading (see [70] for the definition of (α, μ) -fading). For spectral efficiency of 2 bits/s/Hz at the same BER, the improvement of hybrid FSO-RF system over the FSO system is 3.32 dB in Rayleigh fading, 3.72 dB in Nakagami $m = 2$ fading, and 3.86 dB in $\alpha = 3, \mu = 2$ fading.

2.6 Multidimensional coded modulation for FSO communications

In order to satisfy high-bandwidth demands of future optical networks and solve interoperability problems, while keeping the system cost and power consumption reasonably low, the use of multidimensional coded modulation schemes, initially proposed for fiber-optic networks [75],[84] is described in this section. The key idea behind this approach is to exploit various degrees of freedom already available for the conveyance of information on a photon such as frequency, time, phase, amplitude, and polarization to improve the photon efficiency, while keeping the system cost reasonably low. This scheme was called generalized hybrid subcarrier/amplitude/phase/polarization (GH-SAPP) modulation in [80]. The GH-SAPP is composed of three or more hybrid amplitude/phase/polarization (HAPP) subsystems modulated with different subcarriers that are multiplexed together. At any symbol rate and code rate, GH-SAPP is capable of achieving the aggregate rate of the individual HAPP systems it is composed of, without introducing any BER performance degradation, as long as orthogonality among subcarriers is preserved. In this section, coding is done using QC LDPC codes, described in Section 2.2. The QC LDPC codes are chosen to simplify decoder implementation, and to reduce the encoding complexity in comparison to the random codes, as encoding is done using linear shift register circuitry. This technique is demonstrated by 32-GH-SAPP, which in combination with commercially available 10 Gb/s equipment achieves the aggregate data rate of 110 Gb/s, thus representing the 100G Ethernet enabling FSO technology.

In a GH-SAPP system, N input bit streams from different information sources are divided into L groups of bit streams with N_l streams for the l th group. The selection process for N_1, N_2, \dots, N_L is governed by two factors, the required aggregate rate, and the polyhedron of choice. Each of the N_l streams in the l th group is then used as input to an HAPP transmitter, where it is modulated with a unique subcarrier. The outputs of the L HAPP transmitters are then forwarded to a power combiner in order to be sent over the FSO channel. At the receiver side, the signal is split into L branches and forwarded to the L HAPP receivers. Figure 2.16(a) shows, without loss of generality, the block diagram of the 32-H-SAPP system configuration where $N = 11$ and $L = 4$; N_1, N_2, N_3 and N_4 are 4, 2, 2, and 3 respectively; N_1 and N_2 represent a dodecahedron of 20 vertices and 12 faces, and N_3 and N_4 represent the dual icosahedron of 12 vertices. Figure 2.16(b) shows the block diagram of the coded HAPP transmitter: N_l input bit streams from l different information sources, pass through identical encoders that use structured LDPC codes with code rate $r = k/n$ where k represents the number of information bits, and n represents the codeword length. As shown in the figure, the outputs of the encoders are interleaved by an $N_l \times n$ bit-interleaver where the sequences are written row-wise and read column-wise. The output of the interleaver is sent, N_l bits at a time instant i , to a mapper. The mapper maps each N_l bits into a 2^{N_l} -ary signal constellation point on a vertex of a polyhedron inscribed in a Poincaré sphere based on a lookup table (LUT). The ensemble of all the vertices of the HAPP systems forms the vertices of the regular polyhedron and its dual in GH-SAPP. The signal is then modulated by the

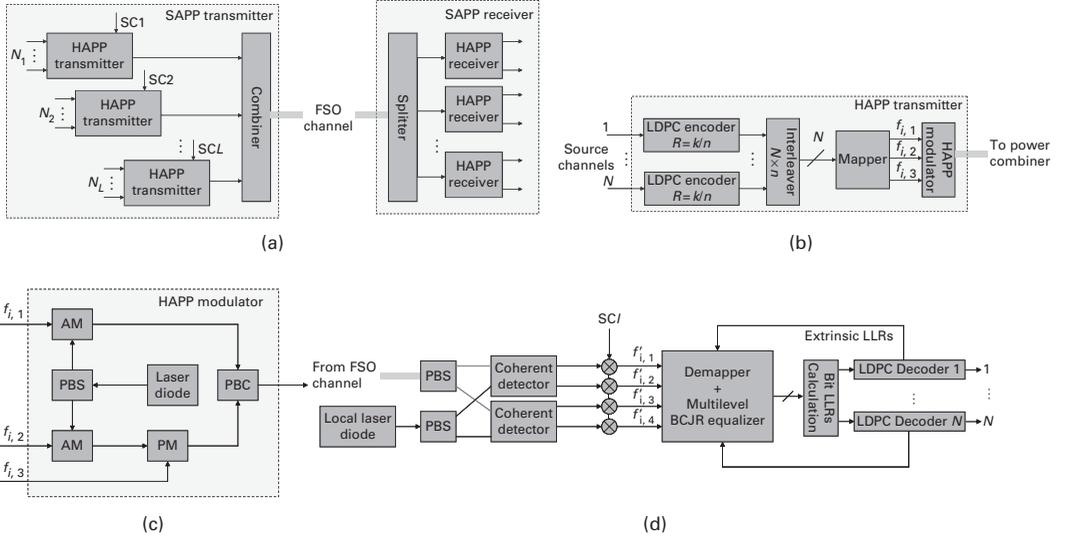


Figure 2.16 GH-SAPP bit-interleaved LDPC-coded modulation block diagrams: (a) 32-H-SAPP system, (b) HAPP transmitter, (c) HAPP modulator, and (d) HAPP receiver configurations. SC: Subcarrier, AM: Amplitude modulator, PM: Phase modulator, PBS: Polarization beam splitter and PBC: Polarization beam combiner.

HAPP modulator. The HAPP modulator, shown in Figure 2.16(c), is composed of two amplitude modulators and one phase modulator. The three voltages ($f_{1,i}, f_{2,i}, f_{3,i}$) needed to control these modulators are defined in an LUT based on Equation (2.30),

$$s_1 = a_x^2 - a_y^2, \quad s_2 = 2a_x a_y \cos \delta, \quad s_3 = 2a_x a_y \sin \delta, \quad \delta = \phi_x - \phi_y, \quad (2.30)$$

where

$$E_x(t) = a_x(t) e^{j[\omega t + \phi_x(t)]}, \quad E_y(t) = a_y(t) e^{j[\omega t + \phi_y(t)]}. \quad (2.31)$$

Since the designed polyhedrons are inscribed in a Poincaré sphere, Stokes parameters are used to define the coordinates of the vertices. The Stokes parameters shown in Equation (2.30), are then converted into amplitude and phase parameters according to Equation (2.31). Without loss of generality, we can assume that $\phi_x = 0$ and hence $\delta = -\phi_y$. This yields a system of three equations with three unknowns that can easily be solved. Figure 2.16(d) shows the block diagram of the HAPP receiver. The signal from the FSO channel is passed into two coherent detectors as shown in the figure, the four outputs of the detectors provide all the information needed for the amplitudes and phases for both polarizations. These outputs are then demodulated by the subcarrier specified for the corresponding HAPP receiver, then sampled at the symbol rate to be forwarded to the APP demapper. The output of the APP demapper is then forwarded to the bit LLRs calculator which provides the LLRs required for the LDPC decoding process. The extrinsic information is then iterated back and forth between the LDPC decoder and the APP demapper to improve BER performance.

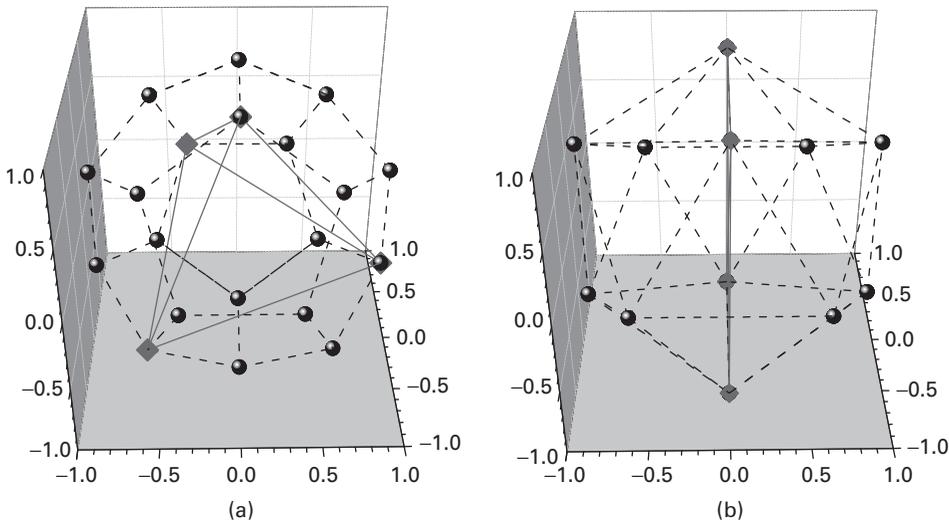


Figure 2.17 32-GH-SAPP constellation points corresponding to: (a) dodecahedron and (b) icosahedron.

The 32-GH-SAPP constellation diagram, shown in Figure 2.17, has constellation points placed in vertices of a dodecahedron (Figure 2.17(a)) and its dual (Figure 2.17(b)). This configuration utilizes four subcarriers; the first two subcarriers are used to modulate the points of the dodecahedron vertices, and the other two subcarriers are used for the vertices of the dual (icosahedron). The 32-GH-SAPP uses 11-bitstream input, grouped into four groups. The first group maps the input from the first four bitstreams onto 16 points of the 20 of the dodecahedron. The second group maps the input of two bitstreams onto the four vertices that form a tetrahedron. The selection of vertices for a subcarrier is done to maximize the distance between the points on the same subcarrier. The third group maps the input from the two bitstreams onto 4 points of the 12 of the icosahedron, while the fourth group maps the input of the remaining three bitstreams onto the remaining 8 vertices. By using 11 different 9.1 Gb/s data streams the aggregate rate of 100.1 Gb/s can be achieved, and 100 G Ethernet can be delivered to different nodes in the FSO network.

In Figure 2.18, we show the BER performance of the 32-GH-HAPP scheme in weak, medium, and strong atmospheric turbulence regimes. This scheme is compared against 32-QAM and polarization-multiplexed (PolMUX) 16-QAM. It is interesting to notice (see Figure 2.18(a)) that both 32-QAM and PolMUX 16-QAM exhibit BER floor phenomenon even in the weak turbulence regime, while the 32-GH-HAPP scheme does not exhibit the error floor in the observed region of SNRs (up to 35 dB) even under strong turbulence regime. Notice that for symbol rate of 10 Giga symbols/s (10 GS/s), the aggregate rate of 32-QAM is $R_D = 50$ Gb/s, the aggregate rate of PolMUX 16-QAM is 80 Gb/s, while the aggregate rate of 32-GH-SAPP is 110 Gb/s. The comparison of the proposed scheme was done with conventional modulation schemes having the same number of constellation points, so that the comparison is fair with respect

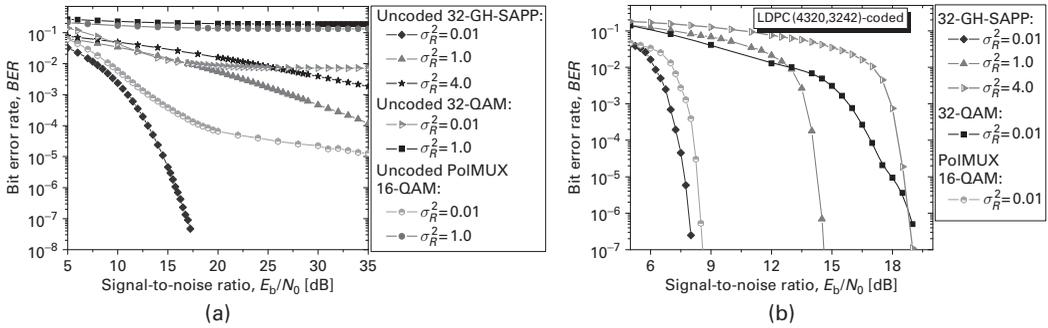


Figure 2.18 BER performance of proposed 32-GH-SAPP FSO system in weak ($\sigma_R^2 = 0.01$), medium ($\sigma_R^2 = 1$) and strong ($\sigma_R^2 = 4$) turbulence regimes: (a) uncoded and (b) QC LDPC(4320,3242)-coded cases. (After ref. [80]; ©IEEE 2010; reprinted with permission.)

to the atmospheric turbulence influence. Therefore, only the 32-GH-HAPP scheme is compatible with 100 G Ethernet while employing commercially mature 10 Gb/s technology. In Figure 2.18(b), we provide simulation results for the LDPC-coded case, obtained for 25 LDPC decoder iterations and 3 APP demapper-LDPC decoder iterations. The 32-GH-HAPP scheme is able to operate in any atmospheric turbulence regime (ranging from weak to strong), PolMUX 16-QAM can only operate in weak turbulence regime, while 32-QAM cannot operate even in weak turbulence regime due to error floor phenomenon.

The GH-SAPP scheme can be used to solve various problems that current optical networking is facing. It can be used to: (i) enable ultra-high-speed transmission to end-users, (ii) allow interoperability of various RF and optical technologies, (iii) reduce installation costs, (iv) reduce deployment time, and (v) improve the energy efficiency of a communication link. The 32-GH-SAPP modulation, which employs mature 10 Gb/s optical technology, is 100 G Ethernet enabling technology. This scheme can operate under various atmospheric turbulence regimes (ranging from weak to strong), as indicated above, while still enabling 100 Gb/s traffic. The corresponding PolMUX 16-QAM scheme can only operate under the weak turbulence regime with aggregate data rate of 80 Gb/s.

2.7 Free-space optical OFDM communication

Optical OFDM systems can support high data rates by splitting a high-rate data stream into a number of low-rate data streams and transmitting these over a number of narrowband subcarriers [44]–[53]. The narrowband subcarrier data streams experience smaller distortions than high-speed ones and require no equalization. Moreover, most of the required signal processing is performed in the RF domain. This is advantageous because microwave devices are much more mature than their optical counterparts and because the frequency selectivity of microwave filters and the frequency stability of

microwave oscillators are significantly better than that of corresponding optical devices. Furthermore, the phase noise levels of microwave oscillators are significantly lower than that of distributed feedback (DFB) laser diodes, which means that RF coherent detection is easier to implement than optical coherent detection. This, in turn, allows a system architect to directly apply the most advanced coherent modulation formats already developed for wireless communication. The basic FSO-OFDM transmitter and receiver configurations are shown in Figure 2.19(a) and (b) respectively. The corresponding FSO link is shown in Figure 2.19(c). The input high-speed data-stream is de-multiplexed into lower data rate streams, each encoded by identical LDPC encoders, so that it can be implemented in currently existing hardware. The LDPC encoded outputs are further demultiplexed, and parsed into groups of B bits. The B bits in each group

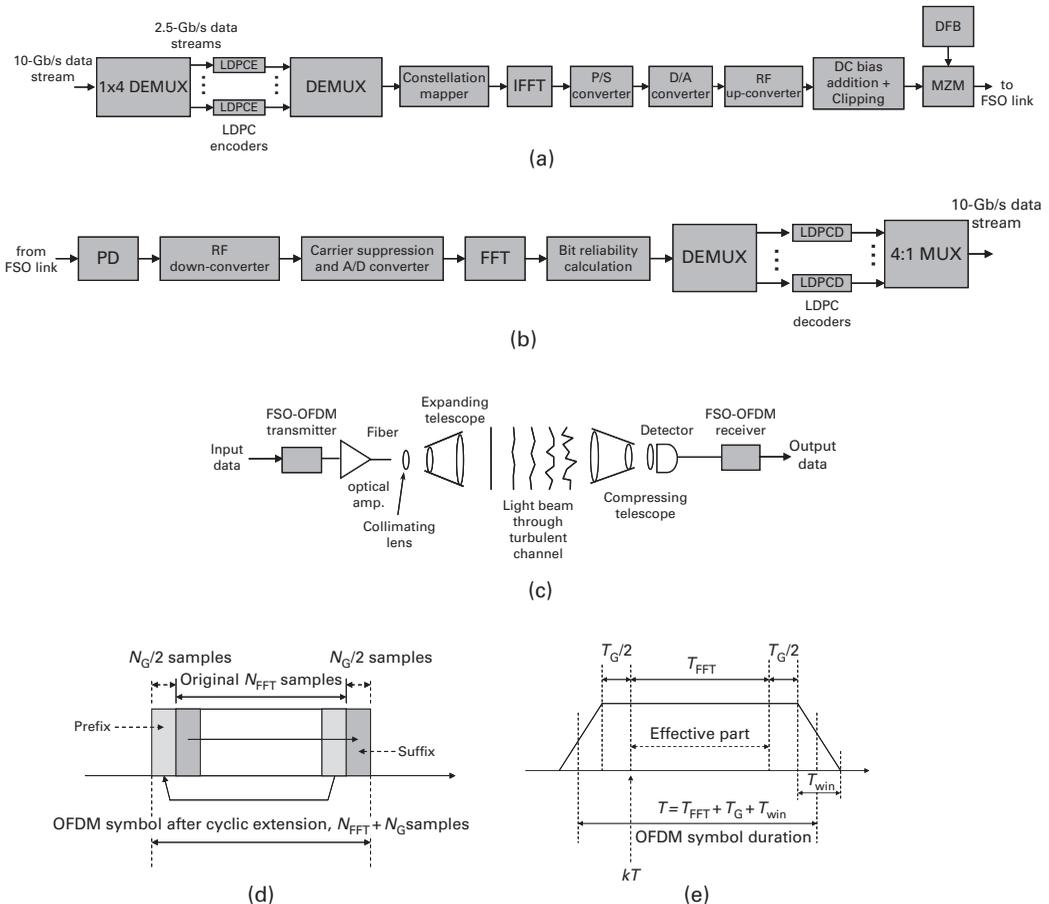


Figure 2.19 FSO-OFDM system: (a) transmitter configuration, (b) receiver configuration, (c) FSO link, (d) OFDM symbol cyclic extension, (e) OFDM symbol after windowing. LDPC: LDPC encoder, LPDCD: LDPC decoder, DFB: distributed-feedback laser, MZM: dual-drive Mach-Zehnder modulator, P/S: parallel-to-serial.

(frame) are subdivided into K subgroups with the i th subgroup containing b_i bits. The b_i bits from the i th subgroup are mapped into a complex-valued signal from a 2^{b_i} -point signal constellation such as QAM. The complex-valued signal points from K subchannels are considered to be the values of the discrete Fourier transform (DFT) of a multicarrier OFDM signal. Therefore, the symbol length (the time between two consecutive OFDM symbols) in an OFDM system is $T = KT_s$, where T_s is the symbol-interval length in an equivalent single-carrier system.

The transmitted OFDM signal can be written as

$$s(t) = s_{\text{OFDM}}(t) + D, \quad (2.32)$$

where D is DC bias and

$$s_{\text{OFDM}}(t) = \text{Re} \left\{ \sum_{k=-\infty}^{\infty} w(t - kT) \sum_{i=-N_{\text{FFT}}/2}^{N_{\text{FFT}}/2-1} X_{i,k} \cdot e^{j2\pi \frac{i}{T_{\text{FFT}}}(t-kT)} e^{j2\pi f_{\text{RF}} t} \right\}, \quad (2.33)$$

which is defined for $kT - T_G/2 - T_{\text{win}} \leq t \leq kT + T_{\text{FFT}} + T_G/2 + T_{\text{win}}$. In the above expression $X_{i,k}$ denotes the k th OFDM symbol in the i th subcarrier, $w(t)$ is the window function, and f_{RF} is the RF carrier frequency. The duration of the OFDM symbol is denoted by T , while T_{FFT} is the FFT sequence duration, T_G is the guard interval duration (the duration of cyclic extension), and T_{win} is the length of the windowing interval.

The details of the resulting OFDM symbol are shown in Figure 2.19(d)–(e). The symbols are generated as follows. $N_{\text{QAM}} (= K)$ consecutive input QAM symbols are zero-padded to obtain $N_{\text{FFT}} (= 2^m, m > 1)$ input samples for inverse fast Fourier transform (IFFT), then N_G samples are inserted to create the guard interval T_G and finally the OFDM symbol is multiplied by the window function (raised cosine function is used in [47], but the Kaiser, Blackman–Harris and other window functions are also applicable). The purpose of the cyclic extension is to preserve the orthogonality among subcarriers when the neighboring OFDM symbols partially overlap, and the purpose of the windowing is to reduce the out-of-band spectrum. The cyclic extension, illustrated in Figure 2.19(d), is performed by repeating the last $N_G/2$ samples of the FFT frame (of duration T_{FFT} with N_{FFT} samples) as the prefix, and repeating the first $N_G/2$ samples (out of N_{FFT}) as the suffix. After a D/A conversion and RF up-conversion, we convert the RF signal to the optical domain using one of two options: (i) for symbol rates up to 10 Gsymbols/s the OFDM signal directly modulates the DFB laser, and (ii) for symbol rates above 10 Gsymbols/s the OFDM signal drives the dual-drive Mach-Zehnder modulator (MZM). The DC component (D in Eq. (2.32)) is inserted to enable incoherent recovery of the QAM symbols. In the remainder of this section, three different OFDM schemes are described. The first scheme is based on intensity modulation, and shall be referred to as the “biased-OFDM” (B-OFDM) scheme. Because bipolar signals cannot be transmitted over an IM/DD link, it is assumed that the bias component D is sufficiently large so that when added to $s_{\text{OFDM}}(t)$ the resulting sum is non negative. The main disadvantage of the B-OFDM scheme is the poor power efficiency. To improve the power efficiency we describe two alternative schemes [45]. The first of

these, which we shall refer to as the “clipped-OFDM” (C-OFDM) scheme, is based on single-side band (SSB) transmission, with clipping of the negative portion of the OFDM signal after bias addition. The bias is varied to find the optimum one for fixed optical launched power. It was found that the optimum case is one in which $\sim 50\%$ of the total electrical signal energy before clipping is allocated for transmission of a carrier. To convert the signal from double-side band (DSB) to SSB, we have two options: (i) to use Hilbert transformation of the in-phase signal as the quadrature signal in electrical domain, and (ii) to perform DSB to SSB transformation by an optical filter. The second power-efficient scheme, which we shall refer to as the “unclipped-OFDM” (U-OFDM) scheme, is based on SSB transmission employing LiNbO₃ MZM. To avoid distortion due to clipping, the information-bearing signal is transmitted by modulating the electrical field (instead of intensity modulation employed in the B-OFDM and C-OFDM schemes) so that the negative part of the OFDM signal is transmitted to the photodetector. Distortion introduced by the photodetector, caused by squaring, is successfully eliminated by proper filtering, and the recovered signal distortion is insignificant. Notice that U-OFDM is less power efficient than C-OFDM because the negative portion of the OFDM signal is transmitted and then discarded. For U-OFDM the detector non-linearity is compensated by post-detection filters that reject (potentially useful) signal energy and compromise power efficiency. Despite this drawback we find that U-OFDM is still significantly more power efficient than B-OFDM. Note that the DC bias shifts the average of the C-OFDM signal towards positive values, while in the case of B-OFDM a much larger bias is needed to completely eliminate the negative portion of the signal.

The point-to-point FSO system considered here, shown in Figure 2.19(c), consists of an FSO-OFDM transmitter, propagation medium and an FSO-OFDM receiver. The modulated beam is projected toward the receiver using the expanding telescope. At the receiver, an optical system collects the light, and focuses it onto a detector, which delivers an electrical signal proportional to the incoming optical power. The receiver commonly employs the transimpedance design, which is a good compromise between noise and bandwidth. A PIN photodiode plus preamplifier or an avalanche photodiode are typically used as optical detectors. During propagation through the air, the optical beam experiences amplitude and phase variations caused by scattering, refraction caused by atmospheric turbulence, absorption, and building sway. The photodiode output current can be written as

$$\begin{aligned} i(t) &= R_{PD} |a(t) s_{OFDM}(t) + a(t) D|^2 \\ &= R_{PD} \left[|a(t) s_{OFDM}(t)|^2 + |a(t) D|^2 + 2R_e \{a(t) s_{OFDM}(t) a^*(t) D\} \right], \quad (2.34) \end{aligned}$$

where $|a(t)|^2$ denotes the intensity fluctuation due to atmospheric turbulence, and R_{PD} denotes the photodiode responsivity.

After RF down-conversion, the A/D conversion and cyclic extension removal, the transmitted signal is demodulated by the FFT algorithm. The soft outputs of the FFT demodulator are used to estimate the bit reliabilities that are fed to four identical LDPC

iterative decoders based on the sum-product algorithm. The parameters of the overall OFDM-FSO system must be carefully chosen so that the reconstructed sequence constellation diagram suffers minimal distortion in a back-to-back configuration. LDPC codes have been shown to significantly outperform turbo-product codes in bursty-error prone channels such as the fiber-optics channel in the presence of intrachannel nonlinear effects [54]. The quasi-cyclic LDPC codes similar to those described in Section 2.2 are employed here. In FSO communications the receiver electronics noise is commonly modeled as a Gaussian noise (see e.g., [44]–[46],[55]). If r_I is the in-phase demodulator sample, and r_Q is the quadrature demodulator sample, then the *symbol* LLRs are calculated by

$$\lambda(s = (s_I, s_Q)) = -\frac{(r_I - s_I)^2}{2\sigma^2} - \frac{(r_Q - s_Q)^2}{2\sigma^2}, \quad (2.35)$$

where s_I and s_Q are the coordinates of a transmitted signal constellation point and the AWGN variance (σ^2) is determined from the required electrical signal-to-noise ratio (SNR) per bit E_b/N_0

$$\frac{E_b}{N_0} = \frac{E\{s_{i,k}\} P_0}{\log_2 M \sigma^2}. \quad (2.36)$$

P_0 is the normalized received power, and $s_{i,k}$ denotes the QAM symbol in the k th sub-carrier channel of the i th OFDM frame. (With M we denote the number of points in the corresponding constellation diagram.) Notice that the definition of electrical SNR per bit, common in digital communications (see [47],[56],[57]), is different from that used in [1]. The Gaussian assumption in Eq. (2.35) may lead to BER performance degradation because the joint distribution is actually a convolution of a Gaussian and a gamma-gamma PDF. In order to reduce complexity, we use the Gaussian approximation in the calculation of symbol reliabilities. Nevertheless, a dramatic performance improvement of an LDPC-coded FSO-OFDM system over an LDPC-coded FSO OOK system is obtained.

Simulation results of an LDPC-coded SSB U-OFDM system for two different turbulence strengths, and zero inner scale, are shown in Figure 2.20. For BPSK and QPSK, the coding gain improvement of an LDPC-encoded FSO-OFDM system over an LDPC-encoded FSO OOK system increases as the turbulence strength increases. However, the 16-QAM FSO-OFDM system is not able to operate in the regime of strong turbulence because the required SNR is too high to be of practical importance. For weak turbulence ($\sigma_R = 0.6$) (see Figure 2.20(a)) the coding gain improvement of LDPC-coded FSO-OFDM system with 64 subcarriers over the LDPC-encoded FSO OOK system is 8.47 dB for QPSK and 9.66 dB for BPSK, at the BER of 10^{-5} . For strong turbulence ($\sigma_R = 3.0$) (see Figure 2.20(b)) the coding gain improvement of the LDPC-coded FSO-OFDM system over the LDPC coded FSO OOK system is 20.24 dB for QPSK and 23.38 dB for BPSK. In both cases, the QC-LDPC (4320,3242) [58] of rate 0.75 is employed.

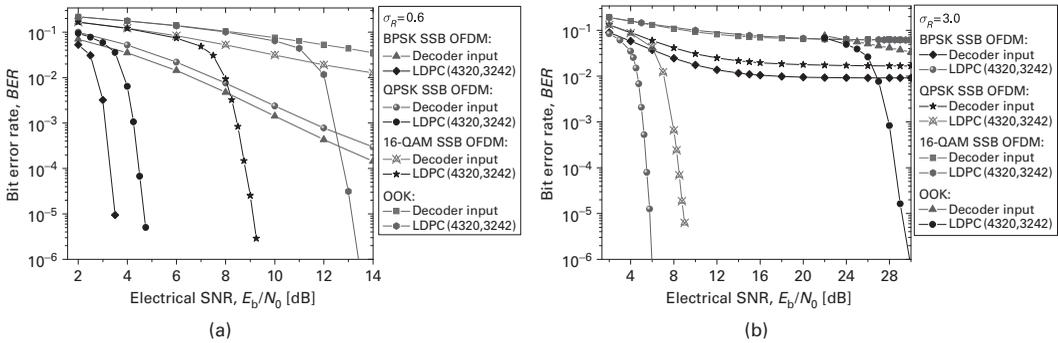


Figure 2.20 BER performance of LDPC-coded SSB U-OFDM system with 64-subcarriers under: (a) the weak turbulence ($\sigma_R = 0.6$), and (b) strong turbulence ($\sigma_R = 3.0$).

2.8 Heterogeneous optical networks (HONs)

Internet traffic has continued its rapid growth over the last few years due to the increased popularity of the Internet. Other major contributors to this growth are the new applications that are emerging, and continuing the demand for higher bandwidths [16], such as VoIP, YouTube, IPTV, and HDTV. For instance, a single uncompressed HDTV stream (1080i) requires a data rate of about 1.5 Gb/s [76]. Such data rates are too high for conventional wireless LAN systems. This bottleneck can be solved by employing ultra wideband (UWB) communications operating in 3.1–10.6 GHz range [77] or more recently proposed 60 GHz radio operating in 57–64 GHz range (in North America) [76]. Combining transmission of UWB/60GHz signal over fiber optical links by employing radio-over-fiber (RoF) is a promising technology for extending the coverage of UWB/60 GHz radios. However, this approach requires expensive optical components in addition to frequent electrical-to-optical (E-O) and optical-to-electrical (O-E) conversions. It is also inefficient in terms of optical bandwidth utilization, as it transmits low-speed wireless signals over high-bandwidth optical channel. Another approach to deliver high-speed signal to an end-user is through the use of passive optical networks (PONs) [78]. In order to reduce the system cost of PONs and speed up the installation cost, we proposed recently the use of heterogeneous optical networking (HON) [48],[49]. The HON can provide high-bandwidth and solve interoperability problems of future optical networks, while keeping the system cost and power consumption reasonably low. With HON, we can deliver a high-speed optical signal, i.e. 40 Gb/s and beyond, to an end-user by using free-space optical (FSO), plastic optical fiber (POF)/multimode fiber (MMF) and indoor optical wireless communication (OWC) links (also known as IR links). The HON scheme provides solutions to two major obstacles; bandwidth limitation and atmospheric turbulence. To deal with bandwidth limitations of POF/MMF and OWC links we employ power-variable rate-adaptive LDPC-coded OFDM. On the other hand, to deal with atmospheric turbulence we utilize the spatial diversity.

Various applications of interest of the proposed heterogeneous optical networking scenario, depicted in Figure 2.21, include: (i) in cellular systems to establish the connection between mobile telephone switching office (MTSO) and base stations (BSs) (see Figure 2.21(a)), (ii) in WiMAX to extend the coverage and reliability by connecting WiMAX BSs with FSO, hybrid (FSO-RF) or heterogeneous (FSO-POF/MMF) links (see Figure 2.21(b)), (iii) in UWB communications to extend the wireless coverage range by using radio over FSO/fiber technologies (see Figure 2.21(c)), (iv) in access networks (Figure 2.21(d)) to increase data rate and reduce system cost and deployment time, (v) in ground-to-satellite/satellite-to-ground FSO communications to increase the data rate (Figure 2.21(e)), (vi) in intersatellite FSO communications (Figure 2.21(e)), and (vii) in aircraft-to-satellite/satellite-to-aircraft communications (Figure 2.21(e)). In order to reduce system installation and maintenance costs for indoor applications, the POFs or MMFs can be used from residential gateway to either fixed or mobile wireless units inside the building. The HON schemes offer many advantages over wireless networks such as low attenuation loss, large bandwidth, improved security, reduced power consumption, and easy installation and maintenance. The heterogeneous communication system is also an excellent candidate to be used instead of PON applications, to substitute various MMF or SMF links, while reducing system cost and speeding up the installation process. With the FSO link being used as the transmission media, free-space optical access networks (FSO-ANs) can offer much higher bandwidth and better energy efficiency, while supporting various communication services. Instead of optical couplers used in PONs we can use amplify-and-forward (AF) FSO relays based on semiconductor

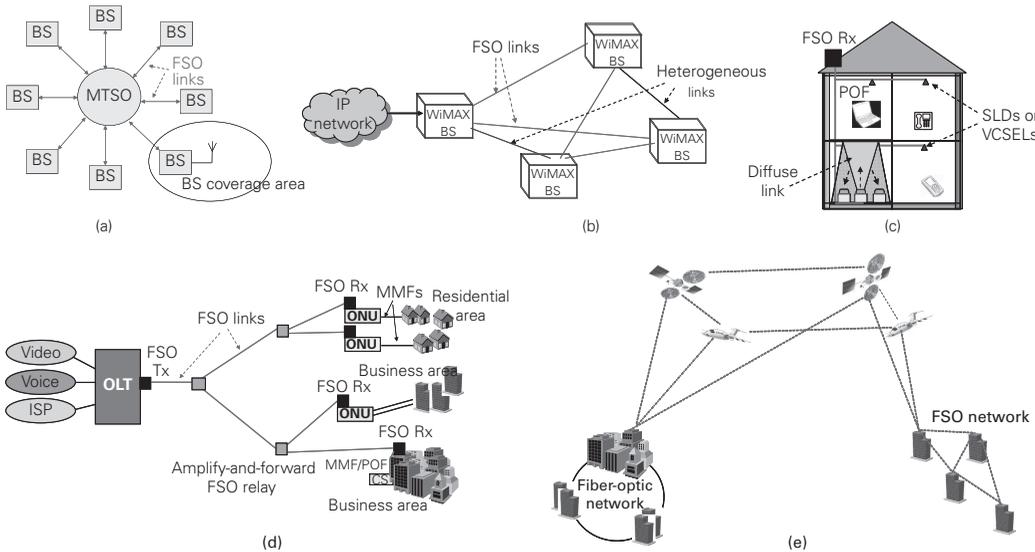


Figure 2.21 Heterogeneous optical networking: (a) cellular network, (b) WiMAX network, (c) indoor optical network, (d) optical access network, and (e) satellite-to-ground/ground-to-satellite, inter-satellite, and satellite-to-aircraft/aircraft-to-satellite FSO communications. OLT: optical line terminal, ONU: optical network unit, CS: central station.

optical amplifiers. FSO-ANs have many advantages compared to other conventional copper-based access technologies. For example, FSO-ANs offer an improved bandwidth while supporting various communication services. In combination with OFDM, FSO-ANs have unique flexibility in dealing with bandwidth resource sharing and virtualization, in addition to protocol independence, service transparency, scalability, and cost-effectiveness. Namely, in OFDM based FSO-ANs we assign to every particular optical network unit (ONU) a subset of subcarriers. This subset of subcarriers can employ different QAM constellation sizes depending on channel SNR. Differentiated quality of service (QoS) can easily be achieved by assigning high-quality subcarriers (with high SNR) to higher priority users. Moreover, this OFDM scheme can operate with simple medium access control (MAC) with low overhead.

Figure 2.22 shows the block diagram of the heterogeneous optical communication system (one possible link from Figure 2.21(d)). In this setup, the LDPC encoded data enters the buffer in the adaptive OFDM block. Depending on the i th subcarrier signal-to-noise ratio (SNR), m_i bits are taken from the buffer. The mapper that follows the buffer selects a corresponding constellation point from 2^{m_i} -QAM constellation diagram. After pilot insertion and serial-to-parallel (S/P) conversion, the inverse FFT (IFFT) is performed. The cyclic extension is performed by repeating the $N_G/2$ samples (of IFFT frame) as prefix and the first $N_G/2$ samples as suffix. The DC bias is added so that real-valued OFDM signal can be transmitted over the FSO system with direct detection. The modulated beam is projected toward the distant receiver array by using an expanding telescope assembly. By providing that the aperture diameter of each receiver is smaller than spatial correlation width of the irradiance function, the array elements will be sufficiently separated so that they act independently. Possible FSO spatial diversity techniques include: equal gain combining (EGC), maximum gain combining (MGC),

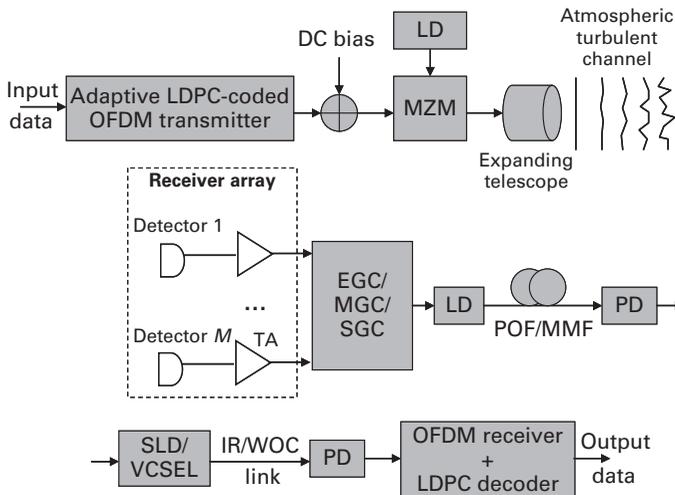


Figure 2.22 Block diagram of possible heterogeneous access optical communication systems. LD: laser diode, PD: photodetector, TA: transimpedance amplifier.

and selection gain combining (SGC). The signal is then re-modulated and transmitted either over MMF or POF. Upon photodetection at the end of the MMF/POF link, the signal is re-transmitted over the indoor OWC/IR link using either a super-luminescent diode (SLD) or VCSEL. If MMF is used instead of POF, it might be possible to transmit the signal from OLT to an end-user even without any intermediate O-E and E-O conversions.

We have recently proposed in [48] to use the adaptive “water-filling” algorithm [81] as a means to deal with the limited bandwidth of POF/MMF and OWC links, while to deal with atmospheric turbulence to use spatial receiver diversity is recommended as illustrated in Figure 2.22. This algorithm is used to determine the optimum power to be allocated to the i th subcarrier P_i as follows

$$P_i/P = \begin{cases} 1/\gamma_{\text{tsh}} - 1/\gamma_i, & \gamma_i \geq \gamma_{\text{tsh}} \\ 0, & \text{otherwise.} \end{cases} \quad (2.37)$$

In Eq. (2.37) P denotes the total available power, and γ_i denotes the signal-to-noise ratio (SNR) in the i th subcarrier. This SNR is defined as $\gamma_i = |H(i)|^2 P/N_0$ where $H(i) = H_{\text{POF}}(i)H_{\text{OWC}}(i)$ denotes the transfer function of the i th subcarrier obtained by the product of the transfer functions of POF, $H_{\text{POF}}(i)$, (as defined by Yabre [82]) and OWC links, $H_{\text{OWC}}(i)$; N_0 denotes the power spectral density of the receiver electronics noise. The variable-rate adaptation can be achieved by choosing the maximum product of integer m_i , corresponding to the number of bits per i th subcarrier, and code rate R of the corresponding LDPC code as follows,

$$m_i R \leq K \log_2 (\gamma_i/\gamma_{\text{tsh}}) = C_i, \quad (2.38)$$

where C_i (bits/channel use) denotes the channel capacity of the i th subcarrier, and K is the constant dependent on modulation format being used. The rate-adaptive LDPC code is designed using the quasi-cyclic or progressive edge growth (PEG) code design [79]. The BER performance of a rate-adaptive LDPC code of codeword length 16935 with possible code rates $\{0.9, 0.875, 0.85, 0.8, 0.75, 0.7\}$ is shown in Figure 2.23.

Figure 2.24 shows the spectral efficiencies of the proposed heterogeneous FSO-POF-OWC system, corresponding to OFDM signal bandwidth of 7.5 GHz with 128 subcarriers. We report the corresponding plots for 100 m of GIPOF50 (from Thorlabs) at 1330 nm and for different FSO turbulence strengths. The OWC link typically contains LOS and diffuse components. In Figure 2.24, we observe the worst-case scenario, where only the diffuse component is present (see Figure 2.21(c)). The diffuse component is commonly modeled as the low pass filter [16] of bandwidth B_{OWC} . To model the propagation of signal over the POF links we employ a comprehensive model due to Yabre [82]. This model includes mode coupling and mode attenuation. We see that beyond 40 Gb/s transmission over the proposed heterogeneous optical system is possible for $M = 2$ FSO receivers and EGC at moderate SNRs. To concentrate on channel distortions only, we assumed that MZM, LD, and SLD/VCSEL modulation characteristics are ideal. In Figure 2.24(a), we provide the aggregate data rates for various FSO spatial diversity techniques (MGC, EGC, and SGC) under the strong turbulence regime ($\sigma_R^2 = 4$). For sufficient OWC link bandwidth MGC and EGC perform comparably

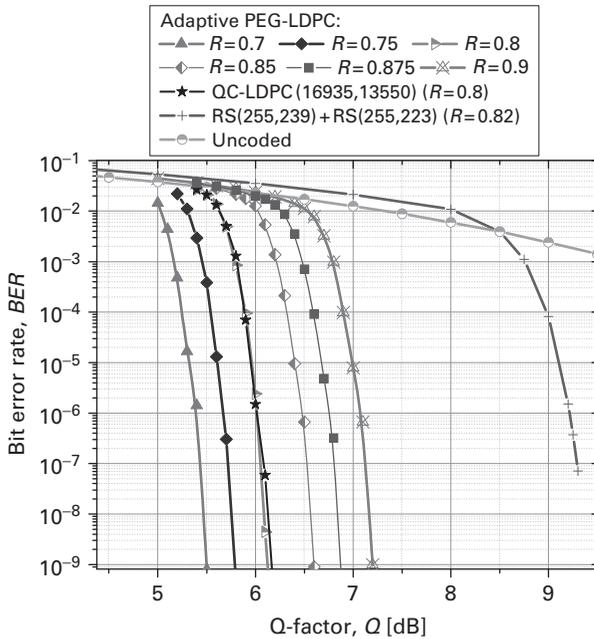


Figure 2.23 BER performance of rate-adaptive LDPC code. Q-factor, commonly used in optical communications is defined by $Q = (\mu_1 - \mu_0)/(\sigma_1 + \sigma_0)$, where μ_i and σ_i denote the mean value and standard deviation corresponding to bit $i \in \{0, 1\}$. (After ref. [48]; ©IEEE 2010; reprinted with permission.)

and slightly outperform SGC. However, when the OWC link bandwidth is significantly lower than the OFDM signal bandwidth, EGC is more robust compared to MGC. In simulations, we do not employ any optical amplifier, so that the major contributor to noise effects is thermal noise of the TA in the final receiver stage. The SNR in Figure 2.24 is, therefore, defined in the electrical domain upon photodetection in the last receiver stage.

In Figure 2.24(b), we provide a comparison between single FSO receiver and EGC with $M = 2$ FSO receivers for various turbulence regimes for fixed OWC link bandwidth of $B_{\text{OWC}} = 2.5$ GHz. For EGC and $M = 2$ FSO receivers the aggregate data rate of 40 Gb/s can be achieved at medium SNRs, while for single FSO receiver only for high SNR values. It is interesting to notice that communication even in the FSO saturation regime is possible with the heterogeneous communication scheme. In Figure 2.25(c) we compare the channel capacity derived aggregate data rates against LDPC-coded OFDM with QAM. For an aggregate data rate of 40 Gb/s under saturation regime when two FSO receivers are used and EGC we are 2.27 dB away from channel capacity when LDPC code of rate 0.875 is used. Finally, in Figure 2.24(d) we compare the aggregate data rates under the saturation regime for various code rates. When an LDPC code of rate 0.75 is used, for an aggregate rate of 40 Gb/s, we are 1.33 dB away from channel capacity; while for the $R = 0.7$ code we are only 1.04 dB away from channel capacity.

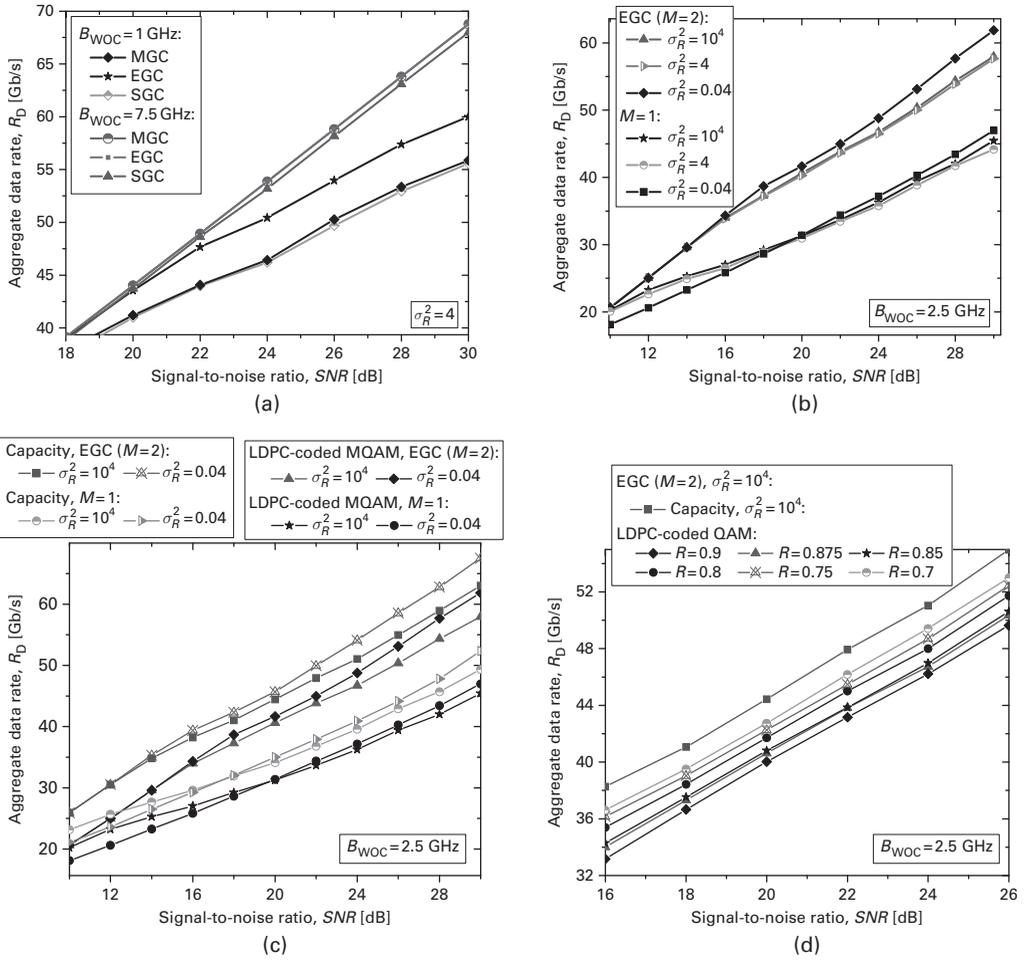


Figure 2.24 Aggregate data rates of heterogeneous access optical communication system at BER of 10^{-9} for LDPC code of rate $R = 0.875$: (a) for different FSO spatial diversity scenarios under the strong turbulence regime, (b) for various atmospheric turbulence regimes, (c) channel capacity derived aggregate data rates against LDPC-coded OFDM ($R = 0.875$), and (d) aggregate data rates in saturation regime for various code rates. (After ref. [48]; ©IEEE 2010; reprinted with permission.)

2.9 Summary

In this chapter, we described several coded-modulation concepts enabling communication over strong atmospheric turbulence channels including coded-MIMO, raptor coding, adaptive modulation and coding, multidimensional coded modulation, and coded-OFDM. We have shown that deep fades due to scintillation can be tolerated by these different approaches. In Section 2.1, we described the uncorrelated gamma-gamma distribution model valid under different turbulence regimes. Section 2.2 was devoted to

different codes on graphs suitable for use in FSO communications. In Section 2.3, we described coded-MIMO, and have shown that the repetition MIMO can be used to deal with scintillation. In Section 2.4, we described the use of raptor codes as a means of enabling the communication over temporal correlated FSO channels. Section 2.5 was devoted to adaptive modulation and coding. Two AMC schemes were described: AMC with RF feedback and a hybrid FSO-RF AMC scheme. Both schemes enable communication under strong atmospheric turbulence. The possibility of using multidimensional coded modulation schemes to enable ultra-high-speed optical communication was discussed in Section 2.6. In Section 2.7, we have shown that deep fades due to atmospheric turbulence can be tolerated by using the coded-OFDM in combination with interleaving. In Section 2.8, we discussed the possibility of implementation of heterogeneous optical networks, composed of various optical links (SMF, MMF, POF, and FSO links).

Acknowledgments

This chapter is supported in part by the NSF under Grants CCF-0952711 and ECCS-0725405.

References

- [1] L. C. Andrews and R. L. Phillips, *Laser Beam Propagation Through Random Media*. SPIE Optical Engineering Press, 1998.
- [2] H. Willebrand, B.S. Ghuman, *Free-Space Optics: Enabling Optical Connectivity in Today's Networks*. Sams Publishing, 2002.
- [3] A. Ishimaru, *Wave Propagation and Scattering in Random Media*. Academic, 1978, vol. 1–2.
- [4] L. C. Andrews, R. L. Phillips, and C. Y. Hopen, *Laser Beam Scintillation with Applications*. SPIE Press, 2001.
- [5] S. Karp, R. Gagliardi, S. E. Moran, and L. B. Stotts, *Optical Channels*. Plenum, 1988.
- [6] S. M. Flatté, C. Bracher, and G.-Y. Wang, “Probability-density functions of irradiance for waves in atmospheric turbulence calculated by numerical simulation,” *J. Opt. Soc. Am. A*, vol. 11, pp. 2080–2092, 1994.
- [7] L. C. Andrews and R. L. Phillips, “I-K distribution as a universal propagation model of laser beams in atmospheric turbulence,” *J. Opt. Soc. Am.*, vol. 2, pp. 160–163, 1985.
- [8] L. C. Andrews and R. L. Phillips, “Mathematical genesis of the I-K distribution for random optical fields,” *J. Opt. Soc. Am.*, vol. 3, pp. 1912–1919, 1986.
- [9] J. H. Churnside and R. J. Hill, “Probability density of irradiance scintillations for strong path-integrated refractive turbulence,” *J. Opt. Soc. Am. A*, vol. 4, pp. 727–733, 1987.
- [10] R. J. Hill and R. G. Frehlich, “Probability distribution of irradiance for the onset of strong scintillation,” *J. Opt. Soc. Am. A*, vol. 14, pp. 1530–1540, 1997.
- [11] J. H. Churnside and R. G. Frehlich, “Experimental evaluation of log-normally modulated Rician and IK models of optical scintillation in the atmosphere,” *J. Opt. Soc. Am. A*, vol. 6, pp. 1760–1766, 1989.

- [12] M. A. Al-Habash, L.C. Andrews, and R. L. Phillips, "Mathematical model for the irradiance probability density function of a laser beam propagating through turbulent media," *Opt. Eng.*, vol. 40, pp. 1554–1562, 2001.
- [13] X. Zhu, and J. M. Kahn, "Free-space optical communication through atmospheric turbulence channels," *IEEE Trans. Commun.*, vol. 50, pp. 1293–1300, 2002.
- [14] S. G. Wilson, M. Brandt-Pearce, Q. Cao, and J.J.H. Leveque, III, "Free-space optical MIMO transmission with Q-ary PPM," *IEEE Trans. Comm.*, vol. 53, pp. 1402–1412, Aug. 2005.
- [15] S. G. Wilson, M. Brandt-Pearce, Q. Cao, and M. Baedke, "Optical repetition MIMO transmission with multipulse PPM," *IEEE Sel. Areas Comm.*, vol. 23, pp. 1901–1910, Sept. 2005.
- [16] W. Shieh and I. Djordjevic, *OFDM for Optical Communications*. Elsevier, 2009.
- [17] I. B. Djordjevic, W. Ryan, and B. Vasic, *Coding for Optical Channels*. Springer, 2010.
- [18] I. B. Djordjevic, L. L. Minkov, and H. G. Batshon, "Mitigation of linear and nonlinear impairments in high-speed optical networks by using LDPC-coded turbo equalization," *IEEE J. Sel. Areas Comm. Optical Comm. and Netw.*, vol. 26, pp. 73–83, 2008.
- [19] I. B. Djordjevic, S. Sankaranarayanan, S. K. Chilappagari, and B. Vasic, "Low-density parity-check codes for 40 Gb/s optical transmission systems," *IEEE/LEOS J. Sel. Top. Quantum Electron.*, vol. 12, no. 4, pp. 555–562, July/Aug. 2006.
- [20] I. B. Djordjevic, O. Milenkovic, and B. Vasic, "Generalized low-density parity-check codes for optical communication systems," *IEEE/OSA J. Lightwave Technol.*, vol. 23, pp. 1939–1946, May 2005.
- [21] B. Vasic, I. B. Djordjevic, and R. Kostuk, "Low-density parity check codes and iterative decoding for long haul optical communication systems," *IEEE/OSA J. Lightwave Technology*, vol. 21, pp. 438–446, Feb. 2003.
- [22] I. B. Djordjevic *et al.*, "Projective plane iteratively decodable block codes for WDM high-speed long-haul transmission systems," *IEEE/OSA J. Lightwave Technol.*, vol. 22, pp. 695–702, Mar. 2004.
- [23] O. Milenkovic, I. B. Djordjevic, and B. Vasic, "Block-circulant low-density parity-check codes for optical communication systems," *IEEE/LEOS J. Sel. Top. Quantum Electron.*, vol. 10, pp. 294–299, Mar./Apr. 2004.
- [24] B. Vasic and I. B. Djordjevic, "Low-density parity check codes for long haul optical communications systems," *IEEE Photon. Technol. Lett.*, vol. 14, pp. 1208–1210, Aug. 2002.
- [25] T. Mizuochi *et al.*, "Forward error correction based on block turbo code with 3-bit soft decision for 10 Gb/s optical communication systems," *IEEE J. Sel. Top. Quantum Electron.*, vol. 10, pp. 376–386, 2004.
- [26] T. Mizuochi *et al.*, "Next generation FEC for optical transmission systems," in *Proc. Opt. Fib. Comm. Conf. (OFC 2003)*, vol. 2, pp. 527–528, 2003.
- [27] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 533–547, Sept. 1981.
- [28] M. P. C. Fossorier, "Quasi-cyclic low-density parity-check codes from circulant permutation matrices," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1788–1793, 2004.
- [29] I. B. Djordjevic, L. Xu, T. Wang, and M. Cvijetic, "Large girth low-density parity-check codes for long-haul high-speed optical communications," in *Proc. OFC/NFOEC*, (IEEE/OSA, San Diego, CA, 2008), Paper no. JWA53.
- [30] H. Xiao-Yu, E. Eleftheriou, D.-M. Arnold, and A. Dholakia, "Efficient implementations of the sum-product algorithm for decoding of LDPC codes," in *Proc. IEEE Globecom*, vol. 2, Nov. 2001, pp. 1036–1036E.

- [31] R. G. Gallager, *Low Density Parity Check Codes*. MIT Press, 1963.
- [32] D. J. C. MacKay, "Good error correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, pp. 399–431, 1999.
- [33] K. Simon and V.A Vilnrotter, "Alamouti-type space-time coding for free-space optical communication with direct detection," *IEEE Trans. Wireless Comm.*, vol. 4, pp. 35–39, Jan. 2005.
- [34] I. B. Djordjevic, B. Vasic, and M.A. Neifeld, "Multilevel coding in free-space optical MIMO transmission with Q-ary PPM over the atmospheric turbulence channel," *IEEE Photon. Technol. Lett.*, vol. 18, no. 14, pp. 1491–1493, July 15, 2006.
- [35] N. Cvijetic, S.G. Wilson, and M. Brandt-Pearce, "Receiver optimization in turbulent free-space optical MIMO channels with APDs and Q-ary PPM," *IEEE Photon. Technol. Lett.*, vol. 19, no. 2, pp. 1491–1493, Jan. 15, 2007.
- [36] V. Tarokh, H. Jafarkani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.
- [37] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.
- [38] I. B. Djordjevic, S. Denic, J. Anguita, B. Vasic, and M. A. Neifeld, "LDPC-Coded MIMO Optical Communication over the Atmospheric Turbulence Channel," *IEEE/OSA J. Lightwave Technol.*, vol. 26, no. 5, pp. 478–487, March 1, 2008.
- [39] S. Denic, I. B. Djordjevic, J. Anguita, B. Vasic, and M. A. Neifeld, "Information theoretic limits for free-space optical channels with and without memory," *IEEE/OSA J. Lightwave Technol.*, vol. 26, no.19, pp. 3376–3384, October 1, 2008.
- [40] S. ten Brink, "Designing iterative decoding schemes with the extrinsic information transfer chart," *AEÜ Int. J. Electron. Comm.*, vol. 54, pp. 389–398, Dec. 2000.
- [41] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
- [42] I. Anderson, *Combinatorial Designs and Tournaments*. Oxford University Press, 1997.
- [43] I. B. Djordjevic, "LDPC-coded MIMO optical communication over the atmospheric turbulence channel using Q-ary pulse-position modulation," *Opt. Express*, vol. 15, no. 16, pp. 10026–10032, 2007.
- [44] I. B. Djordjevic, B. Vasic, and M. A. Neifeld, "LDPC coded orthogonal frequency division multiplexing over the atmospheric turbulence channel," in *Proc. CLEO/QELS 2006*, May 21–26, 2006, Long Beach, California, Paper no. CMDD5. (Invited Paper.)
- [45] I. B. Djordjevic, B. Vasic, and M. A. Neifeld, "LDPC coded OFDM over the atmospheric turbulence channel," *Opt. Express*, vol. 15, pp. 6332–6346, 2007.
- [46] I. B. Djordjevic, B. Vasic, and M. A. Neifeld, "LDPC-coded OFDM for optical communication systems with direct detection," *IEEE/LEOS J. Sel. Top. Quantum Electron.*, vol. 13, pp. 1446–1454, 2007.
- [47] R. Prasad, *OFDM for Wireless Communications Systems*. Artech House, 2004.
- [48] I. B. Djordjevic and H. G. Batshon, "LDPC-coded OFDM for heterogeneous access optical networks," *IEEE Photon. J.*, vol. 2, no. 4, pp. 611–619, Aug. 2010.
- [49] I. B. Djordjevic, "Coded-OFDM in hybrid optical networks," *IET Optoelectronics*, vol. 4, no. 1, pp. 17–28, Feb. 2010.
- [50] I. B. Djordjevic, "LDPC-coded optical communication over the atmospheric turbulence channel," in *Proc. 41st Asilomar Conference on Signals, Systems, and Computers*, Nov. 4–7, 2007, pp. 1903–1909, 2007. (Invited Paper.)

- [51] W. Shieh and C. Athaudage, "Coherent optical frequency division multiplexing," *Electron. Lett.*, vol. 42, pp. 587–589, 2006.
- [52] A. J. Lowery, L. Du, and J. Armstrong, "Orthogonal frequency division multiplexing for adaptive dispersion compensation in long haul WDM systems," in *Proc. OFC Postdeadline Papers*, Paper no. PDP39, Mar. 2006.
- [53] S. L. Jansen, I. Morita, N. Takeda, and H. Tanaka, "20-Gb/s OFDM transmission over 4,160-km SSMF enabled by RF-pilot tone phase compensation," in *Proc. OFC/ NFOEC 2007 Postdeadline Papers*, Paper no. PDP15, March 25–29, 2007, Anaheim, CA, USA.
- [54] I. B. Djordjevic, M. Arabaci, and L. L. Minkov, "Next generation FEC for high-capacity communication in optical transport networks," *IEEE/OSA J. Lightwave Technol.*, vol. 27, no. 16, pp. 3518–3530, August 15, 2009.
- [55] J. A. Anguita, I. B. Djordjevic, M. A. Neifeld, and B. V. Vasic, "Shannon capacities and error-correction codes for optical atmospheric turbulent channels," *OSA J. Opt. Networking*, vol. 4, pp. 586–601, 2005.
- [56] J. G. Proakis, *Digital Communications*. McGraw Hill, 2001.
- [57] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
- [58] O. Milenkovic, I. B. Djordjevic, and B. Vasic, "Block-circulant low-density parity-check codes for optical communication systems," *IEEE J. Sel. Top. Quantum Electron.*, vol. 10, pp. 294–299, 2004.
- [59] N. Levinson, "The Wiener RMS error criterion in filter design and prediction," *J. Math. Phys.*, vol. 25, pp. 261–278, 1947.
- [60] J. Durbin, "Efficient estimation of parameters in moving-average models," *Biometrika*, vol. 46, pp. 306–316, 1959.
- [61] A. T. A Wood and G. Chan, "Simulation of stationary Gaussian processes in $[0,1]^d$," *J. Comp. Graph. Stat.*, vol. 3, 409–432, 1994.
- [62] I. B. Djordjevic, "Coded-OFDM in hybrid optical networks," *IET Optoelectronics*, accepted for publication.
- [63] I. B. Djordjevic, L. Xu, and T. Wang, "Beyond 100 Gb/s optical transmission based on polarization multiplexed coded-OFDM with coherent detection," *IEEE J. Sel. Areas Comm., Optical Comm. and Netw.*, vol. 27, no. 3, Apr. 2009.
- [64] J. A. Anguita, M. A. Neifeld, B. Hildner, B. Vasic, "Rateless coding on experimental temporally correlated FSO channels," *IEEE/OSA J. Lightwave Technol.*, vol. 28, no. 7, pp. 990–1002, 2010.
- [65] M. Luby, "LT codes," *Proc. IEEE Symposium on Foundations of Computer Science (FOCS02)*, 43, 2002.
- [66] A. Shokrollahi, "Raptor codes," *IEEE Inf. Theory*, vol. 52, pp. 2551–2567, June 2006.
- [67] D. J. C. MacKay, "Fountain codes," *IEE Proc. Commun.*, vol. 152, pp. 1062–1068, Dec. 2005.
- [68] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
- [69] I. B. Djordjevic, "Adaptive modulation and coding for communication over the atmospheric turbulence channels," *IEEE Photonics Society Summer Topicals 2009*, 20–22 July 2009, Newport Beach, California, Paper no. TuD3. 3.
- [70] I. B. Djordjevic and G. T. Djordjevic, "On the communication over strong atmospheric turbulence channels by adaptive modulation and coding," *Optics Express*, vol. 17, no. 20, pp. 18250–18262, 28 September 2009.
- [71] I. B. Djordjevic, "Adaptive modulation and coding for free-space optical channels," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 2, no. 5, pp. 221–229, May 2010.

-
- [72] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. 28, pp. 55–67, Jan. 1982.
- [73] G. D. Forney, Jr., "Coset codes – Part I: introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1123–1151, Sept. 1988.
- [74] A. A. Farid, and S. Hranilovic, "Upper and lower bounds on the capacity of wireless optical intensity channels," in *Proc. ISIT 2007*, pp. 2416–2420, Nice, France, June 24–June 29, 2007.
- [75] H. G. Batshon and I. B. Djordjevic, "Beyond 240 Gb/s per wavelength optical transmission using coded hybrid subcarrier/amplitude/phase/polarization modulation," *IEEE Photon. Technol. Lett.*, vol. 22, pp. 299–301, March 1, 2010.
- [76] A. Stöhr, *et al.*, "60 GHz radio-over-fiber technologies for broadband wireless services," *OSA J. Opt. Netw.*, vol. 8, pp. 471–487, May 2009.
- [77] M. Ghavami, L. B. Michael, and R. Kohno, *Ultra Wideband Signals and Systems in Communication Engineering*. John Wiley & Sons, 2007.
- [78] C. F. Lam, *Passive Optical Networks: Principles and Practice*. Amsterdam: Elsevier/Academic Press, 2007.
- [79] X.-Y. Hu, E. Eleftheriou, and D. M. Arnold, "Regular and irregular progressive edge-growth Tanner graphs," *IEEE Trans. Inform. Theory*, vol. 51, no. 1, pp. 386–398, Jan. 2005.
- [80] I. B. Djordjevic and H. G. Batshon, "Generalized hybrid subcarrier/amplitude/phase/polarization LDPC-coded modulation based FSO networking," in *Proc. 12th International Conference on Transparent Optical Networks (ICTON 2010)*, June 27–July 1, 2010, Munich, Germany, Paper no. Th.B3.4.
- [81] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.
- [82] G. Yabre, "Theoretical investigation on the dispersion of graded-index polymer optical fibers," *J. Lightwave Technol.*, vol. 18, pp. 869–877, 2000.
- [83] B. A. Khawaja, I. Djordjevic and M. J. Cryan, "A millimeter wave phase shifter using a wireless hybrid mode locked laser," in *Proc. OFC/NFOEC 2009*, Paper No. OTuM5, San Diego, CA, March 22–26, 2009.
- [84] H. G. Batshon, I. B. Djordjevic, and T. Schmidt, "Ultra high speed optical transmission using subcarrier-multiplexed four-dimensional LDPC-coded modulation," *Opt. Express*, vol. 18, no. 19, pp. 20546–20551, 13 September 2010.

3 Wireless optical CDMA communication systems

Jawad A. Salehi, Babak M. Ghaffari, and Mehdi D. Matinfar

3.1 Introduction

Amidst revolutionary communication technologies of recent decades such as optical fiber and wireless systems other forms of communication technologies are popping up in order to complete or complement the ever-increasing and insatiable need for communications links in today's society. Among many such niche technologies optical wireless communication technology, in particular, is beginning to enjoy a wide range of applications and attention from many industries for short-range interchip applications to interplanetary space applications. Furthermore, optical wireless systems have attracted even further considerable growth in research and development since they can be appropriate alternatives for wireless or fiber-optic communication systems in some specific applications. Indoor wireless LANs, atmospheric optical links, and submarine optical wireless systems have grown in importance where lightwave communications is preferred to radio communications. This preference in various applications can be originated from security requirements, radio interference avoidance, no need for reserving frequency bands and cost of development [1]–[9].

In this chapter we study and analyze a particular and advanced form of optical wireless communication systems namely optical code-division multiple-access (OCDMA) in the context of wireless optical systems. As wireless optical communication systems get more mature and become viable for multi-user communication systems, advanced multiple-access techniques become more important and attractive in such systems. Among all multiple-access techniques in optical domain, OCDMA is of utmost interest because of its flexibility, ease of implementation, no need for synchronization among many users, and soft traffic handling capability.

Intensity modulation/direct detection (IM/DD) schemes are widely used in wireless optical applications. The interest in using IM/DD schemes stems from the fact that they can be easily implemented in transmitter and receiver sides. For instance, in optical wireless indoor systems design a portion of a wireless OCDMA transmitter and receiver based on IM/DD can be established easily using existing and off-the-shelf technology. On the other hand it is shown that in the optical scatter channels it is more efficient to employ incoherent optical modulations [1]. In these cases system structure can be

designed based on on–off keying (OOK) or pulse position modulations (PPM) family [2]. Optical orthogonal codes (OOC), that have been introduced for use in OCDMA systems are binary unipolar codes that have been shown to be optimum in OCDMA systems by means of limiting the multiple-access interference (MAI) noise [10]–[13]. By employing OOC in the OCDMA systems, each user transmits its corresponding data bits on a sparse sequence of optical pulses which constitute its dedicated code. Such a pulsing scheme allows users to communicate simultaneously on the shared channel without the requirement to be synchronized among each other. On the other hand in this case the maximum allowable peak power of optical pulses can be greater than ordinary systems due to multiple pulse OOC structure. This feature is effective in optical wireless systems where the peak power is limited due to the eye safety conditions.

In this chapter we study deployment of OCDMA communication systems in both indoor and outdoor (free-space optical link) environments. In applications where electromagnetic interference is to be avoided, optical networks are preferred to radio networks. The networks inside passenger planes and hospitals are categorized into such group applications. Intra-satellite networks for small scientific spacecrafts is another example that follows this rule too. Employing infrared communications as the intra-satellite network not only prevents electromagnetic interference but also reduces the volume and the weight of the spacecraft. This point is critical since there are several hundreds of wire bundles and connectors, and about one half are dedicated to data transmission [14]. Usually there are several transmit/receive nodes in these networks that need moderate bit rate. Employing a multiple-access technique in this case in order to support communications in the shared physical channel is necessary. OCDMA is an excellent choice since the implementation of the receiver is simple and all optical devices operate on the same wavelength and there is no need to synchronize various nodes.

Free-space optical (FSO) links are extremely important in many applications especially when installation of fiber between various nodes is expensive or is not feasible for various reasons. In fact an FSO link can be part of a bigger optical network that is using OCDMA as the multiple-access technique. We study and investigate the performance of an OCDMA FSO link. In our analysis we consider a realistic model that has been implemented recently. We use 1550 nm wavelength in order to integrate FSO link with high-speed fiber optic network. At this wavelength, EDFA is a good choice to use as the post- and (or) pre-amplifier. However in the case that the amplifier gain is high enough, employing high power post-amplifier (booster) can be more efficient than pre-amplifier since post-amplifier only amplifies the signal but not the noise as the pre-amplifier does.

In what follows we study OCDMA principles along with the structure of OOC and follow up with the introduction of two receiver structures for wireless OCDMA systems.

3.2 OCDMA system description

In OCDMA systems each user transmits its corresponding data bits by its corresponding code. The codes in optical systems may be constructed in different domains such as

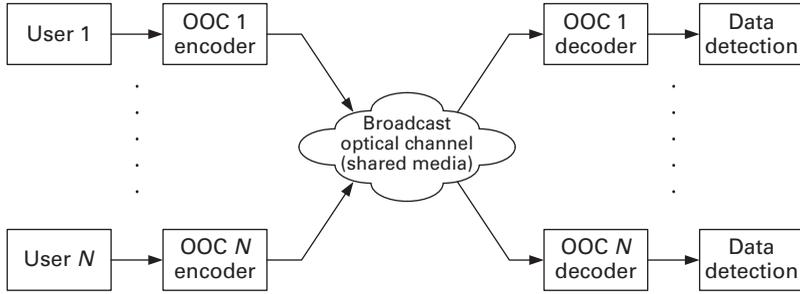


Figure 3.1 OCDMA general block diagram.

time, wavelength, polarization, space, or a combination of them [15]. Due to special conditions of wireless optical channels such as scintillation and phase fluctuations, employing temporal codes in wireless OCDMA system in the context of intensity modulation and direct detection is shown to be the most attractive [17]–[19]. Hence from now on we consider wireless OCDMA systems with their user codes designed in time domain. In such systems a data bit is mapped on a sequence of ‘0’ and ‘1’s in time domain which constitute the corresponding code. The codes are designed such that they are semi-orthogonal. In this case the correlation at the receiver side can distinguish data bits belonging to different users. We have shown the general block diagram of an OCDMA system with N users in Figure 3.1. For each user at the transmitter there is an encoder which generates the coded data stream. We assume optical orthogonal codes (OOC) as the signature sequence in the system. Data stream of users are combined, i.e., added, in the shared media. At the receiver side for each user there is a decoder that attempts to regenerate the data bit of the corresponding user among all users’ data bits.

3.2.1 OOC

Optical orthogonal codes (OOC) are a family of $(0,1)$ -sequences with good auto- and cross-correlation properties providing asynchronous multiple-access communications with easy synchronization and efficient performance in OCDMA communication networks.

An optical orthogonal code $(L, w, \lambda_a, \lambda_c)$ is a family C of $(0, 1)$ -sequences of length L with constant Hamming-weight w satisfying the following two properties:

1. Auto-correlation property: for any codeword $x = (x_0, x_1, \dots, x_{L-1}) \in C$, the inequality $\sum_{i=0}^{L-1} x_i x_{i \oplus \tau} \leq \lambda_a$ holds for any integer $\tau \not\equiv 0 \pmod{L}$
2. Cross-correlation property: for any two distinct codewords $x, y \in C$, the inequality $\sum_{i=0}^{L-1} x_i y_{i \oplus \tau} \leq \lambda_c$ holds for any integer τ ,

where the notation \oplus denotes the modulo- L addition [20]. The small value for λ_a and λ_c guarantees easy synchronization (of each transmitter–receiver pair) and limited multiple-access noise, respectively. When $\lambda_a = \lambda_c = \lambda$, we denote the OOC by (L, w, λ) for simplicity. The number of codewords in a family is called the size of the

optical orthogonal code. From a practical point of view, a code with large size is desired [11]. To find the best possible codes, we need to determine an upper bound on the size of an OOC family with the given parameters. Let $\Phi(L, w, \lambda_a, \lambda_c)$ be the largest possible size of an $(L, w, \lambda_a, \lambda_c)$ -OOC. It is easy to show that if $w(w-1) > \lambda_a(L-1)$ then $\Phi(L, w, \lambda_a, \lambda_c) = 0$ and if $w^2 > \lambda_c L$ then $\Phi(L, w, \lambda_a, \lambda_c) \leq 1$ [21]. However, based on the Johnson bound for constant-weight error-correcting codes [22], we have the following bound [20],

$$\Phi(L, w, \lambda) \leq \left\lfloor \frac{1}{w} \left\lfloor \frac{L-1}{w-1} \left\lfloor \frac{L-2}{w-2} \left[\dots \left\lfloor \frac{L-\lambda+1}{w-\lambda+1} \left\lfloor \frac{L-\lambda}{w-\lambda} \right\rfloor \dots \right\rfloor \right] \right\rfloor \right\rfloor \right\rfloor \quad (3.1)$$

where the notation $\lfloor \cdot \rfloor$ denotes the integer floor function. Also, it is clear from the definition that $\Phi(L, w, \lambda_a, \lambda_c) \leq \Phi(L, w, \lambda)$ where $\lambda = \max\{\lambda_a, \lambda_c\}$ [20].

From the mutual interference point of view, the most desirable on-off signature sequences are OOCs with $\lambda_c = 1$. However, when we look at the system with a small number of users interacting on each other's OOCs $\lambda_c = 1$ is not necessarily the optimum case. The main reason being that there is a trade-off between the code weight, w and λ_c . Larger λ_c allows more interference between users but on the other hand allows having larger code weight which by itself increases the coding gain hence decreasing the total interference. Briefly, the number of users in the network, receiver structure, and target bit error rate define the optimum values of OOC parameters [12].

In what follows we study two main structures for OCDMA receiver and investigate their characteristics in an optical channel.

3.2.2 Correlator structure

The simplest structure for an OCDMA receiver, namely, correlator includes a matched filter, corresponding to its code pattern, and an integrate and dump circuit. We have shown the structure of a OCDMA transmitter including encoder, and the correlator receiver in Figure 3.2. In this block diagram the decoder is matched to the corresponding code generated by the encoder. In fiber-optic CDMA systems the matched filter can be implemented by fiber tapped-delay lines at the receiver side. Delay lines in each branch serve to compensate for the corresponding delays induced at the encoder which constitute the corresponding OOC. The properly delayed replicas add to construct the output correlator which is then fed to the photodetector, integrate, dump, and threshold electronic circuit. In Figure 3.2 the white pulses at the input of the desired decoder are the received signal from undesired users. The major advantage of the correlator structure is that the limiting high-speed chip time decoding operation is done passively and optically. On the other hand in digital electronic implementation, delays can be generated by employing flip flops. If we assume the sampling rate per chip equal to n_s , $n_s L$ flip flops are sufficient to construct the matched filter of the corresponding OOC code [18]. If we denote the code length, the weight and the chip duration by L , w , and T_c respectively, and if the positions of the pulsed mark chips corresponding to the code sequence is c_1, c_2, \dots , and c_w , then the optical matched filter (passive optical tapped-delay line) consists of w fiber delay lines each of which make a delay equal to $(L - c_1)T_c$, $(L - c_2)T_c, \dots$, and

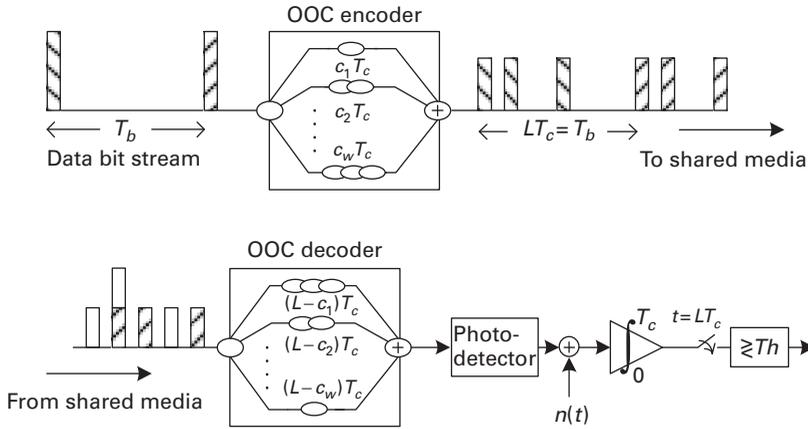


Figure 3.2 OCDMA transmitter, and correlator receiver.

$(L - c_w)T_c$ respectively. Considering only one bit duration, the output of the matched filter which is the combination of all w lines of the output signal has a peak at the time LT_c if the transmitted bit is on. However, when the number of interfering users is equal to or greater than w , we may observe such a peak in the absence of the desired user's signal, i.e., the transmitted bit is off, and this causes an error.

Although the correlator receiver seems trivial, in the sense that it is the optimum receiver in the case of single user communication with no interference, it can be much improved taking into account other interfering users using some other OOC codes on the same optical channel.

3.2.3 Chip-level AND logic gate structure

The correlator structure is not resistant enough against multi-access interference since it accumulates all the received energy from the marked chips independent of the number of interfering pulses on each specified marked chip. The first improvement to the correlator structure was suggested in [11]. The idea was further studied in depth in [23] and [24]. In this method in every definite chip time just that amount of intensity transmitted for a "1" bit may carry information and any excess intensity is due to interference. Thus an optical hard-limiter, which limits such excessive amount of intensity, if placed prior to the conventional correlator receiver will block some interfering patterns from causing errors and improve the system performance by at least two orders of magnitude. This implies that we first decide on the marked chips by placing the hard-limiter. In the second stage, the receiver bases its decision on the transmitted bit based on the detected marked chips. One possible scheme for the decision is that we decide on bit "1" if all the marked chips are detected as pulse present, and bit "0" is detected otherwise. Interestingly, this is the best method for bit detection in many cases. The reason is that the dominant noise in OCDMA systems is MAI which converts chips "0" to "1" but not the reverse, i.e., "1" to "0". Since all marked chips are needed to be detected as "1" or pulse present in

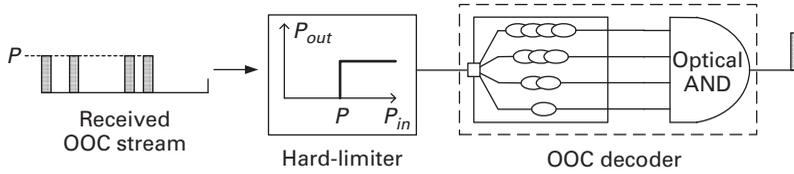


Figure 3.3 Chip-level AND logic gate receiver.

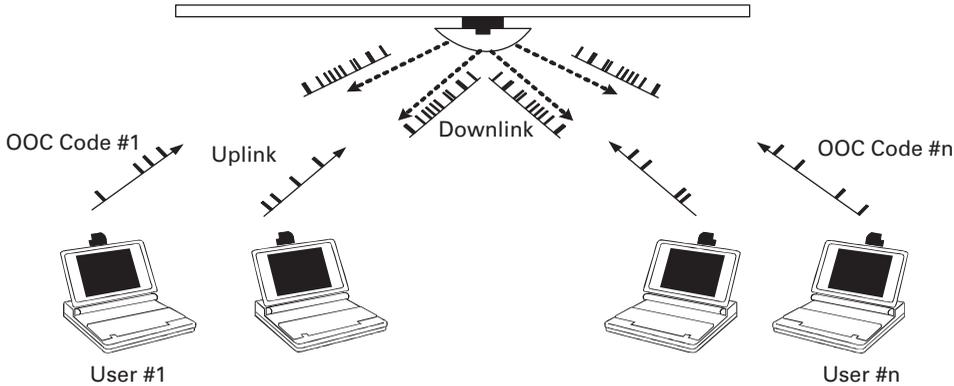


Figure 3.4 Typical indoor wireless optical CDMA LAN containing a base station and n users.

order to result to the detection of bit “1”, we call such receiver structure as chip-level AND logic gate and that can be implemented by AND gate logically. We have shown the receiver structure in Figure 3.3.

3.3 Indoor wireless optical CDMA LAN

In this section a typical wireless OCDMA Local Area Network (LAN) is investigated. A typical wireless OCDMA communication system with a small number of stations communicating to a single base station through the optical channel is considered. Usually it is difficult to implement an irregular and ad hoc network architecture for a wireless optical LAN due to the limitations of an optical wave’s propagation properties. Especially, one or more base stations are needed to act as a bridge when a local network needs to be connected to a global network such as the Internet. To perform the communication of multi-users in a single share channel there is a need for a multiple-access technique such as OCDMA.

Figure 3.4 shows a typical system architecture which includes one base station, with a single pair of optical transmitter and receiver, and N user stations, each having a pair of optical transmitter and receiver. With respect to the maximum number of permitted active users in the network and the desired code length, a set of OOCs is defined and each code is assigned to each active user. In this system each active user is assigned

a signature sequence from a set of OOC which is specified by $(L, w, \lambda_a, \lambda_c)$. All active users send their data bits asynchronously on the common channel. Therefore, the received signal is comprised of the analog summation of all users' optical intensity transmitted signals. Hence, the received signal can be written as follows,

$$Y(t) = \sum_{k=1}^N y^{(k)}(t - \tau_k) \quad (3.2)$$

where, $y^{(k)}(t)$ is the received signal due to the k th user and τ_k indicates its delay time with respect to the base station reference clock and can be written as follows,

$$y^{(k)}(t) = \sum_{i=-\infty}^{\infty} \sum_{j=1}^L u_i^{(k)} c_j^{(k)} P_{T_c}(t - iT_b - jT_c) \quad (3.3)$$

where $\{u_i^{(k)}\}_i$ is the bit stream of user k , $\{c_j^{(k)}\}_{j=1}^L$ denotes the k th OOC code pattern, and $P_{T_c}(t)$ is the pulse function with a duration equal to T_c .

In the design of any wireless CDMA communication network the near-far problem needs to be addressed [26]. The near-far problem in wireless OCDMA LAN requires specific algorithms to mitigate its effect such as the use of a proper automatic gain control (AGC) circuit. Hence, in this section it is assumed the existence of power control algorithms. Therefore we can safely ignore the near-far problem such that all users have equal power in the uplink channel. In what follows we discuss the transmitter and two viable and relevant receiver structures for a wireless OCDMA system. Further we discuss the pros and cons of such transmitter and receiver structures in the context of their digital design and their implementation. Finally their performance is obtained and discussed.

3.3.1 Wireless OCDMA transmitter

In an OCDMA network, using OOC, the network's active users transmit their corresponding data bits asynchronously. But at the receiver side prior to detecting data bits it is first essential to obtain bit-level synchronization between the corresponding transmitter–receiver pair. OOC coding takes place in the physical layer. So by adding some redundant bits, namely training sequence, to information data bits prior to coding, the receiver will be able to synchronize its code with its corresponding transmitted code using the knowledge of the training bits pattern.

3.3.2 Wireless OCDMA receiver

In the design of the receiver and its implementation more complexity is encountered when compared to the transmitter especially in the synchronization module. Similar to radio CDMA systems, synchronization in OCDMA receivers precedes the data detection block and it is applied in two stages namely acquisition and tracking. In a typical synchronization process prior to tracking state, the received data is first passed through the

acquisition stage. Following acquisition, the receiver is synchronized to the spreading code with an accuracy which is within one-half of chip duration.

The tracking circuit performs more precise timing using, as its input, the synchronized code from the acquisition block. Note that the acquisition circuit is active only at the startup phase or when the receiver exits the synchronization state, while the tracking loop remains online at all times. The correct phase code produced by the synchronization block is forwarded to the detection block. Prior to studying the synchronization algorithms it is more suitable to investigate the two viable and relevant receiver structures for the detection block of a wireless OCDMA system. In our analysis of the performance of the receiver detection scheme it is assumed that the synchronization has been established successfully.

Correlator

The correlator receiver is the most propounded structure for OCDMA systems. In this structure, a decision is made based on the sum or the accumulation of the received energy in all marked chips of the corresponding OOC code.

From an implementation point of view this simple receiver involves a matched filter, corresponding to its code pattern, and an integrator followed by a sampler. In optical fiber CDMA systems the matched filter can be implemented by fiber tapped-delay lines at their receivers. However, in a typical wireless OCDMA LAN system where the speed of operation is not very high and the distances are relatively short, correlation may take place after the photodetection at the receiver. In such a receiver the signal at the output of the receiver's photodetector is sampled by an analog-to-digital converter (ADC) at a rate equal to n_s times the OOC chip rate, i.e., $1/T_c$, as shown in Figure 3.5(a). The correlation of the received signal with the desired users' code can be evaluated by the summation of wn_s samples of the received signal in all the marked pulsed chips of one bit duration using a simple adder. A buffer after the ADC operation saves the results of addition at each sampling time instant and resets at the end of the bit time. At the end of

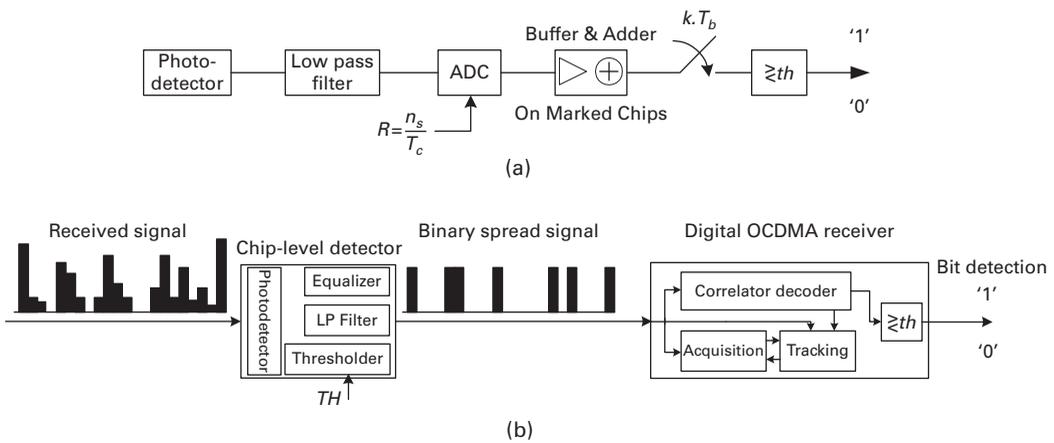


Figure 3.5 Digital structures for an optical CDMA receiver (a) correlator (b) chip-level AND logic gate.

the bit duration, the accumulated value that is saved in the buffer, i.e., correlation value, is compared with an optimum threshold.

Chip-level AND logic gate structure

In this structure, there are two levels or stages of decision. In the first stage, the receiver decides on the marked chips, one by one. In the second stage, a decision on the transmitted bit is made based on the first stage decisions. It is shown that multi-user interference effects can be reduced using this method which can be implemented simply by an optical hard-limiter and optical AND logic gate in a fiber-optic CDMA receiver using unipolar codes such as OOCs, [12, 19]. Using a simple hard-limiter for chip-level decision not only mitigates the undesirable effect of MAI but also reduces the sensitivity of system to power control. In a typical wireless OCDMA LAN receiver, the hard-limiter on the marked chips can be placed after the photodetector instead of being placed before the photodetector. A simple hard-limiting takes place using an analog circuit as shown in Figure 3.5(b). Thus, each sample from wn_s samples has a binary value at the sampler output. The remaining operations are similar to the correlator receiver. In this structure, due to hard-limiting, the effect of MAI is substantially reduced. From a digital design and implementation point of view, it simplifies the implementation since the sampling operation can be done along with other processes for example in a field programmable gate array (FPGA) without using a separate ADC circuitry.

BER analysis

The BER analysis is studied separately for the two above-mentioned receiver structures and BER expressions for an OCDMA LAN using an OOC family with $\lambda_c = 1$ as the signature sequence is studied. The wireless channel is considered to be LOS and is modeled with the DC loss imposed on the transmitted power signal (we assume that the communication rate is chosen properly so as not to cause pulse dispersion due to multipath effect). The maximum number of codes in an OOC set with $\lambda_a = \lambda_c = 1$ which limits the network capacity is obtained from (3.1).

In the digital receiver structure which is considered in this section, a lowpass filter followed by an ADC is placed after the receiver photodetector, as described in the previous section. Assuming the lowpass filter as a square shape response equal to $P_{T_f}(t)$, $T_f = T_c/n_s$, the digital receiver acts as an integrate and dump circuit. Although, for chip-synchronous multi-user interference pattern, $n_s = 1$ is sufficient for a proper detection, for implementation reasons and for tracking circuit more samples per chip are required as will be discussed in the tracking section. Hereafter, for BER analysis the value for n_s is assumed to be equal to one, i.e., $n_s = 1$.

The vector $\vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_w)$ is considered as the interference pattern occurred on the marked chips. In other word α_j is the number of interferences on the j th marked chip.

a. Correlator Structure

In a correlator structure, the signal generated by an integrate and dump circuit has a Poisson distribution function with a mean equal to m_d ,

$$m_d(l) = [(wd + l)\gamma_s + w\gamma_b]T_c \quad (3.4)$$

where $d \in \{0, 1\}$ indicates transmitted data bit, $l = \sum_{j=1}^w \alpha_j$ denotes the total number of interferences that occurred on the pulsed mark chips of the desired OOC code, $\gamma_s \triangleq (\frac{2L}{w})P_{av} \frac{\eta}{h\nu}$ and $\gamma_b \triangleq P_b \frac{\eta}{h\nu} + \frac{i_{dc}}{q}$. P_{av} is the average received power on the photodetector area, P_b is ambient noise power, i_{dc} and η are the photodetector dark current and quantum efficiency, respectively. Also q , ν , λ , h indicate electron charge, optical frequency, optical wavelength, and Planck constant, respectively.

The probability of error for the correlator structure conditioned on l is obtained as follows [27],

$$P(E|l) = \frac{1}{2} \sum_{n=0}^{th} \frac{e^{-m_1(l)} (m_1(l))^n}{n!} + \frac{1}{2} \sum_{n=th}^{\infty} \frac{e^{-m_0(l)} (m_0(l))^n}{n!} \quad (3.5)$$

since the received energy is considered to have Poisson distribution. In the above equation th is an optimum threshold that minimizes the total probability of error. The total probability of error, P_E is computed by averaging over l as follows,

$$P_E = \sum_{l=0}^{N-1} p_l(l) P(E|l) \quad (3.6)$$

where l has a Binomial distribution [27].

$$p_l(l) = \binom{N-1}{l} \left(\frac{w^2}{2F}\right)^l \left(1 - \frac{w^2}{2F}\right)^{N-1-l} \quad (3.7)$$

b. Chip-level AND logic gate structure

For chip-level AND logic gate receiver, probability of error depends on interference pattern in addition to their accumulation. We can write the conditional error probability as follows,

$$P(E|l, \vec{\alpha}) = 1/2 \left\{ \sum_{|\vec{n}|=w} \prod_{j=1}^w p_0(\alpha_j)^{n_j} [1 - p_0(\alpha_j)]^{(1-n_j)} + 1 - \sum_{|\vec{n}|=w} \prod_{j=1}^w p_1(\alpha_j)^{n_j} [1 - p_1(\alpha_j)]^{(1-n_j)} \right\} \quad (3.8)$$

where \vec{n} is the binary vector (detected after hard-limiter) caused on w marked chips, $|\vec{n}|$ is the Hamming weight of \vec{n} and we have,

$$m_d(\alpha_j) = [(d + \alpha_j)\gamma_s + \gamma_b] T_c, \quad d \in \{0, 1\}$$

$m_d(\alpha_j)$ denotes the average number of incident photons on the j th pulsed mark chip assuming data bit $d \in \{0, 1\}$, and we have defined:

$$p_0(\alpha_j) = \sum_{n=TH}^{\infty} \frac{e^{-m_0(\alpha_j)} (m_0(\alpha_j))^n}{n!}, \quad p_1(\alpha_j) = \sum_{n=0}^{TH} \frac{e^{-m_1(\alpha_j)} (m_1(\alpha_j))^n}{n!}$$

where TH is the threshold of hard-limiter and $\vec{\alpha}$ indicates the interference pattern caused on the corresponding OOC's marked chips.

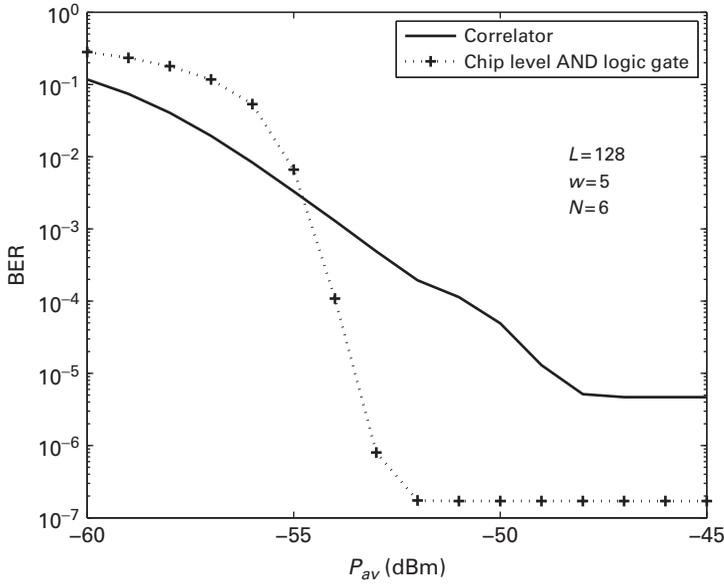


Figure 3.6 Performance curves due to various types of receivers vs. received optical power

The probability of error for chip-level AND logic gate structure is computed as follows,

$$P_E = \sum_{l=0}^{N-1} \sum_{\sum_{j=1}^w \alpha_j = l} P(l, \vec{\alpha}) P(E|l, \vec{\alpha}) \quad (3.9)$$

where $P(l, \vec{\alpha})$, the probability to have l active interfering users with the interference pattern $\vec{\alpha}$, is obtained as follows [27],

$$P(l, \vec{\alpha}) = \frac{(N-1)!}{w^l (N-1-l)! \prod_{j=1}^w (\alpha_j)!} \left(\frac{w^2}{2L} \right)^l \left(1 - \frac{w^2}{2L} \right)^{N-1-l}. \quad (3.10)$$

Using the typical values $L = 128$, $w = 5$, $N = 6$ as the OOC code parameters and $\lambda = 870$ nm, $I_b = 100 \mu\text{W}/\text{cm}^2$, $i_d = 10$ nA respectively as the light wavelength, background light irradiance, and photodetector dark current and $\eta = 0.8$, $A_d = 1$ cm² as the photodetector quantum efficiency and area respectively, the performance of the two above receiver structures is sketched in Figure 3.6.

Acquisition algorithms

In order to extract data bits properly from a spread stream of pulses, which constitutes the desired OOC receiver, one needs to know the correct initial reference time or the phase of the specified code pattern that carries the desired information bit.

Assuming a chip-synchronous scenario, the receiver encounters an ambiguity equal to L cells, each corresponding to one possible cyclic shift of the indicated OOC code. At the end of the acquisition level the correct cell, or equivalently the correct shift is found. This

can be obtained for example by using training bits which are inserted at the beginning of each frame. Some algorithms are suggested to be used in acquisition state. For each algorithm, a different number of training bits is required to meet acquisition with a probability near 1. It is evident that the less training bits required the more efficient the algorithm in use will be. In the following, three recently suggested algorithms for OOC based OCDMA are presented.

a. Simple serial-search method

In this approach, one cell is randomly selected and it is assumed to be the correct cell. A correlation between the received data and the selected code over a bit time duration (dwell time) is obtained and the outcome is compared with an optimum threshold. If the output of the correlation is greater than the optimum threshold, then the first randomly selected cell is the correct cell [28]. Otherwise, the next cell is examined. By next cell it is implied a chip-sized rotation replica of the initial cell or phase of the specified code. So by continuing this algorithm and in maximum L stages the correct cell is obtained in the ideal case. This method is relatively slow but its implementation in ultra-fast fiber optic systems is viable.

b. Multiple-shift method

To reduce the average time of acquisition, the multiple-shift method was proposed recently [29]. In this method L cells are divided into Q groups each containing M cells. In the first stage of the algorithm, all M cells of group 1 are examined simultaneously. If the output of the correlation is greater than a predetermined optimum threshold then it will proceed to the second stage, otherwise it will proceed to examine cells which belong to group 2. In continuing this process and from [29] in maximum Q steps the true group is obtained. In the second stage the M possible cells will be examined one by one similar to the simple serial-search method. So in the ideal case the corrected cell can be found in maximum $Q + M$ stages. So acquisition time is reduced considerably when compared to the simple serial search method from maximum L stages to $2\sqrt{L}$ stages.

c. Matched filter method

Assume a matched filter containing w delay lines. Each, delay line value is proportional to the position of pulsed chips corresponding to its respective OOC. When training bits enter the matched filter then the output signal is a periodic training pulse with a width equal to T_c and a period equal to T_b . This implies that in one bit duration the peak correlation of the incident signal and code pattern can be observed without any knowledge of the correct initial time reference. As an example the digital implementation of a matched filter is depicted for an OOC stream with $w = 3$ in Figure 3.7. Considering this idea, matched filtering automatically searches all the L cells within one bit duration instead of L bits duration as in the simple serial-search method. This method is extremely fast and can be easily implemented in wireless OCDMA systems that have been implemented in digital electronics.

Tracking circuit

The tracking circuit performs two key operations in a typical OCDMA receiver. In general, the received code from acquisition block has an ambiguity in phase or initial time reference which is less than $T_c/2$. The first task of the tracking circuit is to minimize this

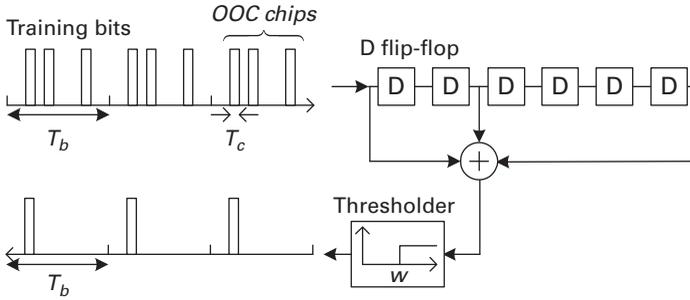


Figure 3.7 An example of a digital matched filter for an OOC stream with $w = 3$.

ambiguity. The second task for the tracking circuit is to alarm the *out of synchronization state* to the acquisition block. Early-late method is typically used for implementation of the tracking circuit [30]. In this method two separate early and late replica of the reference code, which is produced by the acquisition block, is generated by the tracking block. If $c(t)$ denotes the reference code, then $c(t + T_c/2)$ and $c(t - T_c/2)$ are the corresponding early and late codes, respectively. Now, correlation action takes place between the training bit stream and the early and late codes. If the output of both early and late correlators are less than a threshold determined by the detection block that is computed such that (3.6) is minimized for L sequential bits, then *out of synchronization state* is activated. This signal is passed to the acquisition block to switch on the acquisition state. To track the original code, the values of early and late correlator outputs are subtracted, and three actions may be initiated based on the subtraction result. Selecting a proper threshold, say th_c , then a shifted replica to the early correlator is selected as the new reference code if the subtraction result is greater than th_c . A shifted replica to the late correlator is selected if the subtraction result is less than $-th_c$, otherwise the main code is selected. To select the proper threshold value, th_c , two main issues need to be considered: (1) th_c value must not be so low to cause the tracking circuit to oscillate and (2) not so high in order for the tracking circuit not to be able to follow the phase shifts fluctuations. Figure 3.8 shows the impact of this parameter in the performance of the correlator receiver structure [18]. By increasing the number of samples per chip, the performance improves at the cost of increasing the complexity.

Power budget

In what follows we discuss the methodologies used to obtain the minimum power level of an optical transmitter that guarantees proper operation of a typical short range wireless OCDMA receiver in the presence of multi-user interference signal. Power budget is obtained under two conditions. First, we consider IEC's class A standard to meet eye safety conditions as proposition [31]. Second, in the worst case misalignment for Tx-Rx from the ideal line of sight in the presence of all interfering users while the system must operate at a BER less than a predefined value for instance 10^{-9} . Received optical power can be considered as follows [5],

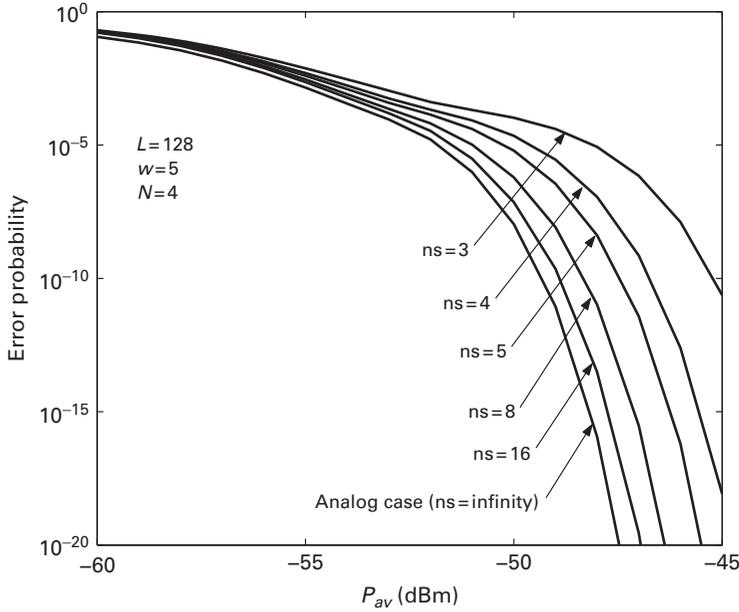


Figure 3.8 Performance curves for different number of samples taken at the receiver for detection process.

$$P_r = \left(\frac{P_t R(\theta, \phi)}{d^2} \right) A_{eff} \quad (3.11)$$

where P_t and d are transmitted optical power and Tx-Rx distance, respectively; A_{eff} indicates lens effective area and $R(\theta, \phi)$ is the propagation pattern function of the optical transmitter; θ and ϕ are the elevation and azimuth angles, respectively. The propagation function of a Lambertian source with order m is defined as [5],

$$R(\theta, \phi) = \frac{m+1}{2\pi} \cos^m \theta. \quad (3.12)$$

If the propagation half-angle of an optical transmitter is equal to $\theta_{1/2}$ then m is obtained as follows,

$$m = \frac{\ln(1/2)}{\ln(\cos \theta_{1/2})}. \quad (3.13)$$

As an example for a practical indoor environment and for a chip-level AND logic gate receiver structure to obtain a BER less than 10^{-7} , the received optical power P_r needs to be at least -52 dBm referring to Figure 3.6 results. If the distance of transmitter and receiver is assumed to be equal to 3 meters and the divergence angle of optical transmitter $\theta = 30^\circ$, then for the worst-case scenario, transmitter–receiver distance and misalignment are obtained to be $d = 3$ m and $\theta = 30^\circ$ respectively. For such parameters it can be deduced that $\theta_{1/2} = 15^\circ$ and using (3.13) Lambertian order m is equal to 20. Using (3.12) $R(\theta, \phi) = 0.188$ for the worst case. From (3.11) the transmitted average power per bit is obtained as $P_t = 3.01$ mW. Using the results of [31], it can be observed

that the maximum allowable transmission power for a typical optical transmitter operating at 870 nm and a 15° propagation half angle is equal to 28 mW. The optical power for each OOC chip pulse in an infrared CDMA transmitter is obtained as follows,

$$P_{t/OOC \text{ chip pulse}} = \frac{2L}{w} P_t. \quad (3.14)$$

For typical values of an OOC set with $L = 128$, $w = 5$, $N = 6$, the peak power of optical pulses can be as high as 1.433 W. This shows that in wireless OCDMA systems the peak power can become higher than ordinary OOK systems for around one to two order(s) of magnitude. One can improve the system performance especially for chip-level AND logic gate structure by increasing the SNR.

3.4 Free-space optical CDMA systems

In optical CDMA systems, an asynchronous communication between users of the network, usually in an all-optical structure, is established. The coded optical signal may be carried out by various waveguides such as optical fibers or a free-space optical (FSO) link. The network structure may consist of fiber and free-space links simultaneously. The structure of the free-space link(s) can be a point-to-point or a mesh or ring network [7, 32]. However in the performance analysis of physical layer it is fairly sufficient to consider a single point-to-point link.

We have shown the structure of a typical FSO-CDMA link in Figure 3.9. The distance between two nodes for a typical FSO link can be in order of a few hundred meters up to a few kilometers. To support such ranges it is required to use relatively high-power lasers or optical amplifiers. Employing erbium doped fiber amplifier (EDFA) as the post-amplifier at the transmitter side can be a very attractive option in this application. One reason for this attractiveness is that there are various kinds of EDFA with different desired gains in the market. In this case by employing just one high-power EDFA at the transmitter side, it can support the link budget for ranges that can extend to a few kilometers. On the other hand using 1550 nm wavelength can easily integrate the FSO link with the other part of the network that probably uses optical fibers and other available devices working in the same wavelength band [33, 34].

At the receiver side a telescope with high enough aperture size is employed to mitigate the channel losses due to the transmitted beam divergence. To analyze the performance of a FSO-CDMA link, the channel characteristics should be considered as well as the inherent characteristics of a typical OCDMA system. In this part we repeat our study on two different structures namely correlator and chip-level AND logic gate structure in the context of FSO-CDMA systems.

It is usually necessary to employ adaptive power control (APC) for FSO-CDMA systems. Employing APC in such systems guarantees that the power budget of the link is in the desired range and on the other hand in clear weather EDFA transmits at a lower power level. Furthermore in this case the performance of the receiver is less sensitive to the threshold set for detecting optical chips (in chip-level structure) or bits (correlator

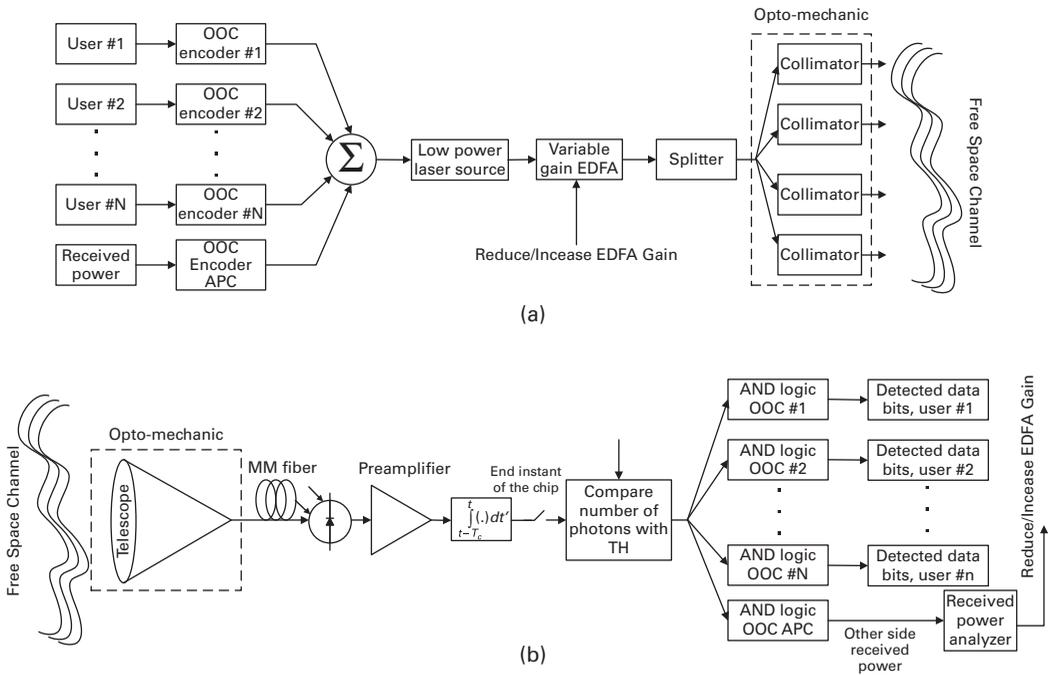


Figure 3.9 Typical FSO-CDMA system with N users: (a) transmitter part, (b) receiver part.

structure). To implement APC one assigns a dedicated OOC code for transmitting low rate sampled and digitized received power in each node to the other node in the FSO link. In node 1, the information of optical power is received by node 2 using the dedicated OOC assigned to APC. By using this information it is easy to tune the gain of EDFA belonging to node 1 in order to transmit at a power level such that the received power in node 2 is in the desired range. This scheme is performed for both directions. The power control algorithm for evaluating the optimum EDFA gain in each node is an iterative algorithm that reduces or increases the EDFA gain according to the received power information corresponding to the other node in a step-by-step procedure. In fact a desired range for the received power is defined. If the received power is higher than the range, the EDFA gain reduces by a predefined constant value and if received power is lower than the desired range, the EDFA gain increases by the same predefined value. As a practical example, assume that a variable gain EDFA with maximum 30 dBm output power is employed. For some typical commercial EDFA the third stage variable gain lets the output power to range between 15 dBm and 30 dBm according to the external command that sends to the amplifier via its communication port. By using the aforementioned EDFA, it is possible to consider a desired range of 3 dB and to change the EDFA output power by 1 dB in each step.

The transmitter and receiver structures are similar to the one used in indoor wireless OCDMA system which was discussed in the earlier section. In the following subsection,

performances of these receiver structures in the context of FSO-CDMA system are studied. Analytical solutions to the error probability for various channel models using PIN photodetector are obtained. In the following analysis, the effects of atmospheric turbulence, ambient light, thermal noise, and multi-user interference are considered in the context of the semi-classical photon counting approach. For the sake of mathematical simplicity, the Gaussian approximation is used for performance evaluation.

3.4.1 BER analysis

The following analysis is based on semi classical photon counting where the photon count statistics is obtained separately in each stage. As it is shown in Figure 3.9, we place an optical amplifier in the FSO system which is not necessarily present in indoor wireless optical CDMA systems. High-power EDFA amplifier along with a low-power laser source, e.g., with $P_{t,1} = 0dBm$ output power, act as a high-power laser source at the transmitter side. Similar to indoor wireless OCDMA system OOK signalling is employed.

Low-power laser source

The photon count in one chip duration due to low power laser is indicated by n_1 which is assumed to have a Poisson distribution. The mean photon counts of low-power laser pulses representing zero and one are respectively, $m_{t,0,c} = 0$ and $m_{t,1,c} = P_{t,1}T_c/h\nu$. Where $P_{t,1}$, T_c , h , and ν are the peaked output power of the low-power laser source, chip duration, Planck's constant and optical frequency, respectively. Therefore, the moment generating function (MGF) corresponding to the photon count of low-power laser output with Poisson distribution is as follows,

$$\phi_{t,d,c}(z) = \exp[1 - m_{t,d,c}(1 - z)]. \quad (3.15)$$

Optical amplifier

The dominant noise source in optical amplifiers is ASE (Amplified Spontaneous Emission). By assuming G , $K = n_{sp}(G - 1)$, and M as the gain, the noise factor, and the number of spontaneous modes, respectively, the optical amplifier output photon count, denoted by n_2 , is obtained as a Laguerre distribution with parameters $(m_{t,d,c}G, K, M)$ [35]. Hence, the MGF corresponding to the photon count of optical amplifier output is obtained as follows,

$$\phi_{EDFA,t,d,c}(z) = \frac{1}{[1 + K(1 - z)]^M} \times \phi_{t,d,c}\left(1 - \frac{G(1 - z)}{1 + K(1 - z)}\right). \quad (3.16)$$

FSO channel model

Atmospheric channel can be simply modeled as an attenuator with random attenuation coefficient. Channel attenuation can be expressed as the multiplication of two parameters: one a deterministic parameter and another a random variable. FSO channel attenuation is caused by both molecular absorption and aerosol scattering which

randomly affects the light intensity traversing through the channel. The deterministic part can be written as follows [1],

$$a = \frac{A}{\pi \left(\frac{\theta L_0}{2}\right)^2} e^{-\beta L_0}. \quad (3.17)$$

In the above equation L_0 and β are the link distance and atmospheric extinction coefficient, A and θ are the telescope area and transmitted beam divergence in radians, respectively. The random attenuation of the channel is induced by the random variations of the channel refractive index which can be modeled as a *log normal* distribution e^{2x} in which x is a normal random variable. The random term is assumed to be normalized such that $E[e^{2x}] = 1$. So, the mean and the variance of the normal random variable x are considered as $-\sigma_x^2$ and σ_x^2 , respectively. Hence, if we assume that the channel attenuation is deterministic, i.e., condition on known x , the MGF of the photon count due to the channel output, denoted by n_3 , can be written as follows,

$$\phi_{s,d,c}(z|x) = \phi_{EDFA,t,d,c} \left[1 - ae^{2x}(1-z) \right]. \quad (3.18)$$

Furthermore the ambient light can also be modeled as an additive independent photon count with Poisson distribution. So, the MGF due to the total photon count coupled into the multi-mode fiber is obtained as follows,

$$\phi_{r,d,c|x}(z) = \phi_{s,d,c}(z|x) \phi_{a,c}(z) \quad : \quad \phi_{a,c}(z) = \exp \left[1 - m_{a,c}(1-z) \right] \quad (3.19)$$

where $m_{a,c}$ in the above equation indicates the mean photon count of the ambient light in one chip duration and it is obtained as follows,

$$m_{a,c} = \frac{W(\lambda) A \Omega_{FOV} \Delta \lambda T_c}{h\nu} \quad (3.20)$$

where Ω_{FOV} is the receiver field of view (FOV) in steradian, $\Delta \lambda$ is the optical filter bandwidth, and $W(\lambda)$ is the spectral radiance function, defined as the power radiated at wavelength per cycle of bandwidth in a unit solid angle per unit of source area.

Photodetector model

The photodetector is modeled as an attenuator with attenuation coefficient equal to its quantum efficiency indicated by η . So, the MGF corresponding to the photo electron count at the photodetector output is obtained as follows,

$$\phi_{PD,d,c}(z|x) = \phi_{r,d,c} [1 - \eta(1-z)|x]. \quad (3.21)$$

As it was mentioned earlier, the Gaussian approximation is employed for the BER analysis. The BER analysis is studied hereafter for two different receiver structures, namely, simple correlator and chip-level AND logic gate.

3.4.2 Correlator receiver structure

Usually, the coherence time of fading in FSO links is on the order of several milliseconds [1], while usual bit durations are less than several microseconds; therefore, we can

assume that the fading is constant during a bit period. Considering the structure of the correlator receiver, independence of counts for different chip intervals, and the fact that the fading is approximately constant during a bit period, mean and the variance of the decision variable for transmitting bit “1” and “0” can be written as follows,

$$m_d = \sum_{i=1}^w m_{r,d,c}^{(i)} \quad , \quad \sigma_d^2 = \sum_{i=1}^w \sigma_{i,r,d,c}^2 \quad ; \quad d = 0, 1 \quad (3.22)$$

where $\left\{ m_{r,d,c}^{(i)} \right\}_{i=1}^w$ and $\left\{ \sigma_{i,r,d,c}^2 \right\}_{i=1}^w$ are the means and variances corresponding to w marked chips. Assuming l active interfering users, in the desired user’s marked chips, the decision variable mean $m_{d,l}$ and variance $\sigma_{d,l}^2$ are obtained as follows,

$$\begin{aligned} m_{d,l} &= (d \times w + l) \times m_{PD,d,c} + w \times m_{dc,c} \\ \sigma_{d,l}^2 &= (d \times w + l) \times \sigma_{PD,d,c}^2 + w \times m_{dc,c} + \sigma_{th}^2 \quad d = 0, 1 \end{aligned} \quad (3.23)$$

where, $m_{PD,d,c}$, $\sigma_{PD,d,c}^2$ are the mean and variance of the photoelectrons in one chip duration; $m_{dc,c}$, σ_{th}^2 are the variance of the photodetector dark current per chip and thermal noise photoelectron count, respectively [36],

$$m_{dc,c} = \frac{i_d T_c}{q} \quad , \quad \sigma_{th}^2 = \frac{2K_b T_r T_c}{R_L q^2} \quad (3.24)$$

where, K_b , T_r , and R_L are Boltzman constant, receiver equivalent temperature, and the resistance of the receiver load. Also, i_d , T_c , and q are the photodetector dark current, chip time, and the electron charge, respectively. Using the aforementioned mathematical model in each stage, the mean and variance of the photoelectron count in one chip duration is obtained as follows,

$$\begin{aligned} \overline{m}_{PD,d,c} &= \eta \overline{m}_{a,c} + (M \times K + m_{t,d,c} \times G) \times \eta a e^{2x} \\ \sigma_{PD,d,c}^2 &= \eta \overline{m}_{a,c} + M \times K \times \eta a e^{2x} (1 + \eta a e^{2x} \times K) \\ &\quad + m_{t,d,c} \times G \times \eta a e^{2x} (1 + 2\eta a e^{2x} \times K). \end{aligned} \quad (3.25)$$

Finally, using Gaussian approximation, the BER of the simple correlator receiver structure is obtained as follows,

$$\begin{aligned} PE(x) &= \\ &= \frac{1}{2} \sum_{l=0}^{N-1} \binom{N-1}{l} \left(\frac{w^2}{2L} \right)^l \left(1 - \frac{w^2}{2L} \right)^{N-1-l} \left[Q \left(\frac{m_{1,l} - th}{\sigma_{1,l}} \right) + Q \left(\frac{th - m_{0,l}}{\sigma_{0,l}} \right) \right], \end{aligned} \quad (3.26)$$

where th in the above equation is the optimum threshold for the correlator that minimizes the total error probability; $Q(u)$ is the normalized Gaussian function defined as $Q(u) = (1/\sqrt{2\pi}) \int_u^\infty e^{(-v^2/2)} dv$. Note that the above expression evaluates the conditional error probability given the fading coefficient x . Hence the total error probability is obtained by averaging $PE(x)$ over the fading coefficient x as follows,

$$PE = \int_{-\infty}^{\infty} PE(x) f_X(x) dx \quad (3.27)$$

where $f_X(x)$ is the pdf of the fading coefficient defined as follows,

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x+\sigma_x^2)^2}{2\sigma_x^2}}.$$

3.4.3 Chip level AND logic gate

The conditional error probability given the turbulence coefficient x can be evaluated similar to the approach used in indoor wireless OCDMA section. The conditional error probability given the interference number, the interference pattern, and the fading coefficient is obtained as follows,

$$P(E|l, \vec{\alpha}, x) = \frac{1}{2} \left[\sum_{|\vec{n}|=w} \prod_{j=1}^w p_0(\alpha_j, x)^{n_j} (1 - p_0(\alpha_j, x))^{(1-n_j)} + 1 - \sum_{|\vec{n}|=w} \prod_{j=1}^w p_1(\alpha_j, x)^{n_j} (1 - p_1(\alpha_j, x))^{(1-n_j)} \right] \quad (3.28)$$

where \vec{n} is the binary vector (detected after hard-limiter) caused on w marked chips, $|\vec{n}|$ is the hamming weight of \vec{n} , $p_0(\alpha_j, x)$ and $p_1(\alpha_j, x)$ are the chip level error probabilities on the j th chip for transmitting bit '0' and bit '1', while $\vec{\alpha}$ and x are the known interference pattern and the fading coefficient on that chip duration, respectively. Assuming the photoelectron counts in the j th chip to be Gaussian, the error probabilities $p_0(\alpha_j, x)$ and $p_1(\alpha_j, x)$ are obtained as follows,

$$p_0(\alpha_j, x) = Q\left(\frac{TH - m_{0,j}}{\sigma_{0,j}}\right), \quad p_1(\alpha_j, x) = Q\left(\frac{m_{1,j} - TH}{\sigma_{1,j}}\right) \quad (3.29)$$

with

$$\begin{aligned} m_{d,j} &= (d + \alpha_j) \times m_{PD,d,c} + m_{a,c} \\ \sigma_{d,j}^2 &= (d + \alpha_j) \times \sigma_{PD,d,c}^2 + m_{a,c} + \sigma_{ih}^2, \quad d = 0, 1 \end{aligned}$$

where TH is the optimum threshold for chip level detection and is obtained to minimize the total error probability. The total error probability is obtained by averaging the error probability $P(E|l, \alpha_j, x)$ over the number of interfering users l , the interference pattern $\{\alpha_j\}_{j=1}^w$ and the turbulence coefficient x as follows,

$$PE = \int_{-\infty}^{\infty} \left[\sum_{l=0}^{N-1} \sum_{\sum_{j=1}^w \alpha_j = l} P(l, \vec{\alpha}) P(E|l, \vec{\alpha}, x) \right] f_X(x) dx \quad (3.30)$$

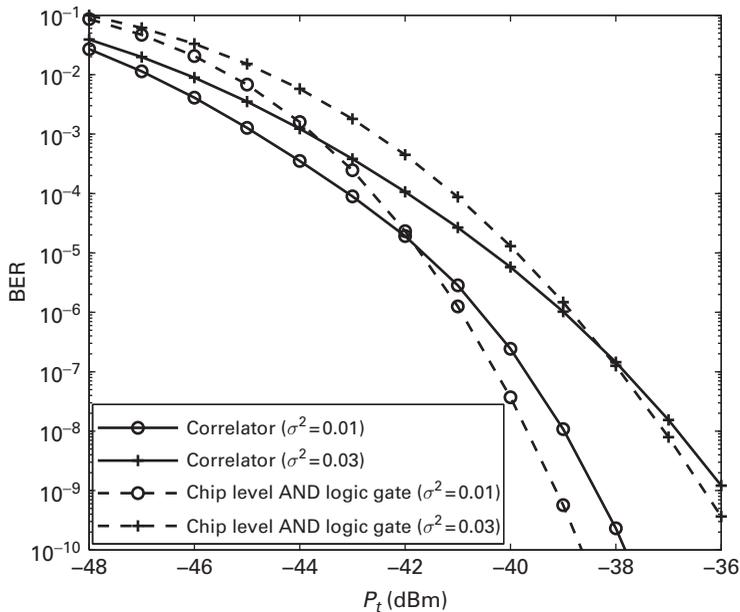
where $P(l, \vec{\alpha})$, the probability to have l active interfering users with the interference pattern $\vec{\alpha}$, is obtained as follows,

$$P(l, \vec{\alpha}) = \frac{(N-1)!}{w^l (N-1-l)! \prod_{j=1}^w (\alpha_j)!} \left(\frac{w^2}{2L}\right)^l \left(1 - \frac{w^2}{2L}\right)^{N-1-l}. \quad (3.31)$$

As an example, consider FSO link with parameters listed in Table 3.1. Performances of the aforementioned receiver structures are evaluated using the above represented expressions for the error probabilities. The result for $\sigma_x^2 = 0.01$ and $\sigma_x^2 = 0.03$ as the fading variance is shown in Figure 3.10. In a communication system the variance

Table 3.1 Typical values for the FSO OCDMA link.

| | | |
|-----------------|---------------------------------------|---------------------|
| L | OOC length | 128 |
| w | OOC weight | 5 |
| N | Active users | 4 |
| i_d | Photodetector dark current | 10nA |
| η | Photodetector quantum efficiency | 0.8 |
| λ | Optical wavelength | 1550 nm |
| T_c | Chip duration | 4 ns |
| R_L | Load resistance | 10K Ω |
| $\Delta\lambda$ | Optical filter bandwidth | 1nm |
| θ_{FOV} | Receiver field of view angle | 5mrad |
| θ | Transmitter beam divergence | 1mrad |
| G | EDFA gain | 30dB (1000) |
| n_{sp} | Spontaneous emission parameter | 1.1 |
| T_r | Receiver temperature | 300°K |
| L_0 | Channel length | 3Km |
| β | Attenuation coefficient for clear air | 0.1Km ⁻¹ |

**Figure 3.10** Performance comparison of simple correlator with chip level AND logic gate receiver structures for FSO-CDMA link.

of an interference or noise signal is a measure of its strength. For this reason increasing the fading variance, BER increases as depicted in (Figure 3.10). The figure compares the performance of simple correlator receiver and chip level AND logic gate receiver. As it is shown in this figure, simple correlator outperforms the chip level AND logic

gate structures in low power regions because, the system is shot noise limited. As the transmitted power increases the system becomes interference limited. So, chip level AND logic gate structure outperforms the simple correlator.

3.5 Modulation

With the advent of wireless and fiber-optic OCDMA technology, a large body of research activity has been carried out on finding powerful code structures and effective modulations and signalings that can enhance such systems' performance and capacity. In particular, we introduced one such powerful code structure, namely, OOC. On the other hand for effective modulation on-off keying (OOK) and pulse position modulation (PPM) are among the most desired and employed modulation techniques in typical intensity modulation/direct detection (IM/DD) optical communication systems. The immense interest in OOK and PPM modulations stems from two basic facts, namely, the ease with which they can be transmitted and received (implementation), and secondly it is shown that by invoking divergence theorem in a typical optical channel (Poisson channel) the aforementioned signal modulations are optimum [2].

In OOK-OCDMA systems bit '1' is mapped on to L (code length) sequential chips in which w (code weight) chip positions that are determined by the code structure are *on* and the remaining $L - w$ chips are *off*. Furthermore bit '0' is mapped on to L zero sequential chips. The implementation of this modulation is simple and the performance is acceptable in most applications. Computing the optimum threshold in the correlator structure is a necessity for OOK modulation. This implies that after the correlator a hard-limiter should decide on the transmitted bit by comparing the detected energy to an optimum threshold. It is shown that in some applications the performance of the receiver is very sensitive to the threshold especially in such cases that the received power is not high enough. As an example in free-space optical links there is a considerable performance difference of systems with and without adaptive thresholding [37].

In M-ary PPM, the other propounded modulation in IM/DD optical communication systems, optical pulses are transmitted in smaller time durations and in different symbols that are distinguished by separating the time domain. Inherently M-ary PPM is more efficient in energy but less efficient in bandwidth when compared to OOK. To implement M-ary PPM in a typical OCDMA system, each user's symbol interval (slot) is divided into M spreading intervals where each interval could contain the corresponding OOC. To send the m th symbol the OOC of the corresponding user is placed in the m th interval and the remaining intervals remain *off* as shown in Figure 3.11. At the receiver end the correlated energy in each symbol duration is obtained and the maximum is selected as the transmitted symbol. So if at the receiver side we use correlator structure there is no need to compute an optimum threshold.

In general the use of M-ary PPM, which is among the class of energy-efficient but bandwidth-inefficient modulation methods, reduces the system throughput. However its use in conjunction with OOC-based OCDMA could effectively reduce the multiaccess interference thereby enhancing the performance of a typical OCDMA system.

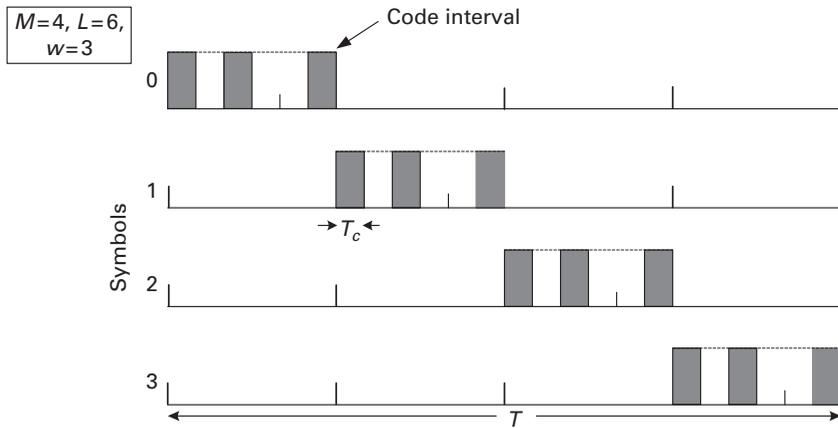


Figure 3.11 M-ary PPM OCDMA symbols, $M = 4$.

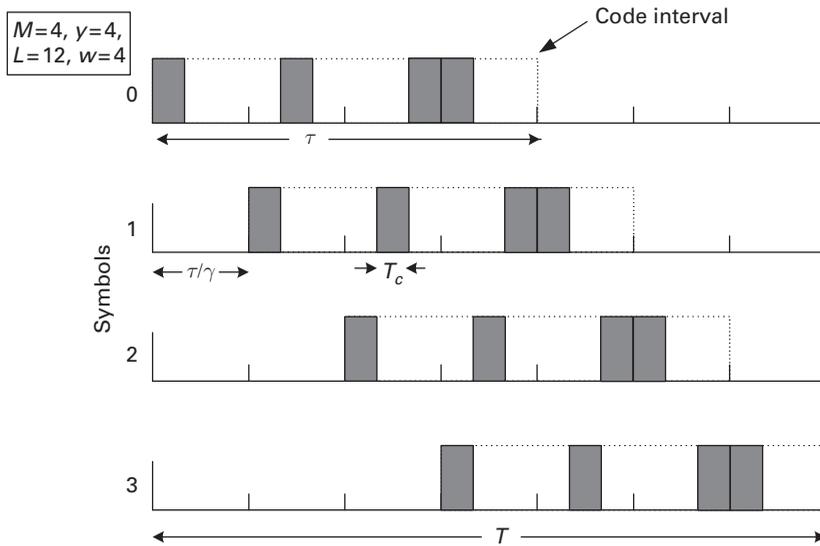


Figure 3.12 OPPM OCDMA symbols.

To mitigate the degradation in throughput efficiency of M-ary PPM, Overlapping PPM (OPPM) was suggested for optical systems [40]–[42]. Figure 3.12 shows a typical OPPM OCDMA system modulation. In this modulation, adjacent symbols are not completely disjoint in time domain but they are overlapped. The amount of the overlap is defined by the modulation parameter. The greater this parameter the better the throughput efficiency, but, on the other hand, there is more interference between symbols. Therefore one can make a trade-off between throughput efficiency and interference level in this modulation. In OPPM similar to M-ary PPM there is no need for threshold setting in the correlator receiver. M-ary PPM can be considered as a special case of

OPPM if we set the modulation parameter to unity. Here if we consider an OOC code with length L and given laser pulse duration T_c , PPM multiplicity M , and index of overlap γ , we divide a spreading interval of duration $\tau = LT_c$ into γ smaller subintervals with width τ/γ and define OPPM modulator unit delay as LT_c/γ . In a single OPPM frame each spreading interval is located at a unit delay with respect to the previous one. Thus adjacent spreading intervals have $(1-1/\gamma)\tau$ seconds overlap and since there are in total M symbols, an OPPM frame length equals in seconds:

$$T = (M - 1)\frac{\tau}{\gamma} + \tau = (M - 1 + \gamma)\frac{L}{\gamma}T_c. \quad (3.32)$$

Note that in order to fit unit delay properly in the symbol interval, the following constraint should be met

$$\frac{\tau}{\gamma} = \text{integer} \times T_c \quad \text{or} \quad \frac{L}{\gamma} = \text{integer}. \quad (3.33)$$

One should note that M-arry PPM is a special case of M-ary OPPM signalling in which the index of overlap, γ , equals 1 that means that there is no overlap between the adjacent spreading intervals. As shown in Figure 3.12, to send the m th symbol we place the OOC of the desired user in the m th spreading interval and make all remaining chips *off* in the corresponding frame of information. Considering n_i as the number of collected photons in the i th spreading interval, the receiver detects the m th symbol if for all $i, 0 \leq i \leq M-1$ and $i \neq m$ we have $n_m > n_i$.

Studying the performance of OPPM in OCDMA systems will lead us to different results than typical optical systems. It can be shown that using OPPM in OOC-based OCDMA systems improves the performance in both throughput and energy efficiency. To explain the reason intuitively we can say that OOC codes are typically sparse. Therefore they allow us to overlap the codes without increasing the inter-symbol interference by much. Furthermore it is shown that by choosing proper parameters for the modulation, OPPM outperforms OOK and M-ary PPM [43].

To compare the performance of the three mentioned modulations in OCDMA systems it is sufficient to consider the case of interference limited (high enough SNR) and the chip-level AND logic gate receiver structure.

3.5.1 Performance analysis

An upper bound on the error probability of the above-mentioned modulations in OCDMA systems using OOC codes and chip-level AND logic gate as the receiver structure is obtained. For the performance evaluation we only consider MAI noise in order to obtain the limits or the floor on the performance for different modulation OOC-based OCDMA systems.

For a system with N users where each user is assigned an OOC with code weight w , code length L , and cross-correlation coefficient $\lambda = 1$ the probability of error for OOK modulation can be found as follows [12],

$$P_{e/OOK} = \frac{1}{2} \left[1 + \sum_{k=1}^w (-1)^k \binom{w}{k} \left(1 - k \frac{w}{2L} \right)^{N-1} \right]. \quad (3.34)$$

For M-ary PPM modulations the upper bound of error probability can be written as follows:

$$P_{e/M-PPM} \leq \frac{M}{2} \left[1 + \sum_{k=1}^w (-1)^k \binom{w}{k} \left(1 - k \frac{w}{ML} \right)^{N-1} \right] \quad (3.35)$$

and for OPPM with some approximation the following expression can be obtained [43]:

$$P_{e/OPPM} \leq \frac{M}{2} \left[1 + \sum_{k=1}^w (-1)^k \binom{w}{k} \left(1 - k \frac{\gamma w}{(M-1+\gamma)L} \right)^{N-1} \right]. \quad (3.36)$$

To compare the performance of OOK and PPM systems assume a fixed value for chip duration, i.e., T_c . Also for a fair comparison assume that for all systems, the amount of information per chip that each user carries is constant. If this value is shown by R_0 then $R_0 = 1/2L_{OOK}$ bits/chip for OOK, $R_0 = \log_2 M/ML_{PPM}$ bits/chip for M-ary PPM and $R_0 = \gamma \log_2 M/(M-1+\gamma)L_{OPPM}$ bits/chip for OPPM. R_0 is directly related to the physical throughput of the system. We obtain the total throughput of the modulation by $R_0 \times N$.

For comparing the probability of error $R_0 = 0.002$ is kept constant for all modulation. The results are plotted in Figure 3.13 where the cross-correlation value is $\lambda = 1$ for the OOC and $M = 16$ and $\gamma = 16$ for PPM modulations. Results show that OPPM

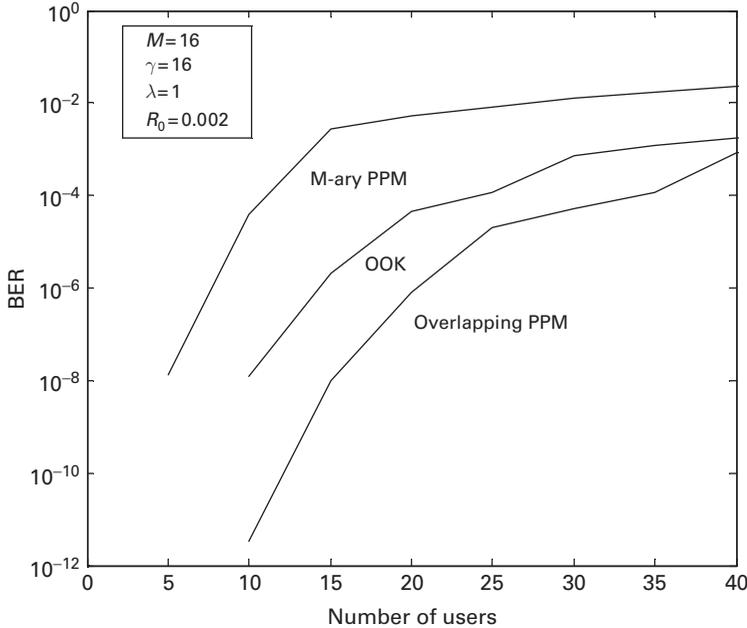


Figure 3.13 BER versus number of users for OOK, M-ary PPM, and OPPM.

outperforms the other modulations by means of decreasing the error probability of the system, i.e., accommodating more users.

3.5.2 Generalized OOC

The first attempt to define and establish optical orthogonal codes was based on minimizing the amount of interference between each two codes in such a way that there is no need for network synchronization. In this case it is evident that the minimum amount of interference or cross-correlation between each two unipolar binary streams is 1 when there is no timing reference. This is the best case when one is looking at the mutual interference. However, in this case the value of code weight is restricted by the number of codes according to the Johnson bound. From a throughput efficiency point of view it is desired that the code length be as short as possible. So in the case of strict OOC ($\lambda = 1$), the product of users and throughput per user of the system is limited. To increase the value of users \times throughput one can relax the cross-correlation coefficient. It can be seen from the Johnson bound that by increasing λ the cardinality of codes will increase if we keep the code weight as before. In the first view one can deduce that by increasing λ , the multiple-access interference increases and the system performance degrades. But this is not necessarily true since in this case we can increase the code weight which in itself could improve the performance by increasing the coding gain. Briefly, there is a trade-off between λ and w and there is an optimum point for the code parameters according to the system conditions.

There are many studies on using generalized OOC in OCDMA systems and finding their performance in different conditions. In [12], [16], [44], attempts were made to explore the performance of OCDMA systems that employ OOCs with $\lambda \geq 1$. In particular, in [16], by obtaining lower and upper bounds on the system's performance, it was hinted that OOCs with $\lambda = 2$ could, under certain conditions, outperform OCDMA employing OOCs with $\lambda = 1$, with a cardinality that could be a hundred to a thousand times longer. Similarly, using a structure based on a frame time-hopping ultra-wideband CDMA system, it was shown that the system's performance degrades gradually with respect to OOCs with $1 \leq \lambda < w$, while offering many possible signature sequences. In [12] performance analysis of generalized OOC in an OCDMA system with chip-level AND logic gate receiver structure was obtained. It was shown that the optimum value for λ is between 1 and 3 depending on the design criteria.

To compare the performance of OOCs in OCDMA systems as a function of the code's cross-correlation coefficient we follow the approach presented in [12]. One can assume that the probability of interference of an undesired code on the i th specified marked chip positions of the desired OOC is p_i . It is obvious that the value of p_i directly depends on the code set and the position of the marked chips belonging to the desired and interfering codes. However, as an approximation we assume that p_i is the same for all interfering codes. In this case we have,

$$\sum_{i=1}^{\lambda} i \binom{w}{i} p_i = \frac{w^2}{2L}. \quad (3.37)$$

The probability of error caused by MAI is obtained as follows,

$$P_e = \frac{1}{2} \Pr(\text{error}|1) + \frac{1}{2} \Pr(\text{error}|0). \quad (3.38)$$

Neglecting shot noise the first term on the right-hand side of the above equation is equal to zero. However an error occurs by MAI noise only when the transmitted bit is 0. If we show the number of interference on the i th marked chip of the desired OOC by α_i then the probability of error is as follows [12],

$$\begin{aligned} P_e &= \frac{1}{2} \Pr(\text{error}|0) = \frac{1}{2} \Pr(\alpha_1 > 0, \alpha_2 > 0, \dots, \alpha_w > 0) \\ &= 1 - \Pr(\alpha_1 = 0 \text{ or } \alpha_2 = 0 \text{ or } \dots \text{ or } \alpha_w = 0) \\ &= 1 + \sum_{k=1}^w (-1)^k \binom{w}{k} \Pr(\alpha_1 = \alpha_2 = \dots = \alpha_k = 0). \end{aligned} \quad (3.39)$$

The λ events that are related to p_1, p_2, \dots , and p_λ are disjoint. Furthermore all $N - 1$ interfering users are independent and hence we can write:

$$\Pr(\alpha_1 = \alpha_2 = \dots = \alpha_k = 0) = \left[1 + \sum_{l=1}^{\lambda} (-1)^l \binom{k}{l} \rho_l \right]^{N-1} \quad (3.40)$$

where ρ_l is defined as:

$$\begin{aligned} \rho_l &= \Pr(\alpha_1 = \alpha_2 = \dots = \alpha_l = 1 \mid \text{one interfering user}) \\ &= p_l + \binom{w-l}{1} p_{l+1} + \dots + \binom{w-l}{\lambda-l} p_\lambda = \sum_{i=l}^{\lambda} \binom{w-l}{i-l} p_i. \end{aligned} \quad (3.41)$$

Substituting (3.41) in (3.40) and using (3.39) we can rewrite (3.38) as follows:

$$P_e = \frac{1}{2} \left\{ 1 + \sum_{k=1}^w (-1)^k \binom{w}{k} \left[1 + \sum_{l=1}^{\lambda} \sum_{i=l}^{\lambda} (-1)^l \binom{k}{l} \binom{w-l}{i-l} p_i \right]^{N-1} \right\}. \quad (3.42)$$

However, it is shown that if OOC codes are chosen such that $p_k = 0$ for $1 \leq k \leq \lambda - 1$ and thus only $p_\lambda \neq 0$, we obtain an upper bound on the performance of such an OCDMA system with generalized OOCs.

To analyze the effect of cross-correlation of codes on the system performance we obtain the maximum achievable throughput of the system versus number of users in the case which the bit error rate does not exceed a predefined value, e.g., 10^{-9} . The goal is to find the minimum value for the code length, assuming a fixed value for number of users by finding the optimum value for cross-correlation and code weight using the Johnson bound. The results have been sketched in Figure 3.14 where it can be seen that the optimum value for λ varies between 2 and 3. It is noteworthy that for $\lambda > 1$ we consider the worst case, i.e., $p_k = 0$ for $1 \leq k \leq \lambda - 1$ and $p_\lambda = w^2/2L\lambda \binom{w}{\lambda}$. Such codes may not necessarily exist but from a mathematical point of view it gives an upper bound on the performance of all possible and existing codes.

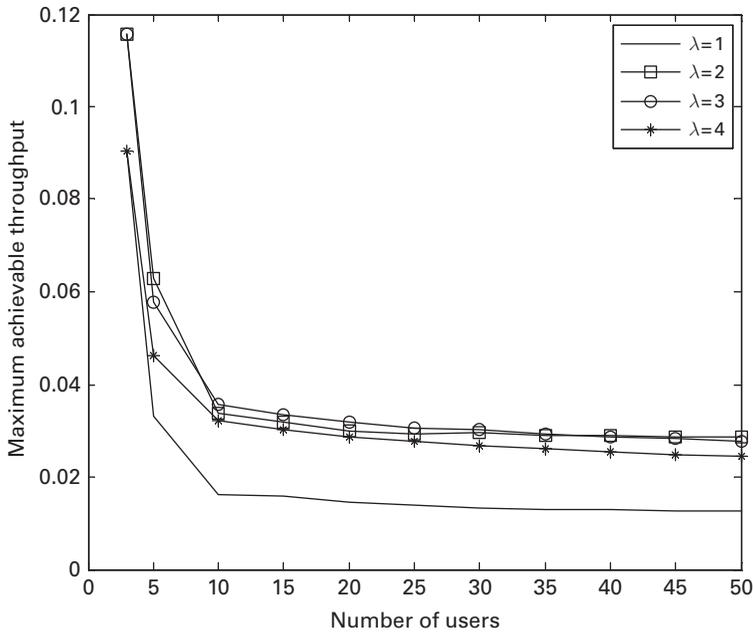


Figure 3.14 Maximum achievable throughput vs. number of users for OCDMA chip-level AND logic gate receiver.

3.6 Experimental prototypes

3.6.1 Indoor wireless optical CDMA LAN

We have shown the implemented block diagram of wireless OCDMA indoor prototype in Figure 3.15 [19]. In the transmitter side we have used dual port memory in which the training bits and matched pattern are put in the header section. Data bit stream is written in the remaining section of the memory while to generate the whole frame, including data bits and header, data is read from the first to last cell of memory. The relation between the speed of the read and write clocks is equal to the relation of the total length of frame to the payload part of frame. The corresponding OOC is loaded in a shift register. The chip rate clock feeds the shift register. The output of the shift register is combined with the output of the memory through an AND gate. This structure simply generates the coded bits.

The outcome of the digital part of the transmitter feeds the analog part of the transmitter that includes fast transistors as the driver prior to the LEDs operating in infrared wavelength (here 870 nm).

In the receiver part we have used PIN photodiode following by a three-stage amplifier circuit. The gains of the three stages are computed such that they support the nominal parameters of the system. In the implemented indoor system we have designed the analog circuit in the receiver such that it supports the distance range of transmitter and

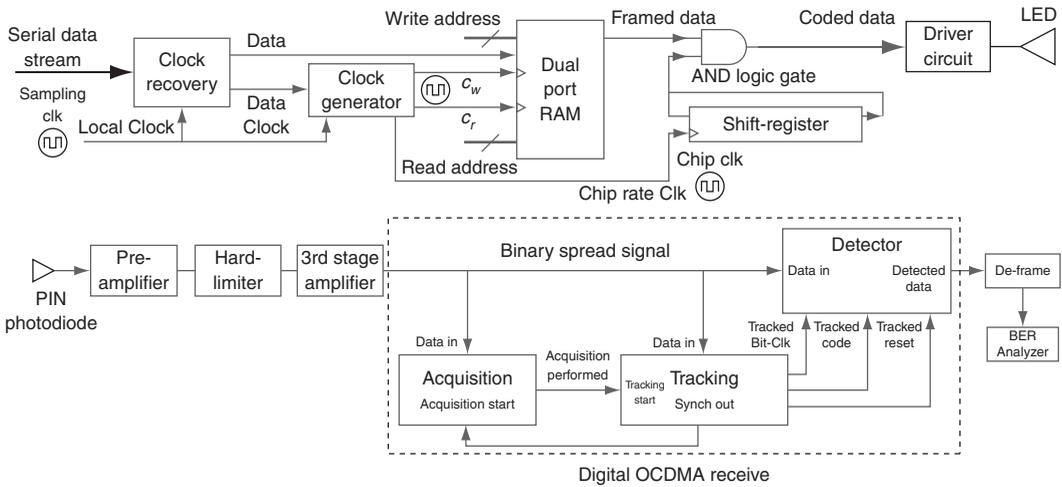


Figure 3.15 Transmitter/receiver block diagram of wireless indoor OCDMA system.

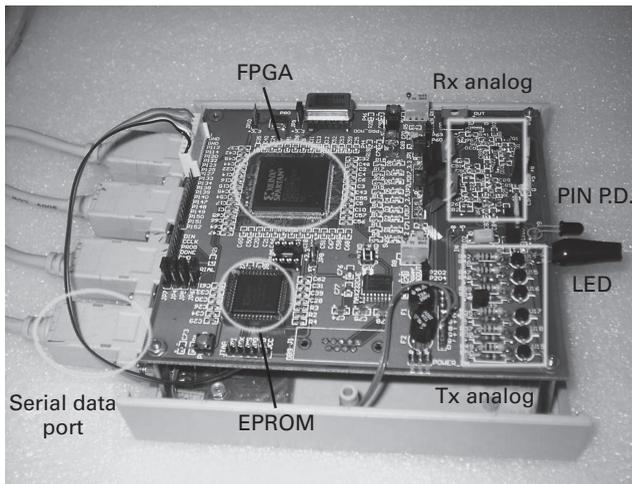


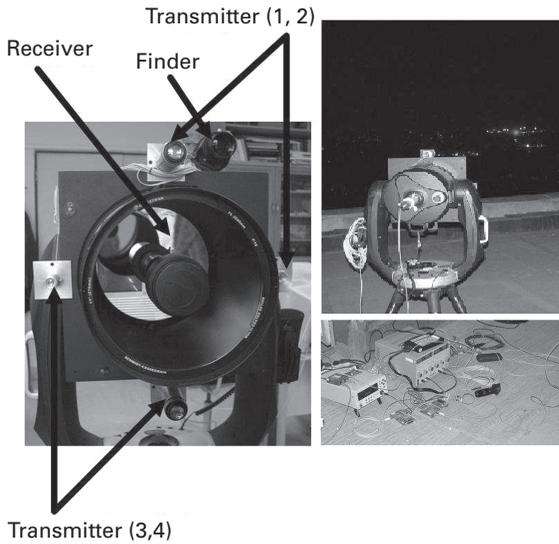
Figure 3.16 The electronic board designed for the base station, supporting up to four serial data terminals.

receiver 0.5 to 3.0 meters and half-angle of LED equal to 10° . The second stage of amplifier acts as the hardlimiter, i.e., clips the received pulses from the first stage. The third stage amplifies the pulses coming from the second stage to the proper level for the digital part, i.e., FPGA.

Binary detected chips go to the digital section where we have acquisition, tracking, and detection blocks. We described the operation of digital blocks in the previous sections. A simple de-frame block extract data bits from the frame. We have shown the implemented prototype in Figure 3.16.

Table 3.2 Specifications of FSO link

| Parameter | Specification |
|--------------------------|----------------|
| Operating wavelength | 1550 nm |
| Telescope aperture size | 11.25 inch |
| Diversity in transmitter | 2 or 4 |
| EDFA output power | 15-30 dBm |
| Coupling in receiver | MMF |
| Beam divergence | ± 0.4 mrad |

**Figure 3.17** Prototype of FS-OCDMA system.

3.6.2 FSO-CDMA link

The block diagram of the implemented prototype of the FSO system is shown in Figure 3.9. In the transmitter side we have used ordinary high-speed SFP laser modules operating at 1550 nm wavelength. The laser is modulated by the data stream coming from another digital board. A high power EDFA with variable adjustable gain amplifies the power of laser to the desired power level. The maximum output of EDFA in the prototype is 30 dBm. However, we use the range 15 dBm to 30 dBm using the proposed power control schema. The output of EDFA is split to four branches, by employing passive SMF couplers, in order to use diversity in the transmitter side.

To support distance of 5 Km we employ an 11.25 inch telescope in the receiver side. The received light is coupled in a multimode fiber and then goes for further processes in the following blocks. The system supports fast and gigabit Ethernet as well as STM1. It also supports E1-4E1 with OCDMA and different QoS. A photograph of the prototype is shown in Figure 3.17 and the system specifications are given in Table 3.2.

Acknowledgment

The authors would like to thank the Iran National Science Foundation (INSF) for supporting this research.

References

- [1] S. Karp, R. M. Gagliardi, S. E. Moran, and L. B. Stotts, *Optical Channels Fibers, Clouds, Water and the Atmosphere*. Plenum, 1980.
- [2] R. M. Gagliardi and S. Karp, *Optical Communications*. John Wiley, 1976.
- [3] F. R. Gfeller and U. H. Bapst, "Wireless in-house data communication via diffuse infrared radiation," *Proc. IEEE*, vol. 67, no. 11, pp. 1474–1486, Nov. 1979.
- [4] J. R. Barry, *Wireless Infrared Communications*. Kluwer, 1994.
- [5] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proc. IEEE*, vol. 85, no. 2, pp. 265–298, Feb. 1997.
- [6] J. Strohbehn, ed., *Laser Beam Propagation in the Atmosphere*. Springer, 1978.
- [7] B. T. Binder, P. T. Yu, J. H. Shapiro, and J. K. Bounds, "An atmospheric optical ring network," *IEEE Trans. Commun.*, vol. 38, no. 1, pp. 74–81, Jan. 1990.
- [8] M. R. Pakravan, M. Kavehrad, and H. Hashemi, "Indoor wireless infrared channel characterization by measurements," *IEEE Trans. Vehic. Tech.*, vol. 50, no. 4, pp. 1053–1073, July 2001.
- [9] S. Hranilovic, *Wireless Optical Communication Systems*. Springer, 2005.
- [10] J. A. Salehi, "Code division multiple-access techniques in optical fiber networks—part I: fundamental principles," *IEEE Trans. Commun.*, vol. 37, pp. 824–833, Aug. 1989.
- [11] J. A. Salehi and C. A. Brackett, "Code division multiple-access techniques in optical fiber networks—part II: system performance analysis," *IEEE Trans. Commun.*, vol. 37, pp. 834–842, Aug. 1989.
- [12] S. Mashhadi and J. A. Salehi, "Code division multiple-access techniques in optical fiber networks—part III: optical AND gate receiver structure with generalized optical orthogonal codes," *IEEE Trans. Commun.*, vol. 45, pp. 1457–1468, Aug. 2006.
- [13] S. Mashhadi and J. A. Salehi, "Optimum code structures for positive optical CDMA using normalized divergence maximization criterion," *IEEE Trans. Commun.*, vol. 55, pp. 1414–1421, Sept. 2008.
- [14] I. Arruego *et al.*, "OWLS: A ten-year history in optical wireless links for intra-satellite communications," *IEEE J. of Selected Areas in Commun. (JSAC)*, vol. 27, no. 9, pp. 1599–1611, Dec. 2009.
- [15] P. J. Winzer, "Modulation and multiplexing in optical channels," Lasers and Electro-Optics, 2009 and 2009 Conference on Quantum Electronics and Laser Science Conference. CLEO/QELS 2009, pp. 1–2.
- [16] M. Azizoglo, J. A. Salehi, and Y. Li, "Optical CDMA via temporal codes," *IEEE Trans. Commun.* vol. 40, 1162–1170, 1992.
- [17] J. M. H. Elmurghani and R. A. Cryan, "Hybrid PPM-CDMA systems utilising OOC for indoor wireless infrared communication," *Microwave and Opt. Tech. Lett.*, vol 8, no. 1, pp. 44–47, Jan. 1995.
- [18] B. M. Ghaffari, M. Matinfar, J. A. Salehi, "Wireless optical CDMA LAN: digital design concepts," *IEEE Trans. Commun.* vol. 56, no. 12, pp. 2145–2155, Dec. 2008.

-
- [19] B. M. Ghaffari, M. D. Matinfar, and J. A. Salehi, "Wireless optical CDMA LAN: digital implementation analysis," *IEEE J. of Selected Areas in Commun. (JSAC)*, vol. 27, no. 9, pp.1676–1686, Dec. 2009.
- [20] F. R. K. Chung, J. A. Salehi, and V. K. Wei, "Optical orthogonal codes: design, analysis, and applications," *IEEE Trans. Inf. Theory* vol. 35, no. 3, pp. 595–604, May 1989.
- [21] H. Chung and P. Vijay Kumar, "Optical orthogonal codes – new bounds and an optimal construction," *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 866–873 Jul. 1990.
- [22] S. M. Johnson, "A new upper bound for error-correcting codes," *IRE Trans. Inf. Theory*, vol. IT-8, pp. 203–207, 1962.
- [23] H. M. Shalaby, "Chip level detection in optical code division multiple access," *J. Lightwave Technol.*, vol. 16, pp. 1077–1087, Jun. 1998.
- [24] T. Ohtsuki, "Performance analysis of direct-detection optical asynchronous CDMA systems with double optical hard-limiters," *J. Lightwave Technol.*, vol. 15, no. 3, pp. 452–457, Mar. 1997.
- [25] U. N. Griner and S. Arnon, "Multiuser diffuse indoor wireless infrared communication using equalized synchronous CDMA," *IEEE Trans. Commun.*, vol. 54, pp. 1654–1662, Sep. 2006.
- [26] A. Aminzadeh-Gohari and M. R. Pakravan, "Analysis of power control for indoor wireless infrared CDMA communication," in *Proceedings of the 25th IEEE Performance, Computing, and Communications Conference (IPCCC)* (IEEE, 2006), pp. 297–302.
- [27] S. Zahedi and J. A. Salehi, "Analytical comparison of various fiber-optic CDMA receiver structures," *J. Lightwave Technol.*, vol. 18, no. 12, pp. 1718–1727, Dec. 2000.
- [28] A. Keshavarizan and J. A. Salehi, "Optical orthogonal code acquisition in fiber-optic CDMA systems via the simple serial-search method," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 473–483, Mar. 2002.
- [29] A. Keshavarizan and J. A. Salehi, "Multiple-shift code acquisition of optical orthogonal codes in optical CDMA systems," *IEEE Trans. Commun.*, vol. 53, no. 4, pp. 687–697, Apr. 2005.
- [30] R. L. Peterson, D. E. Borth, and R. E. Ziemer, *Introduction to Spread Spectrum Communications*. Prentice-Hall, 1995.
- [31] Int. Electrotech. Commission, *CEI/IEC 825-1: Safety of Laser Products*, 1993. J. D. Rancourt, *Optical Thin Film*. Macmillan, 1987.
- [32] R. M. Gagliardi, "Pulse-coded multiple access in space optical communications," *IEEE JSAC*, vol. 13, no. 3, pp. 603–608, 1995.
- [33] T. H. Carbonneau and D. R. Wisely, "Opportunities and challenges for optical wireless; the competitive advantage of free space telecommunications links in today's crowded marketplace," in *Proc. SPIE Wireless Technologies and Systems: Millimeter-Wave and Optical*, vol. 3232, pp. 119–128, Jan. 1998.
- [34] P. F. Szajowski, G. Nykolak, J. J. Auburn, H. M. Presby, and G. E. Tourgee, "High power optical amplifiers enable 1550 nm terrestrial freespace optical data-link operating at 10 Gbs," in *Proc. MILCOM '99*, vol. 1, pp. 687–689, 1999.
- [35] T. Li and M. C. Teich, "Photon point process for traveling-wave laser amplifiers," *IEEE J. Quantum Electron.*, vol. 29, no. 9, pp. 2568–2578 Sep. 1993.
- [36] M. Razavi and J. H. Shapiro, "Wireless optical communications via diversity reception and optical preamplification," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 975–983, 2005.
- [37] M. Jazayerifar and J. A. Salehi, "Atmospheric optical CDMA communication systems via optical orthogonal codes," *IEEE Trans. Commun.* vol. 54, no. 9, pp. 1614–1623 Sept. 2006.

- [38] B. Hamzeh and M. Kavehrad, "OCDMA-coded free-space optical for wireless optical-mesh networks," *IEEE Trans. Commun.*, no. 12, pp. 2165–2174, Dec. 2004.
- [39] T. Ohtsuki, "Performance analysis of atmospheric optical PPM systems," *J. Lightwave Technol.*, vol. 21, no. 2, pp. 406–411, Feb.
- [40] G. M. Lee and G. W. Schroeder, "Optical PPM with multiple positions per pulse-width," *IEEE Trans. Commun.*, vol. COM-25, pp. 360–364, Mar. 1977.
- [41] C. N. Georghiadis, "Modulation and coding for throughput-efficient optical systems," *IEEE Inform. Theory*, vol. 40, pp 1313–1326, Sept. 1994.
- [42] H. M. H. Shalaby, "Maximum achievable throughputs for uncoded OPPM and MPPM in optical direct-detection channels," *J. Lightwave Technol.* vol. 13, no. 11, pp. 2121–2128, Nov. 1995.
- [43] H. M. H. Shalaby, "Direct detection optical overlapping PPM-CDMA communication systems with double optical hardlimiters," *J. Lightwave Technol.* vol. 17, pp. 1158–1165, July 1999.
- [44] J. J. Chen and G. Yang, "CDMA fiber-optic systems with optical hard limiters," *J. Lightwave Technol.*, vol. 19, no. 7, pp. 950–958, Jul. 2001.
- [45] A. R. Forouzan, J. A. Salehi, and M. Nasiri-Kenari, "Frame time-hopping fiber-optic code-division multiple access using generalized optical orthogonal codes," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1971–1982, Dec. 2002.

4 Pointing error statistics

Shlomi Arnon

Optical wireless communication transceivers use in many cases very narrow laser beam divergence and very narrow field of view receiver telescope. The laser transmitter uses narrow laser beam divergence in order to minimize the transmitter power while the receiver uses very narrow field of view telescope in order to reduce background radiation noise [1–7]. As a result the pointing between the transmitter and the receiver is complicated and in order to establish optical communication between two transceivers, the line of sight of their optics must be aligned during the entire communication time. Optical acquisition and tracking system is used to keep line of sight of the transceivers. However, from time to time due to electronic noise, and mechanical vibration (e.g. wind in an urban system [2, 3] or jitter of antenna of a satellite [8, 9]) error in pointing direction occurs. In the following we present a simple model that describes the effect of the statistic of the pointing error on the performance of the communication system.

We assume that the orthogonal direction of error pointing angle models are based on a normal distribution [1]. In that case the elevation pointing error angle is normally distributed with a probability density

$$f(\theta_V) = \frac{1}{\sqrt{2\pi}\sigma_V} \exp\left(-\frac{\theta_V^2}{2\sigma_V^2}\right), \quad (4.1)$$

where σ_V and θ_V are elevation pointing standard deviation and elevation pointing respectively.

The azimuth pointing error angle is normally distributed with probability density

$$f(\theta_H) = \frac{1}{\sqrt{2\pi}\sigma_H} \exp\left(-\frac{\theta_H^2}{2\sigma_H^2}\right), \quad (4.2)$$

where σ_H and θ_H are azimuth pointing standard deviation and azimuth pointing angle, respectively.

The radial pointing error angle is the root square sum of the azimuth and elevation angles

$$\theta = \sqrt{\theta_V^2 + \theta_H^2}. \quad (4.3)$$

Based on symmetry we can assume that

$$\sigma_V = \sigma_H = \sigma. \quad (4.4)$$

We assume that the azimuth and elevation processes are independent and identically distributed so the radial pointing error angle model is Rayleigh distributed with probability density

$$f(\theta) = \frac{\theta}{\sigma^2} \exp\left(-\frac{\theta^2}{2\sigma^2}\right). \quad (4.5)$$

Now, we derive a model that relates the transmitted optical signal from the transmitter to the receiver. The distance between the transmitter and the receiver satellite is z meters. The instantaneous received power as a function of pointing direction error angle θ is

$$P_R(\theta) = P_T \eta_T \eta_R \left(\frac{\lambda}{4\pi z}\right)^2 G_T G_R \exp(-G_T \cdot \theta^2) \quad (4.6)$$

where η_R , η_T , and $G_R = \left(\frac{\pi D_R}{\lambda}\right)^2$ and $G_T = \left(\frac{2\pi W}{\lambda}\right)^2$ are the optical efficiencies and the telescope gain of the receiver and the transmitter respectively. And where D_R is the receiver aperture diameter, W is Gaussian rms width at the transmitter aperture, the wavelength of the laser transmitter is λ , and P_T is the optical laser transmitter power.

The received signal is given by

$$\mu(\theta) = R \cdot P_R(\theta). \quad (4.7)$$

The signal-to-noise ratio (SNR) model is

$$\text{SNR} = \int_0^{\infty} \frac{\mu(\theta)}{\sigma} f(\theta) \cdot d\theta, \quad (4.8)$$

where σ_n is the noise variance. The receiver is assumed to include an optical detector in direct detection mode. In such systems the instantaneous error probability is given by

$$\text{BER}(\theta) \approx \frac{1}{2} \left(\text{erfc} \left(\frac{\mu(\theta)}{\sqrt{2}(2\sigma_n)} \right) \right), \quad (4.9)$$

where the complementary error function is given by

$$\text{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x \exp(-y^2) dy. \quad (4.10)$$

The average bit error probability is given

$$\overline{\text{BER}} = \int_0^{\infty} \text{BER}(\theta) \cdot f(\theta) \cdot d\theta. \quad (4.11)$$

References

- [1] Arnon., S. Rotman, and N. S Kopeika, "The performance limitations of free space optical communication satellite networks due to vibrations – digital case," *Optical Engineering*, vol. 36, no. 11, pp. 3148–3157, (Nov. 1997).
- [2] S. Arnon, "The effects of atmospheric turbulence and building sway on optical wireless communication systems," *Optics Letters*, Vol. 28, no. 2, pp. 129–131 (Jan. 2003).
- [3] D. Kedar and S. Arnon, "Urban optical wireless communication network: The main challenges and possible solutions," *IEEE Optical Communications Supplement to IEEE Communications Magazine*, pp. S1–S7, (May 2004).
- [4] S. Arnon and N. S Kopeika, "Adaptive optical transmitter and receiver for space communication through clouds," *Applied Optics*, vol. 36, no. 9, pp. 1987–1993, (March 1997).
- [5] S. Arnon and N. S. Kopeika, "The performance limitations of free space optical communication satellite networks due to vibrations – analog case," *Optical Engineering*, vol. 36, no.1, pp.175–182, (Jan. 1997).
- [6] S. Arnon and N. S Kopeika, "Probing and monitoring aerosol and atmospheric cloud via an electro-optic oscillator," *Applied Optics*, vol. 35, no. 27, pp. 5427–5434, (Sep. 1996).
- [7] S. Arnon and N. S Kopeika, "Free space optical communication: analysis of spatial widening of optical pulses for propagation through clouds," *Optical Engineering*, vol. 34, no. 2, pp. 511–515, (Feb. 1995).
- [8] S. Arnon and N. S Kopeika, "Laser satellite communication networks-vibration effects and possible solutions," *Proceedings of the IEEE*, vol. 85, no. 10, pp. 1646–1661, (Oct 1997).
- [9] S Arnon, S. Rotman, and N. S Kopeika, "Beam-width and transmitter power adaptive to tracking system performance for free space optical communication," *Applied Optics*, vol. 36, no 24, pp. 6095–6101, (Aug. 1997).

5 Equalization and Markov chains in cloud channel

Mohsen Kavehrad

Abstract: Free-space optical (FSO) communications is a practical solution for creating a three-dimensional global broadband communications grid, offering bandwidths far beyond those possible in Radio Frequency (RF) range. However, attributes of atmospheric turbulence and obscurants such as clouds impose perennial limitations on availability and reliability of optical links. To design and evaluate optimum transmission techniques that operate under realistic atmospheric conditions, a good understanding of the channel behavior is necessary.

In some prior works, the Monte Carlo ray tracing (MCRT) algorithm has been used to analyze the channel behavior. This task is quite numerically intensive. The focus of this chapter is on investigating the possibility of simplifying this task by a direct extraction of state transition matrices associated with standard Markov modeling from the MCRT computer simulations programs. We show that by tracing a photon's trajectory in space via a Markov chain model, the angular distribution can be calculated by simple matrix multiplications. We also demonstrate that the new approach produces results that are close to those obtained by MCRT and other known methods. Furthermore, considering the fact that angular, spatial, and temporal distributions of energy are interrelated, mixing time of Monte Carlo Markov chain (MCMC) for different types of aerosols is calculated based on eigen-analysis of the state transition matrix and possibility of communications in scattering media is investigated. We also consider in this chapter signal processing techniques for airborne FSO wireless communications through clouds. The FSO channel is known to be accompanied by multi-scattering, which causes severe inter-symbol interference in digital transmissions at high data rates. The chapter investigates the feasibility of digital signal processing schemes applied to FSO communications. By having a channel shortening equalizer (CSE) or time equalization (TEQ) at the transmitter and a Viterbi equalizer at the receiver, reliable communication at low-to-mid level optical thickness values is shown to be feasible. We also suggest how existing TEQ algorithms can be modified to fit into TEQ-in-transmitter configurations. Performance is evaluated in average bit error rate (BER), and compared with conventional equalization techniques.

Key words – angular dispersion, channel modeling, Markov chain, Monte Carlo ray tracing, multi-scattering, second largest eigen modulus (SLEM), channel-shortening, Viterbi algorithm,

5.1 Introduction

Free-space optical (FSO) communications is the only practical solution for creating a broadband three-dimensional global communications grid among ground and airborne nodes, due to its ease of deployment, enormous available bandwidth, possibility of bandwidth reuse, and inherent security at physical layer, as laser beams are normally spatially confined and cannot be tapped [1, 2]. In ideal free-space, the total loss due to absorption and scattering is virtually zero. This is because of the spatial confinement of laser beams. However, atmospheric obscurants such as fog, haze, smoke, dust, and clouds turn the propagation environment into a multiple scattering medium and hence introduce laser pulse broadening in space and time. Moreover, even in clear weather, turbulence due to refractive index fluctuations in layers of atmosphere causes wavefront distortion, introducing both random phase and intensity fluctuations across the transmitted light beam [3, 4]. In presence of multi-scattering, both availability and capacity of optical link degrade, significantly.

Turbulent air creates small-scale fluctuations of the refractive index along a waveform propagation path causing random amplitude, phase, and angle-of-arrival variations known as “turbulence.”

Turbulence distorts an optical beam cross-section spatially, causing random beam dancing, phase-front distortion, etc., in imaging. These spatial adverse effects contribute to image blurring. In transmissions or communications using simple Intensity Modulation/Direct Detection, turbulence lends itself approximately to a multiplicative Lognormal fading coefficient (flat fading) – known as scintillation which is a temporal effect. Where there is moisture, turbulent air can create an upward motion of moist air particles and this is a prerequisite in formation of clouds. Once a cloud is formed, light beam photons traveling through along a propagation path are scattered due to Mie scattering. Received optical power loss associated with light beam passing through clouds, for the most part, is a loss to scattering rather than absorption, in parts of the infrared range.

The focus of this section is on deleterious effects of multi-scatterings and aerosols. Turbulence-induced effects are not further discussed in this work, as sufficient link power budget and diversity can overcome this shortfall [3].

In some applications, transmit power is limited, in order to observe eye and skin safety regulations and on the other hand, sensitivity of optical receivers is affected by the shot noise caused by background light [5]. Hence, link budget management becomes very critical and challenging. Furthermore, multi-path delay spreading limits the achievable bit rates [5]. In digital FSO communications, one is usually interested in the peak power of received waveform in order to distinguish the signal from steady background radiation [6, 7]. In a receiver unequipped with appropriate countermeasure techniques, overlapping of pulse tails on adjacent symbol intervals introduces intersymbol-interference

(ISI), and thus limits the achievable bit rate. As a result, the receiver should have a small field-of-view (FOV) and has to resort to line-of-sight (LOS) photons, known as ballistic, and/or near LOS (snake) photons which are highly forward scattered. However, if due to thickness and/or particle density of obscurants, a light pulse enters the diffusion regime, spatial distribution approaches a normal distribution and angular distribution becomes isotropic [8]. In this case, a receiver of small FOV may fail to maintain the desired link margin.

To answer the fundamental question of optical communications feasibility in presence of clouds, one has to accurately model atmospheric optical channel and estimate parameters such as angular, spatial, and temporal dispersions.

As a collimated laser beam propagates through clouds, it spreads and after traveling some distance, it reaches a steady state of being defocused and diffused. The rate of convergence to this steady state determines feasibility of communications through various clouds of different optical thickness values. Optical thickness τ is the average number of scatterings over a given length of cloud whose value is a number with no physical unit. It is defined by multiplying the scattering coefficient (km^{-1}), β_{sca} , of cloud by the physical cloud length, L , in kilometer. We will elaborate on this issue more, in Section 5.2.

So far, Monte-Carlo ray tracing (MCRT) has been used to calculate the channel parameters [6, 9, and 10–12]. This method requires a high computational capacity and a long execution time. In [8], authors find spatial and angular distribution by a statistical approach. However, their results are limited to the first two moments of multiple scattering. Finding angular distribution is of great importance in computer graphics, as well. In [13], a single-scatter impulse response is defined and then the result is generalized to multiple scattering via convolution. Since this impulse response is three dimensional, the convolution process is quite cumbersome. In this work, angular distribution evolution of a laser beam in a multiple scattering medium is characterized by extracting the corresponding simplified Markov chain information from the MCRT algorithm. By calculating the state transition matrix, one can find the probability distribution function of scattering angle after any number of scatterings. We show that the Markov chain model produces values that are close to MCRT results and other proposed analytical methods, yet need much less time to produce the results. Since angular, spatial, and temporal distributions of received power are inter-related, useful information can be extracted from angular distribution about the behavior of the entire system. The remainder of this chapter is organized as follows. In Section 5.2, channel propagation modeling is reviewed. Section 5.3 details our modeling results and eigen-analyses of the state transition matrix for the cloud channel. This is followed by an investigation of advance digital signal processing techniques, in particular, channel shortening algorithms as apply to cloud channel models. Finally Section 5.5 presents a summary and concludes the chapter.

5.2 Channel propagation modeling

When propagating through clouds or fog, a laser beam interacts with medium particles, which are mostly water droplets of sizes comparable to the optical wavelength values. Hence, we need Mie theory to explain these interactions. This theory is the application

of Maxwell's equation to the problem of a homogeneous sphere radiated by a plane wave from a single direction [14]. Knowing the particle size distribution of the medium, one can determine absorption, extinction, and scattering coefficients as well as phase function, using this theory. Assuming that particles are homogeneously distributed in space, distance between two successive scatters turns out to be an exponential random variable with a mean value of $D_{ave} = \frac{1}{\beta_{sca}}$ where β_{sca} is the scattering coefficient. Furthermore, absorption, extinction, and scattering coefficients are related as: $\beta_{sca} + \beta_{abs} = \beta_{ext}$, where β_{abs} and β_{ext} are absorption and extinction coefficients, respectively. Additionally, the ratio of the scattered energy to the total energy is determined by single scatter albedo, $\varpi = \beta_{sca}/\beta_{ext}$.

According to Mie theory, one can define the scatter direction via a three-dimensional probability distribution function (*PDF*) known as the *Phase Function*. In other words, *Phase Function* is the *PDF* of the solid angle $\Omega = (\theta, \phi)$ [14–16]. For example, in [15] we find that:

$$\int_{\text{Over } 4\pi} P(\theta) d\Omega = \int_0^\pi \int_0^{2\pi} P(\theta) \sin(\theta) d\theta d\varphi = 4\pi \quad (5.1)$$

and where $P(\theta)$ in Eq. (5.1) is referred to as the (un-normalized) scattering phase function in the literature.

Figure 5.1 demonstrates the scattering phase function for different types of clouds. These functions are obtained by substituting the modified gamma distribution for radius density of cloud particles in equations extracted from Mie theory for the poly-dispersed phase function [14, 17]. It is clear from this figure that all these phase functions are highly peaked in the forward direction. Actually, since the azimuth scattering angle, φ , is uniformly distributed in $[0, 2\pi]$, the *Phase Function* can be plotted only against the polar scattering angle, θ . Scattering phase function can also be interpreted as the *PDF* of $\cos(\theta)$ [9, 16].

In order to find suitable probabilities to inject into a Monte Carlo model, one needs the probability density function of θ , not the plotted scattering phase functions. The normalized phase function, as generally used in the literature, is normalized so that its integration over all possible scatter angles (4π steradians) is unity, thus qualifying it to be a *PDF*. With no φ dependence of $P(\theta)$, the integration over φ simply contributes a factor of 2π . This leads to the normalization definition for $P(\theta)$ of:

$$2\pi \cdot \int_0^\pi \frac{P(\theta)}{4\pi} \cdot \sin(\theta) d\theta = \int_0^\pi P(\theta) \cdot \frac{\sin(\theta)}{2} \cdot d\theta = 1.$$

Thus, the *PDF* of θ can be extracted from the normalized phase function and be expressed as:

$$f(\theta) = P(\theta) \cdot \frac{\sin(\theta)}{2}. \quad (5.2)$$

Unfortunately, expressing the phase function versus θ has caused a great deal of confusion in the published literature [18, 19]. More specifically, several authors have mistaken

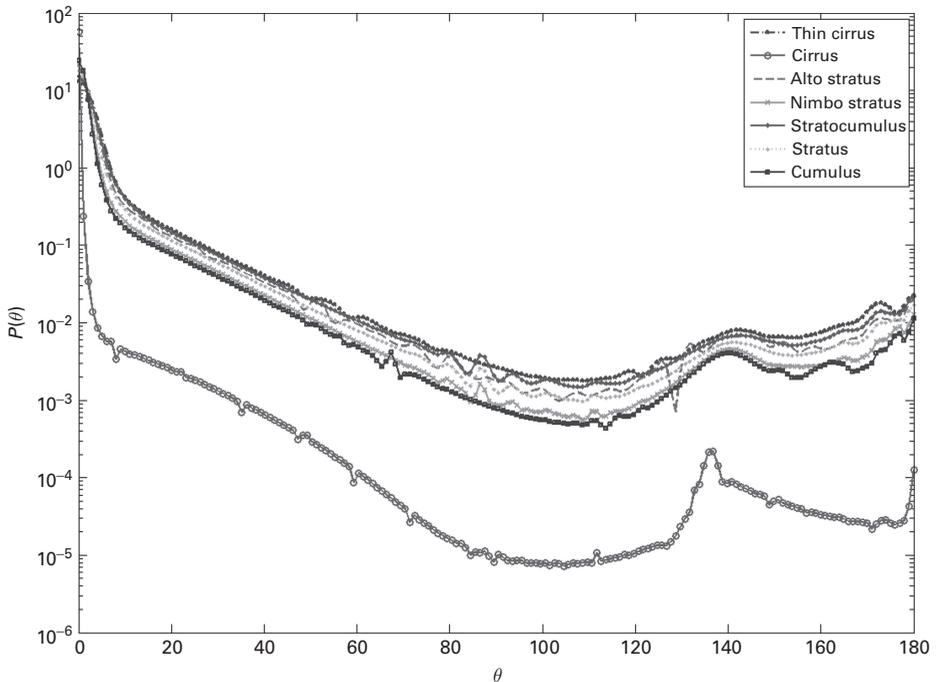


Figure 5.1 Phase functions for different types of clouds.

the phase function for the PDF of θ and obtained an extremely forward-directed angular distribution of scattered beams, which is incorrect. For example, in the case of isotropic scattering, by mistaking the phase function for *PDF* of θ , one may think that θ is uniformly distributed in $[0, \pi]$. However, it is the phase function (*PDF* of solid angle Ω) that is uniform and θ is distributed as $f(\theta) = P(\theta) \cdot \frac{\sin(\theta)}{2}$ [16, 18, 19].

In order to obtain a better understanding of multiple scattering mechanisms, one may wish to use an approximate phase function that can be easily characterized and is more convenient to use than Mie series [6]. The most popular approximate phase function is the Henyey–Greenstein (HG) scattering function [20] and is given by:

$$P(\theta) = \frac{(1 - g^2)}{2\pi(1 - 2g \cos \theta + g^2)^{3/2}}.$$

This approximate phase function is completely characterized by the parameter g , which is the average value of $\cos(\theta)$, and is also called the asymmetric parameter. As we see later in this chapter, the asymmetric parameter is of great importance since it determines the convergence rate of the Markov chain associated with density evolution of polar angle, θ . For most clouds, $\overline{\cos(\theta)} \approx 0.85$, and for more forward-scattering clouds, $\overline{\cos(\theta)} \approx 0.995$ as stated in [8].

Mie theory, although very useful in describing single scattering phenomena, does not provide sufficient insight into the multiple scattering problems by itself. Instead, one

should exploit this theory along with a powerful tool to explain laser beam propagation in a multi-scattering medium. MCRT is one such method that is based on brute-force tracking of all photons' trajectories in a three-dimensional space. As a result, it is quite numerically intensive.

We describe the new approach of channel modeling, which is based on direct extraction of state transition matrices associated with standard Markov modeling from the MCRT computer simulations programs. We can model photon trajectory as a random walk using Markov chain, and investigate evolution of angular beam spreading. If we consider a laser beam to be composed of a large body of coherent photons traveling in the same direction, by tracking these photons, one can account for laser beam propagation in a multiple scattering medium. The MCRT algorithm implements this brute-force tracking. Figure 5.2 shows the geometry of multiple scattering. Note that, the notion of photon simply refers to a packet of energy (or a very small portion of the beam) in the MCRT context and we are not dealing with quantum mechanical aspects of photons.

To describe the MCRT algorithm and our Markov chain model, first we clarify our notations and terminologies. In MCRT, the k th state of a photon, i.e. a photon's position and directions in a global coordinate system right before the $(k + 1)$ th scattering event, is specified by $[x_k, y_k, z_k, \theta_k, \varphi_k]$, where (θ_k, φ_k) represent the traveling direction in global spherical coordinates and $(x_k, y_k, z_k,)$ stands for position in global Cartesian coordinates. Note that, direction can also be expressed by directional cosines (μ_x, μ_y, μ_z) in the global Cartesian coordinate system [16]; however, we choose the former notation because it requires only two variables and fits better in our simplified Markov chain model. After the $(k + 1)$ th collision, a photon is deflected with angles (θ, φ) in local coordinates, (i.e. pre-collision traveling direction is the local z axis) and travels a distance d , until it goes through the next scattering event. Hence, the $(k + 1)$ th state of this photon

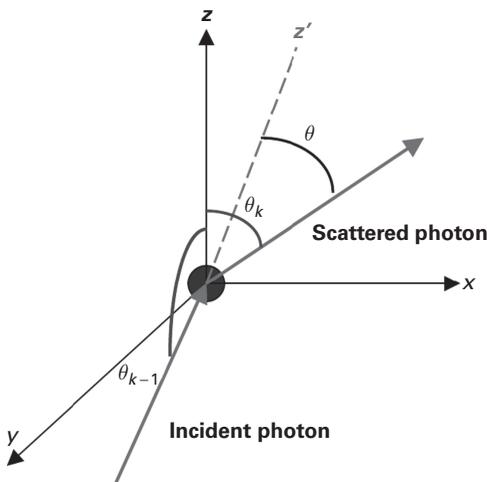


Figure 5.2 Relationship between θ_{k-1} , θ , and θ_k .

can be determined using its previous state and the new input to the system, which is the 3-tuples (d, θ, φ) . We have dropped the index and prime of this 3-tuple for simplicity, as we need the indices to specify the states. Note that, some authors [8] use a different notation in that they index the input 3-tuples (d, θ, φ) measured with respect to the local coordinate system, however, in order to define the Markov chain model, we choose to index the states, which are measured with respect to the global coordinate system. MCRT calculates channel parameters such as angular, spatial, and temporal dispersions using statistics of photons reaching the receiver plane in a post-processing stage.

In the Markov chain of MCRT, variables associated with the position of a photon in a three-dimensional space are neither finite, nor countable. However, if we limit our attention to the photons' direction in spherical coordinates, and quantize θ , a finite state Markov chain is obtained. Note that, due to symmetry, the azimuth traveling direction, φ , is always uniformly distributed in $[0, 2\pi]$. Furthermore, if it is assumed that the laser beam is traveling in a homogeneous medium, the phase function will not change from one state to the next. Thus, a Markov chain with a time-invariant state transition matrix can represent the angular distribution evolution of a laser beam, while it is traveling through a multiple scattering medium.

We claim that the angular dispersion, defined as the average cosine of the incidence angle on the receiver plane, can be calculated by modeling the photon trajectory in a 3D space by a random walk. Moreover, this method provides us with the complete angular distribution, rather than just the moments. Also, it is superior to MCRT since due to its analytical nature, it is more tractable. Furthermore, in MCRT, a large number of photons are sent into the scattering medium with the hope of finding the distribution of photons on the receiver plane. Hence, to account for all possible paths and angles, a large amount of processing is required that may not be desirable. While the computational complexity of MCRT is very high and computer simulation programs need a large execution time, this analytical method can provide us with results for angular distribution through simple matrix multiplications. The details of the proof can be found in reference [3].

Now, we can form a matrix with its rows corresponding to incident angles and its columns corresponding to scattering angles. This matrix can be used as the state transition matrix of the Markov process. Equation (5.3) illustrates this matrix:

$$P = \begin{bmatrix} P(\theta_k = 0 | \theta_{k-1} = 0) & \dots & P(\theta_k = \pi | \theta_{k-1} = 0) \\ \vdots & \dots & \vdots \\ P(\theta_k = 0 | \theta_{k-1} = \pi) & \dots & P(\theta_k = \pi | \theta_{k-1} = \pi) \end{bmatrix}. \quad (5.3)$$

If one wishes to know distribution of a photon's direction after the k th scattering event, knowing that it was initially traveling in the z direction, one has to calculate

$$P_k(\theta) = (P^K)^T \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix}. \tag{5.4}$$

The first row of the state transition matrix is $P(\theta) \sin(\theta)/2$. Hence, we can find the phase function resulting from k scattering events from the first row of the k th power of the state transition matrix, simply by taking out the $\sin(\theta)/2$ factor.

Figure 5.3 shows the state transition matrix for cumulus cloud at a wavelength of $1.55 \mu\text{m}$. We have calculated the state transition matrix with a resolution of $\pi/300$. In other words, P is a 300×300 matrix. We observed that by increasing the resolution to $\pi/1000$, our simulation results do not change. The forward scattering property of cumulus cloud is clear from this picture. That is, the state transition matrix is very close to an identity matrix. Figure 5.4 shows P^{15} which clarifies transition probabilities at an optical thickness value of 15. Here, the value 15 is chosen since for this value, there are less line-of-sight (LOS) photons, and most of the photons that may reach the receiver have gone through multiple scatterings. From Figure 5.4, one can see that after 15 scattering events, the phase surface is very close to the shape of $\sin(\theta)/2$. This corresponds to an isotropic scattering.

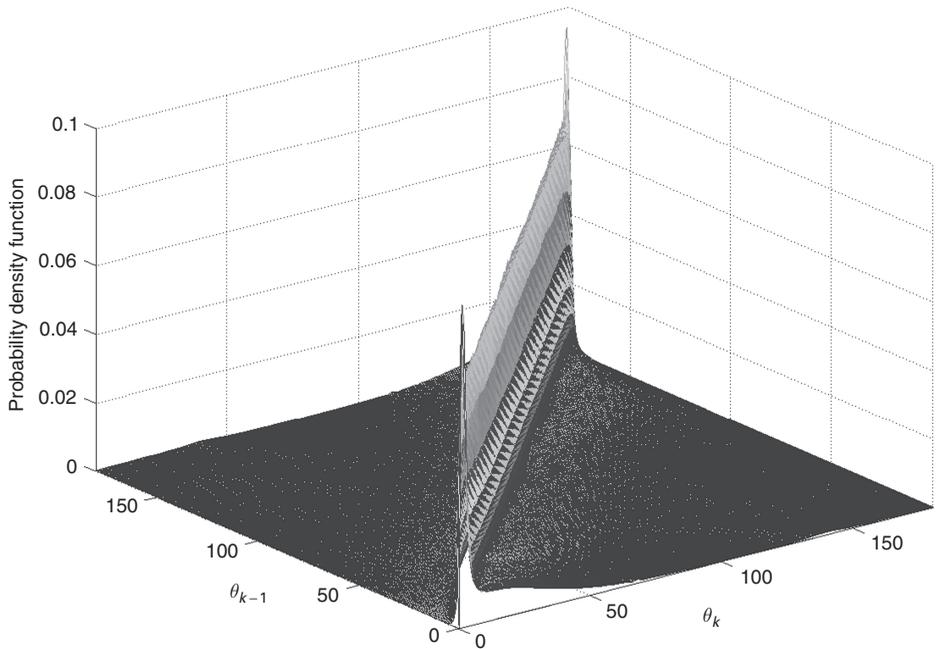


Figure 5.3 State transition matrix of cumulus clouds at a wavelength = $1.55 \mu\text{m}$.

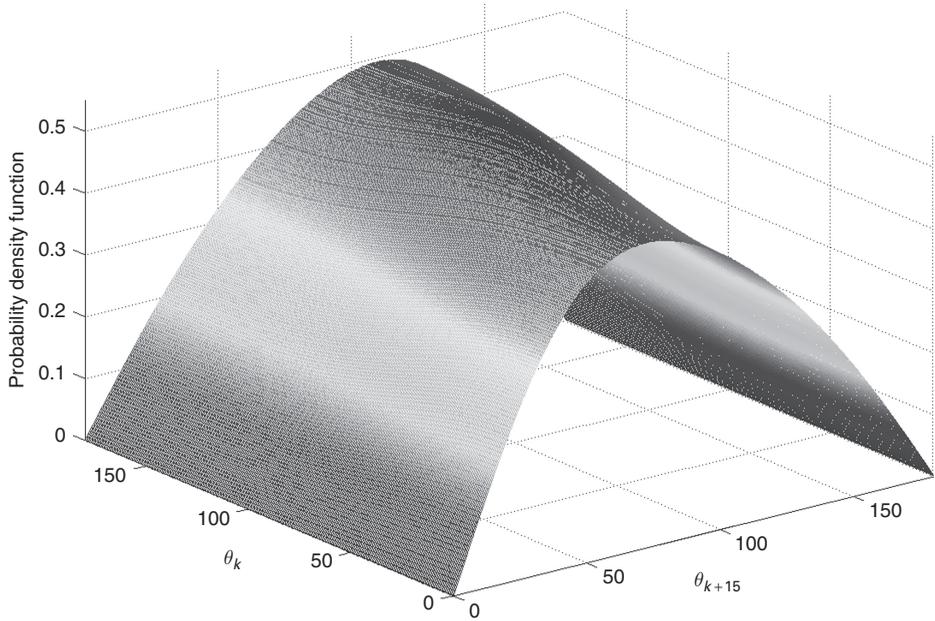


Figure 5.4 State transition matrix of cumulus cloud, raised to 15th power, at a wavelength = 1.55 μm .

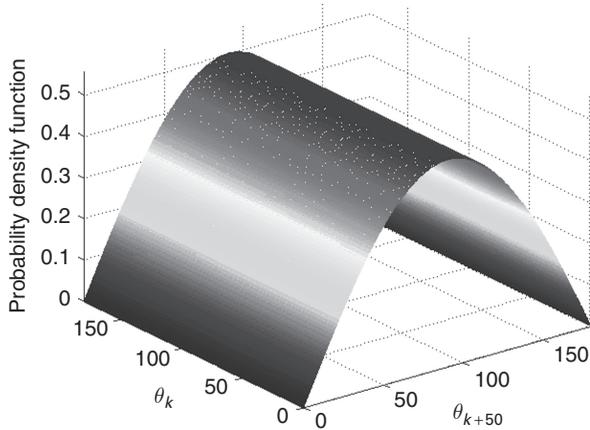


Figure 5.5 State transition matrix of cumulus cloud, raised to 50th power, at a wavelength = 1.55 μm .

Figure 5.5 shows P^{50} , that is, the phase surface after 50 scattering events. This figure clearly illustrates isotropic radiation after 50 scattering events. It can be inferred that after 50 scattering events, irrespective of the initial incident angle value, the scattering angle is distributed as $\sin(\theta)/2$. This is consistent with Bucher's [5] observation of uniform brightness of the cloud bottom.

5.3 Modeling results and eigen analyses

As mentioned earlier, application of Markov chain in MCRT cloud modeling is a shortcut for calculating the angular distribution of energy in space for any arbitrary optical thickness. In this section, we compare the results of our Markov chain model with those of MCRT and another analytical method, which we call the moment technique [8]. To make a fair comparison, we note that while both the MCRT and Markov chain model produce a complete distribution, the moment technique only comes up with the first two moments. Furthermore, MCRT can provide us with the distribution of photons on the receiver plane. However, both the Markov chain model and moment technique generate angular distribution in a three-dimensional space, i.e. on a sphere. Hence, to compare the Markov chain model with MCRT, we should only consider the forward part of the distribution, i.e., $0 < \theta < \pi/2$. Moreover, we should take into account the projected area correction factor of $\cos(\theta)$ for mapping from a 3D distribution on a sphere onto a 2D distribution on the receiver plane [6].

Given a photon at the cloud exit plane has been subject to scattering k times over a cloud length L , its angular distribution would be $P_k(\theta)$ as in Eq. (5.4). To calculate the unconditional angular distribution, we note that the probability of a photon undergoing exactly k scatterings over L is Poisson distributed with a mean τ , where τ is the optical thickness, as defined earlier. That is:

$$P(K(\tau) = k) = \frac{\tau^k}{k!} e^{-\tau}. \quad (5.5)$$

Hence, the angular distribution for this optical thickness value is:

$$P(\theta) = \sum_{k=0}^{\infty} \frac{\tau^k}{k!} e^{-\tau} \cdot P_k(\theta). \quad (5.6)$$

Using Eq. (5.6), angular distribution is calculated for optical thickness values of 1 to 15. Figure 5.6 shows the cumulative distribution function (CDF) of incident angle obtained from the Markov chain model and MCRT, where we have applied the above-mentioned measures in order to make the comparison fair. From Figure 5.6, one can see that the curves obtained from the Markov chain model and MCRT are quite close in numerical values. Furthermore, as optical thickness value increases, angular dispersion increases. One important measure of angular dispersion is the average cosine of incident angle, $\overline{\cos(\theta)}$. Variance of cosine of incident angle is also considered as a measure of angular dispersion in the literature. The moment technique provides us with the first two moments of $\cos(\theta)$ in a three-dimensional space. However, MCRT provides the complete distribution on the receiver plane. Then, one can find the first two moments of $\cos(\theta)$ using this distribution. The Markov chain model, in its original shape, produces the same mean and variance as those of the moment technique. However, as mentioned earlier, it can be truncated and modified to provide the distribution on the receiver plane.

Finally, the time required for each method to generate the results in different optical thickness values is tabulated in Table 5.1. Computer programs were executed on a 3.40 GHz Pentium-4 CPU, with 2.00 GB of RAM. It is clear from this table that

Table 5.1 Running time of three methods for different optical thickness values.

| Optical thickness | MCRT | Markov | Moments |
|-------------------|--------|--------|---------|
| 1 | 656 s | 33 s | 0.06 s |
| 5 | 767 s | 33 s | 0.06 s |
| 10 | 948 s | 33 s | 0.06 s |
| 15 | 1212 s | 33 s | 0.06 s |

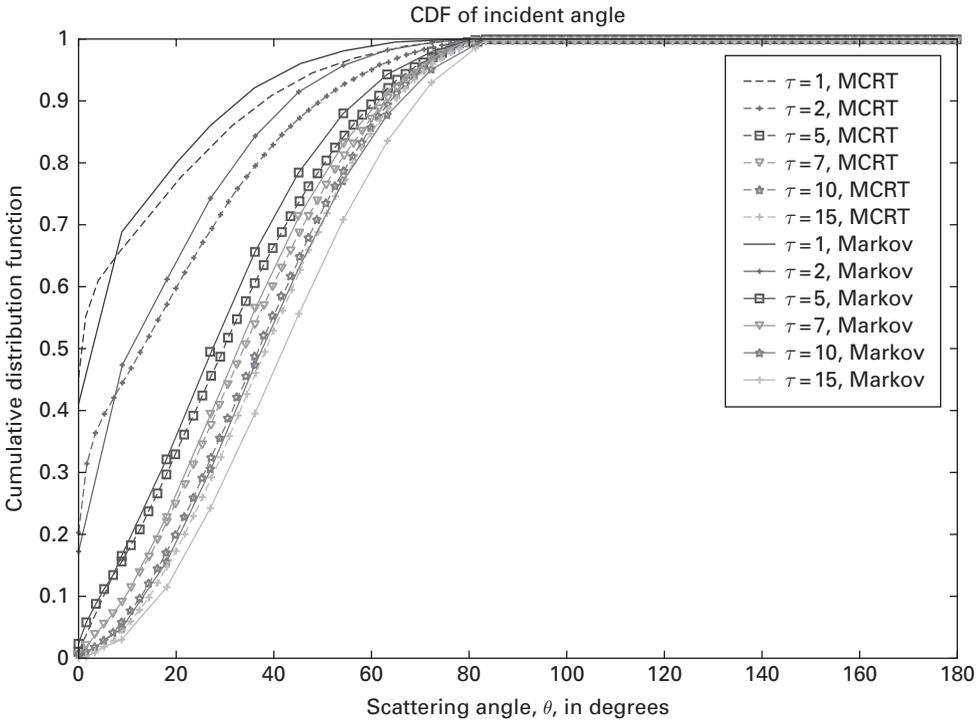


Figure 5.6 Comparison of *CDF* curves for angular distribution.

the computation time for MCRT increases with the optical thickness value. However, neither the Markov chain model, nor the moment techniques show much variability in the computation times, for different optical thickness values. We also note that the execution time required for the Markov chain model is significantly less than MCRT but much more than moment technique. However, moment technique gives only the first two moments, and not the complete distribution. Thus, the Markov chain model is able to demonstrate a more complete picture.

Note that, running time of MCRT increases exponentially with the optical thickness of the scattering medium. For large optical thickness values, the running time will no longer be reasonable. Furthermore, running time of MCRT also depends on the number of photons, the positions and directions of which at each scattering location have to be

recorded. Again, by increasing the number of these photons, running time increases. The program is terminated when the number of received photons is large enough so that the statistical variations at receiving plane become negligible.

The main point of this chapter is not to substitute (or compete with) the comprehensive Monte Carlo method by the less time-consuming Markov chain model. We are seeking new insight into the problem of laser beam propagation in scattering media. The trend of how fast a collimated laser beam diverges into a completely diffuse light source is more easily observed using the proposed Markov chain model. Also, the concept of eigenvalues of this Markov chain process is better understood this way. These are critical facts from a communications point of view; pulse broadening in space and time is and will continue to be the major limitation for high-speed communications in scattering media.

The state transition matrix, P , has no zero entries, and hence is regular. In other words, it is possible to go to all the states from any arbitrary state. It is a well-known fact that for a regular Markov chain, as n approaches infinity, $P^n \rightarrow \Pi$, where Π is a matrix of the form $[\nu, \nu, \dots, \nu]$, with ν being a constant vector. From the previous discussions, we see that this is true about the state transition matrix of our Markov chain P , and as n increases, all the rows of the state transition matrix become identical and proportional to $\sin(\theta)/2$. Now, the question is whether it is possible to predict ν , and thus Π , without using the limits and the answer is affirmative. In fact, Π satisfies the equation:

$$\Pi P = \Pi. \quad (5.7)$$

Notice that, from the Perron–Frobenius theorem [21], ν is the first left eigenvector of P , corresponding to the unique largest eigenvalue, $\lambda_0 = 1$. By examining the first left eigenvector of matrix P , we realize that it is proportional to $\sin(\theta)/2$. Hence, we could have predicted the diffuse behavior of light at the bottom of cloud, just by looking at the left eigen vector of the state transition matrix, to start with.

Convergence of P^k elements means that it becomes more and more difficult to guess k , from p_{ij}^k (the element in the i th row and the j th column of P^k). That is, the chain forgets the length of its history [21]. The fact that limit of P^k has identical rows suggests that the Markov chain forgets the initial direction. When a Markov chain converges to this steady-state, the traveling direction of photons becomes rather isotropic, as opposed to forward-scatter in the initial steps. This suggests that the laser beam is spatially diffused and irrespective of initial traveling direction, photons escape almost uniformly from all boundaries of the cloud. In this case, spatial confinement of transmitted energy is no longer preserved and loss (mostly attributed to scattering) is rather large.

In some applications, due to eye-safety regulations, there is restriction on increasing the transmit power level beyond a certain maximum. Under such circumstances, the receiver may not see much forward scatter (snake) photons and must resort to only the LOS (ballistic) photons, that is, the non-diffused part of the intensity, which is attenuated according to the Beer–Lambert law as:

$$I_{coh} = I_0 e^{-\tau}. \quad (5.8)$$

Needless to say, by using ultra-short laser pulses and cascade amplifiers, this component can be amplified, significantly. In Eq. (5.8), τ is the optical thickness of cloud, as defined earlier.

The number of steps required for the Markov chain to converge to the equilibrium state is of great importance since it determines the depth up to which the laser beam can penetrate before becoming spatially diffused. From [3], one can see that:

$$\overline{\cos(\theta_k)} = \overline{\cos(\theta_{k-1})} \times \overline{\cos(\theta)} = g \overline{\cos(\theta_{k-1})} = g^k. \quad (5.9)$$

In other words, the asymmetric parameter, g , determines the convergence rate of Markov chain. On the other hand, from the Markov chain theory [22], we know that the Second Largest Eigenvalue Modulus (SLEM) determines the ‘‘Mixing Rate’’ of a Markov chain. In other words, the smaller the SLEM is, the faster the spatial memory-loss happens. Hence, ‘‘Mixing-Time’’ of a Markov chain is given by:

$$T = \frac{1}{\log\left(\frac{1}{\lambda_*}\right)} \quad (5.10)$$

where T is the number of steps over which deviation from equilibrium state decreases by a factor e , and λ_* is the SLEM. By examining the eigenvalues of P (state transition matrix of cumulus cloud) we realize that the second largest eigenvalue is indeed the asymmetric parameter g .

Table 5.2 lists the SLEM for Markov chains associated with the Henyey–Greenstein (HG) phase functions [20] of different asymmetric parameters as well as full Mie series phase function for different types of clouds. This table also contains the mixing time and the average distance between two successive scattering events for each cloud. Average distance, D_{ave} , between two successive scattering events is used to convert optical thickness of a specific scattering medium to its physical thickness and vice versa. It also corresponds to the inverse of scattering coefficient, expressed in km^{-1} , which can be obtained by substituting the cloud particle size distribution in equations extracted from Mie theory for poly-dispersed extinction coefficient [17, 23].

Table 5.2 Asymmetric parameter and SLEM for different clouds at a wavelength = 1.55 μm .

| Cloud | SLEM | g | Mixing time | D_{ave} |
|----------------------|------|------|-------------|-----------|
| Thin Cirrus | 0.82 | 0.82 | 5 | 11.3 km |
| Cirrus | 0.87 | 0.87 | 7 | 984 m |
| Alto Stratus | 0.83 | 0.83 | 5.4 | 10.5 m |
| Nimbo Stratus | 0.85 | 0.85 | 6.2 | 12.3 m |
| Stratocumulus | 0.82 | 0.82 | 5 | 26.5 m |
| Stratus | 0.83 | 0.83 | 5.4 | 17.5 m |
| Cumulus | 0.85 | 0.85 | 6.2 | 7.5 m |
| Low altitude haze | 0.73 | 0.73 | 3.2 | 817.8 m |
| Medium altitude haze | 0.78 | 0.78 | 4 | 238.8 m |

From Table 5.2 we see that the second largest eigenvalue is equal to g for all of these phase functions. That is why asymmetric parameter is important in calculating the moments of multiple scattering and channel parameters [6, 8]. Since angular, spatial, and temporal distributions of energy are interrelated, isotropic angular distribution suggests that spatial distribution in transverse coordinates (x and y directions) is Gaussian. Furthermore, spatial memory-loss of Markov chain associated with angular distribution after several scattering events implies that the entire Monte Carlo Markov chain (MCMC) has converged to the equilibrium state. This suggests that the pulse is so broadened in space and time that use of equalization on the pulse might be difficult, unless there is no constraint on the optical transmitted peak power level.

Examining the parameters of Table 5.2, we realize that mixing time is rather short for clouds and haze and a large link margin is necessary, in order to have reliable communications through clouds of optical thickness values much larger than 10 to 15. However, this rather small optical thickness value may translate into a long physical thickness for clouds such as thin Cirrus due to the long average distance between two successive scatterings for these clouds [17]. Note that, all the results presented in this chapter assume a homogeneous body of clouds. In reality, as a waveform moves through clouds, the dynamics are far more rapidly varying. Hence, the predictions in this section are on the conservative side.

5.4 Equalization related issues

A cloudy or foggy optical channel is usually characterized by severe intersymbol interference (ISI) at high data rates, due to multi-scattering. In Section 5.3, we investigated the feasibility of high data rate FSO transmission through a cloud channel. Results suggest that beyond a certain cloud thickness, pulses are broadened in space and time to the extent that use of equalization to counter intersymbol interference (ISI) might be difficult. In this section, we will try to examine the level of challenge and seek ways of countering the severe channel ISI by various equalization techniques. The subject has been of research interest for some time now [24–29]. As it should be clear, unlike RF communications where spectrum is costly and scarce, the availability of wide bandwidth light beams enable reliable wireless communications at very high data rates with enhanced security via direct beams. One way to enhance the link performance in ISI is through linear minimum mean-squared error (LMMSE) equalization or through decision-feedback equalization (DFE).

In this section, as an alternative, we propose an FSO system for through-cloud communications consisting of a channel shortening equalizer (CSE) [30–32] followed by a Viterbi equalizer [33]. Time-domain equalization (TEQ) was originally proposed to reduce the length of the discrete channel impulse response (CIR) in channels such as digital subscriber loop (DSL) or cable. In FSO through clouds, in the same sense, use of TEQ can be justified in the context of heavy scattering dispersion. The ensuing Maximum-likelihood sequential detection (MLSD) scheme is known to be the optimal equalization algorithm, compared with linear minimum mean-squared error (LMMSE)

equalization or Minimum-mean-squared error (MMSE)-DFE, but its complexity constraint and memory requirements indeed require a short channel length, thus further justifying our use of shortened channel TEQ. On the other hand, since the use of TEQ at the receiver after sampling causes colored noise, we cannot use a Viterbi equalizer in the strict sense, even though this problem is bypassed in the existing literature by assuming that we still have white Gaussian noise after the TEQ. This can be solved by placing the TEQ at the transmitter, and can be understood as a discrete-equivalent to the ultra-short FSO scheme proposed in the analog domain in [34].

In this section, we will use two well-known CIR shortening algorithms, maximum shortened signal-to-noise ratio (MSSNR) [30, 31] and MMSE [32], followed by the MLSD algorithm to evaluate error performance of FSO links over an air-to-ground through-cloud channel. FSO channel model reported in [2] will be adopted. Also, in order to adopt the existing channel shortening TEQ solutions to our case of TEQ-in-the-transmitter configuration, we show how the existing solutions' framework can be further adjusted to fit the problem at hand. These approaches have recently been investigated extensively in [29].

Figure 5.7 shows an optical communication system model based on intensity modulation/direct detection (IM/DD). Wireless optical links transmit information by modulating the instantaneous optical intensity $I(t)$ in response to an input electrical current $x(t)$, with two constraints in non-negativity and power satisfied, as in Eqs. (5.11) and (5.12).

$$I(t) \geq 0, \forall t \in R. \tag{5.11}$$

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T I(t) dt \leq P. \tag{5.12}$$

The modulated signal $x(t)$ can be expressed by Eq. (5.13), where x_{DC} is a DC bias term added by bias adjustment (BA) block to ensure non-negativity constraint of Eq. (5.11):

$$x(t) = \sum_k I_k g_T(t - kT_s) + x_{DC}, \quad T_s : \text{symbol period}, I_k: [0, 1], \tag{5.13}$$

$g_T(t)$: a transmit pulse shaping function; square-root raised cosine filter assumed.

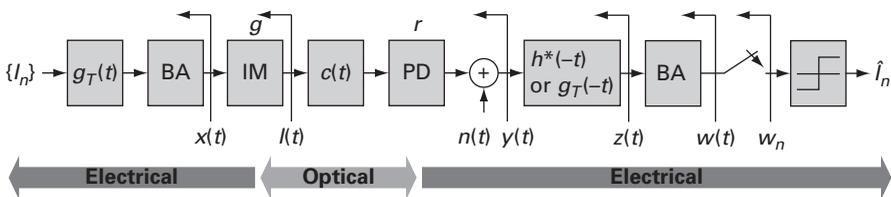


Figure 5.7 A system model for the optical communication system based on IM/DD. IM: intensity modulator, $c(t)$: channel model, $n(t)$: electrical additive noise, BA: bias adjustment, PD: photo-detector.

Denoting the optical gain of electrical-optical conversion by g , the optical intensity modulated signal can be represented by Eq. (5.14):

$$I(t) = gx(t). \quad (5.14)$$

Referring to the impulse response due to multi-scattering by $c(t)$, the received signal at the receiver front end after photo-detection can be represented by a baseband electrical model as in Eq. (5.15). In our notations, for simplicity, we will assume $rg = 1$.

$$y(t) = rI(t) \otimes c(t) + n(t) = rgx(t) \otimes c(t) + n(t), \quad (5.15)$$

r: photo-detector sensitivity, $[A \cdot m^2/W]$, $n(t)$: AWGN.

Finally, using a receive filter $g_R(t) = g_T(t)$, $y(t)$ is filtered to $z(t) = y(t) \otimes g_R(t)$. The $z(t)$ can be mathematically derived in further detail as in Eq. (5.16), using definitions of Eqs. (5.17), (5.18), and (5.19). Alternatively, the equivalent discrete-time impulse response can be expressed by Eq. (5.20), yielding Eq. (5.21).

$$z(t) = y(t) \otimes g_R(t) = [x(t) \otimes c(t) + n(t)] \otimes g_R(t) = \sum_k I_k f(t - kT_s) + \alpha x_{DC} + v(t). \quad (5.16)$$

$$f(t) = g_T(t) \otimes c(t) \otimes g_R(t). \quad (5.17)$$

$$\alpha = \int (c(t) \otimes g_R(t)) dt. \quad (5.18)$$

$$v(t) = n(t) \otimes g_R(t). \quad (5.19)$$

$$f_k = f(t - \Delta) \Big|_{t=kT_s}, \quad \Delta : \text{synchronization delay}. \quad (5.20)$$

$$z_n = \sum_k I_k f_{n-k} + \alpha x_{DC} + v_n. \quad (5.21)$$

Most of the existing DSP techniques such as equalization depend on the autocorrelation property of transmitted symbols. In case of an OOK modulation which does not satisfy Eq. (5.22), one good way is to adjust the DC at the receiver.

$$E(I_n I_m) = \begin{cases} 0, & n \neq m \\ \sigma_I^2, & n = m \end{cases}, \quad I_n, I_m: \text{transmit symbols}. \quad (5.22)$$

At the receiver, the BA block can first work on Eq. (5.21) to remove the DC term αx_{DC} . Furthermore, assuming a frame-by-frame transmission of frame length FR ($\gg L$), and the finite-length discrete-time channel f_k , $k = 0 \dots L$, one can further adjust the DC level. In Eq. (5.23), the transmit symbol constellation is changed from $[0, 1]$ to $[1/2, -1/2]$. However, for some portions of the received frame, $\sum_{m_1}^{m_2} f_k I_{avg}$ is time-varying. Thus, the frame structure needs a preamble longer than $L + 1$ and a post-amble longer than L . One point to note is that I_{avg} in Eq. (5.23) was used as a part of the detection scheme with channel state information at the receiver (CSIR) assumed, whereas the DC term x_{DC} in Eq. (5.16) was used to ensure non-negativity condition at the transmitter.

$$u_n = \sum_{k=m_1}^{m_2} f_k I_{n-k} - \sum_{k=m_1}^{m_2} f_k I_{avg} + v_n = \sum_{k=m_1}^{m_2} f_k (I_{n-k} - I_{avg}) + v_n = \sum_{k=m_1}^{m_2} f_k D_{n-k} + v_n \quad (5.23)$$

where $D_k: [-1/2 \ 1/2]$, $I_{avg} = E(I_n)$ and

- (1) if $0 \leq n \leq L$, then $m_1 = 0, m_2 = n$, and $\sum_{m_1}^{m_2} f_k I_{avg}$ is time-varying
- (2) if $L + 1 \leq n \leq FR - 1$, then $m_1 = 0, m_2 = L$, and $\sum_{m_1}^{m_2} f_k I_{avg}$ is constant
- (3) if $FR \leq n \leq FR + L - 1$, then $m_1 = n - FR + 1, m_2 = L$, and $\sum_{m_1}^{m_2} f_k I_{avg}$ is time varying.

In short, unlike traditional OOK-based FSO systems, the modulation needs the non-negativity constraint of Eq. (5.11) on the optical signal path, and DC bias adjustment on the electrical signal path to facilitate DSP techniques, which hold as long as we send in a frame structure of long-enough length.

5.4.1 Review of channel shortening algorithms

Conventionally, equalizers can be categorized into linear equalizers (LE) such as zero-forcing (ZF) equalizer or MMSE equalizer, and nonlinear equalizers (NLE) such as DFE or MLSD equalizer. DFE suffers more in low SNR due to error propagation than LE. MLSD is known to be optimal, but its complexity grows exponentially with the length of the delay spread, and is thus prohibitively costly in channels of long delay spreads. The philosophy of the conventional linear equalization is to minimize MSE defined as $J = E[I_k - \hat{I}_k]^2$. For MMSE equalization to yield a good error performance, under the constraint of equalizer tap numbers, it expects an equalized CIR that is close to the ideal CIR, or, a Dirac delta function $\delta(t)$, such that the Euclidean distance between the transmitted symbol and the equalized symbol is somehow minimized.

As an alternative, we focus on the channel-shortening equalizer in this work (Figure 5.8). This was originally suggested as a way to relieve the computational burden of MLSD equalization [30]. In the 1990s, channel-shortening TEQ was further applied to time-domain equalizations of discrete multi-tone (DMT) multi-carrier modulation [31] or orthogonal frequency division multiplexing (OFDM) systems. Both for MLSD based single-carrier systems and for multi-carrier modulation based systems. The idea here is not to bring back the CIR gradually close to a single tap Dirac delta function,

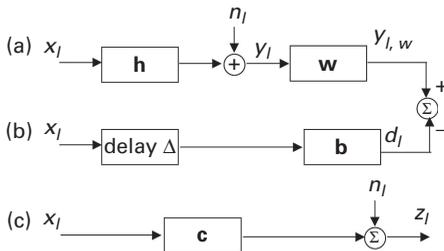


Figure 5.8 Block diagram for MMSE-TEQ: (a) a discrete-time model with TEQ (b) delay Δ and TIR **b** (c) equalized channel, l : index of time instants, x_l : transmit sequence from the transmitter, **h**: channel impulse response, n_l : Gaussian noise, y_l : received sequence, **w**: channel shortening equalizer, Δ : delay in discrete-time domain, **b**: target impulse response, **c**: equalized channel impulse response (or shortened channel impulse response).

but rather to shorten it to a few taps, such that either multi-carrier-based systems or MLSD can handle this shortened channel impulse response (SCIR). One problem in channel shortening is that it causes colored noise. Originally, the MLSD algorithm was proposed under the assumption that we have a white Gaussian noise sequence. If we have a colored noise sequence, then depending on the correlation, we can still use the Viterbi equalizer, but the BER is no longer minimum, on the average. In the case of FSO link, we need TEQs of considerable lengths in order to shorten the CIR, rendering the use of Viterbi equalizer in colored noise far from optimal. If we place a whitening filter at the receiver, it would nullify what the TEQ had done. One way around this dilemma is to place the TEQ at the transmitter, while keeping the Viterbi equalizer at the receiver. For this configuration to work, we need to provide channel state information (CSI) from the receiver to the transmitter, but until we have established a secure FSO link, we can benefit from a separate RF link to deliver CSI. A lengthy discussion of various optimization algorithms to perform these tasks has been provided in [29]. We now present some results applying these techniques.

5.4.2 Channel shortening algorithms: results

Here we will base the results on the channel models presented in [2], and we present some snapshots of the discrete channel model based on the single-gamma function. Since our interest is in application of shortening equalizers for low-to-mid optical thickness values, we consider τ values of 5, 7, 8, 9, 10, 11, 12, 13, 14 and 15, and present a snapshot corresponding to $\tau = 10$ in Figure 5.9. Our discrete channel snapshots consist of 5, 8, 16, 24, 34, 44, 55, 67, 81, and 95 taps for the aforementioned τ values, when 99.9% of the total energy was considered. We also mention that, since we approach FSO in the equalization context, we adopt a channel model with no attenuation considered, assuming that attenuation has been somehow compensated for by power control.

In TEQ design, first, it needs to be considered that the time span length T , the target impulse response (TIR) length $\nu + 1$, and the SCIR length give rise to computational complexity of TEQ and MLSD equalizer, respectively. Also, considering the MLSD stage following TEQ and its memory requirements, $\nu + 1$ should not exceed some reasonable length. In this context, we adopt an inequality constraint such as $T \leq 100$, $T \leq 200$, or $T \leq 400$, depending on the optical thickness value and an equality constraint of $\nu + 1 \leq 20$, and search through the quadratic MSE, or MSSNRn space for the candidate points consisting of T and $\nu + 1$. Second, even though we set $\nu + 1 \leq 20$, it does not always guarantee that the evaluated SCIR is of MLSD-equalizable length. In particular, in some cases when the CSE algorithm cannot afford to constrain the SCIR energy into $\nu + 1$ taps, (mostly encountered in extremely dispersive channel cases), we may end up with a major lobe of $\nu + 1$ taps, followed by one or more minor lobes after some distance from the main lobe. In this case, one cannot implement a Viterbi equalizer, even though the main energy of the SCIR falls within several main taps of $\nu + 1$ length. Thus, we numerically calculate the effective SCIR length by counting in only taps of 99.9% energy and mask out the MSE or maximum shortening-SNR (MSSNRn) space such that only MLSD-equalizable points can be considered. In this section, we

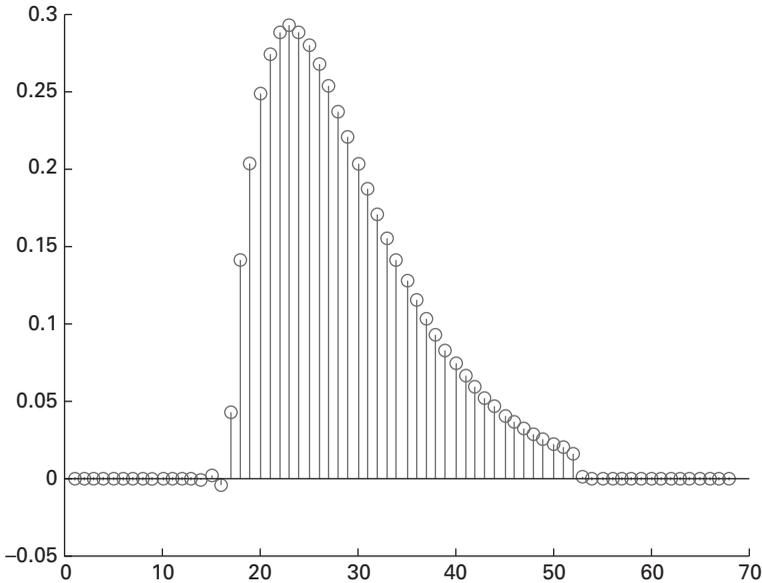


Figure 5.9 A discrete-time channel model at $\tau = 10$ with sampling interval of 0.18751 nsec (a) 3D MSE space (b) distribution of 99.9% effective length of SCIR (c) 3D MSE space with Viterbi-equalizable points.

arbitrarily pick only resulting SCIR values that do not exceed 14 taps in 99.9% energy criterion as Viterbi-equalizable SCIR values.

Figure 5.10 shows MMSE with unit energy criterion (UEC) TEQ performance in 3D MSE space, effective SCIR length space in 99.9% energy criterion, and in 3D MSE space with only Viterbi-equalizable points. One may make several observations. First, T naturally needs to be of some length longer than the CIR, and $\nu + 1$ need to be of some reasonable value, such that MSE can be enhanced. Second, in Figure 5.10(b), similar to our findings in Figure 5.10(a), the effective length distribution depends on both T and $\nu + 1$. If we have T and $\nu + 1$ of beyond some lengths, then the effective SCIR length increases linearly with $\nu + 1$, but not as much with T . If we take the Viterbi-equalizable points from Figure 5.10(b) and project the 2D space of the Viterbi-equalizable points onto the 3D MSE space of Figure 5.10(a), then we obtain Figure 5.10(c) where we verified that we could obtain 38.50% out of the 2D space of $\{T : T \leq 100\} \cap \{\nu : \nu + 1 \leq 20\}$, as the equalizable points. We can choose $T = 80$ and $\nu + 1 = 6$ as the MMSE-UEC-TEQ setting for average BER evaluations, for its advantages in both MSE and effective SCIR length. We further evaluated the performance at even higher optical thickness of $\tau = 12$ with the constraints of $\{T : T \leq 200\} \cap \{\nu : \nu + 1 \leq 20\}$ where, out of a total of 400 points, we could get 4 equalizable points at $(T = 80, \nu + 1 = 11)$, $(T = 90, \nu + 1 = 11)$, $(T = 100, \nu + 1 = 11)$, and $(T = 110, \nu + 1 = 11)$. For τ values of 13 and higher, we could not find any equalizable point in a similar space.

Likewise, similar tests were performed by the MSSNRn-TEQ algorithm, under the same constraints per τ as before. We can make several observations. First, Figure 5.11(a)

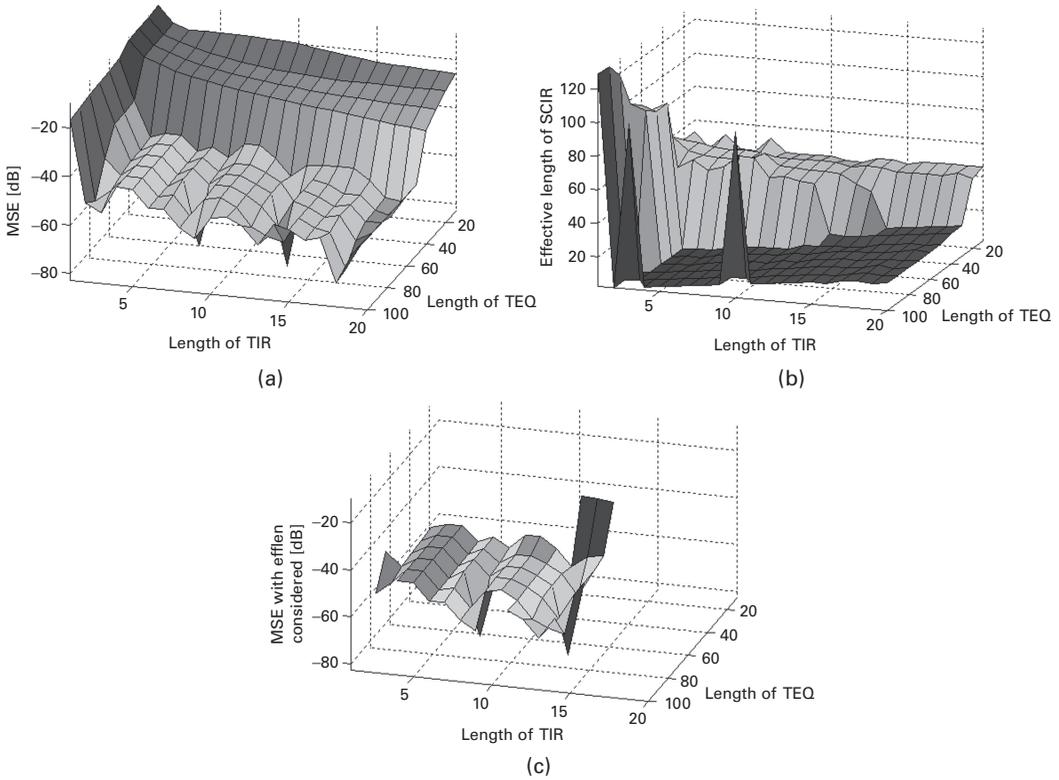


Figure 5.10 MMSE-UEC TEQ, $\tau = 10$, SNR = 30 dB, $T \leq 100$, $\nu + 1 \leq 20$. (a) 3D SSNR space (b) Distribution of 99.9% effective length of SCIR (c) 3D SSNRn space with Viterbi-equalizable points.

shows that the 3D SSNRn space shows dependence on both T and $\nu + 1$. Second, Figure 5.11(a) shows that, to some extent, complexity in terms of T and $\nu + 1$ are exchangeable, although the TEQ's complexity is linear with T whereas the Viterbi equalization's complexity grows exponentially with $\nu + 1$ in true sense. However, in so far as we employ TEQ of enough length and aim at SCIR of beyond a certain length, MSSNR space shows a flat response. Third, Figure 5.11(b) shows that in so far as T and $\nu + 1$ are larger than some values, SCIR's effective length grows linearly with $\nu + 1$ as expected, but almost independent of T . Fourth, Figure 5.11(c) shows that out of the total of 200 points, we obtained 50.3% Viterbi-equalizable points. We chose $T = 100$ and $\nu + 1 = 15$ as the point for BER evaluation.

At even more dispersive channel with $\tau = 15$, out of a total space of 400 points in $\{T : T \leq 200\} \cap \{\nu : \nu + 1 \leq 20\}$, we could obtain 67 equalizable points confined within $\{T : 105 \leq T \leq 200\} \cap \{\nu : 8 \leq \nu + 1 \leq 14\}$. At $\tau = 20$, however, out of 400 points in $\{T : T \leq 400\} \cap \{\nu : \nu + 1 \leq 20\}$, we could not find any equalizable point by the MSSNRn algorithm. To summarize, we can make several observations by considering Figure 5.10 and Figure 5.11. We can use both MMSE-UEC and the MSSNRn algorithms

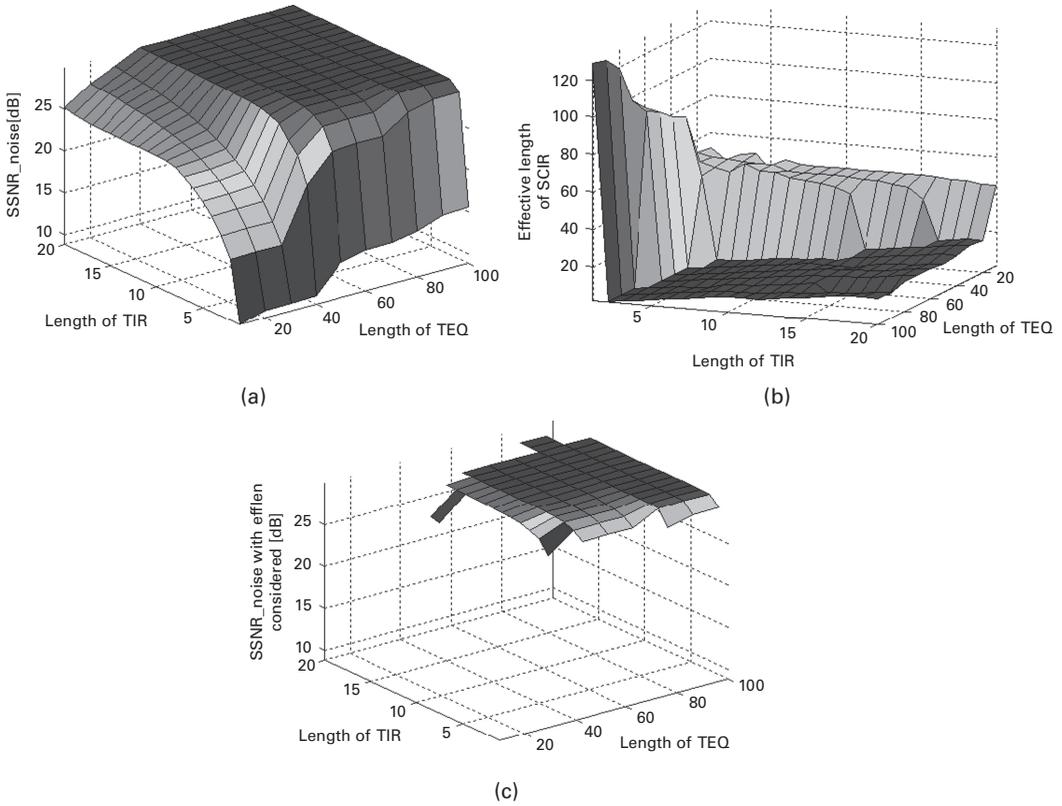


Figure 5.11 MSSNRn TEQ, $\tau = 10$, SNR = 30dB, $T \leq 100$, $\nu + 1 \leq 20$.

for moderate optical thickness values around $\tau = 10$. However, for $\tau = 15$, MSSNRn is a better choice, since MMSE-UEC does not yield a Viterbi-equalizable SCIR, whereas MSSNRn can still return SCIR's of moderate length.

We now incorporate a Viterbi equalizer at the receiver. As for the T and $\nu + 1$ values, we first analyzed the results of 3D MSE and MSSNRn spaces, and then chose one of the points with satisfactory performance and with short SCIR length. Figure 5.12 and Figure 5.13 show average BER curves for several τ values for the MMSE-UEC and the MSSNRn algorithms obtained from Monte Carlo simulations. We can make several observations. First, if we compare these at a BER of $1E-4$, even though both algorithms show similar performance at τ values of 5, the MMSE-UEC algorithm outperforms the MSSNRn algorithm at higher τ values of $8 \sim 12$. This is due to the fact that MMSE-UEC returned shorter SCIR than MSSNRn at these higher τ values. Second, however, since we cannot run MMSE-UEC for $\tau \geq 15$, we conclude that both algorithms have highs and lows. Third, due to limits in computational capability, verifications for smaller BER values than those depicted were not possible. In particular, for the τ value of 12, it is unclear whether the MSSNRn can still enhance the performance or not, by increasing the SNR value. However, at least, MMSE-UEC TEQ still works fine at a τ value of 12.

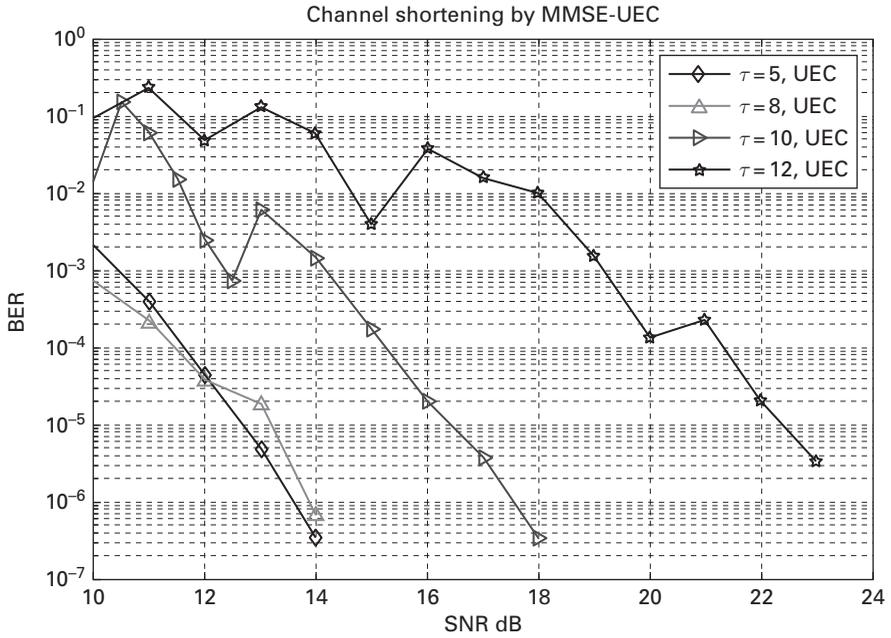


Figure 5.12 Error performance by MLSD with MMSE-UEC-TEQ. $\tau = 5$ case: $T = 20, \nu + 1 = 2$; $\tau = 8$ case: $T = 100, \nu + 1 = 2$; $\tau = 10$ case: $T = 80, \nu + 1 = 6$; $\tau = 12$ case: $T = 170, \nu + 1 = 8$.

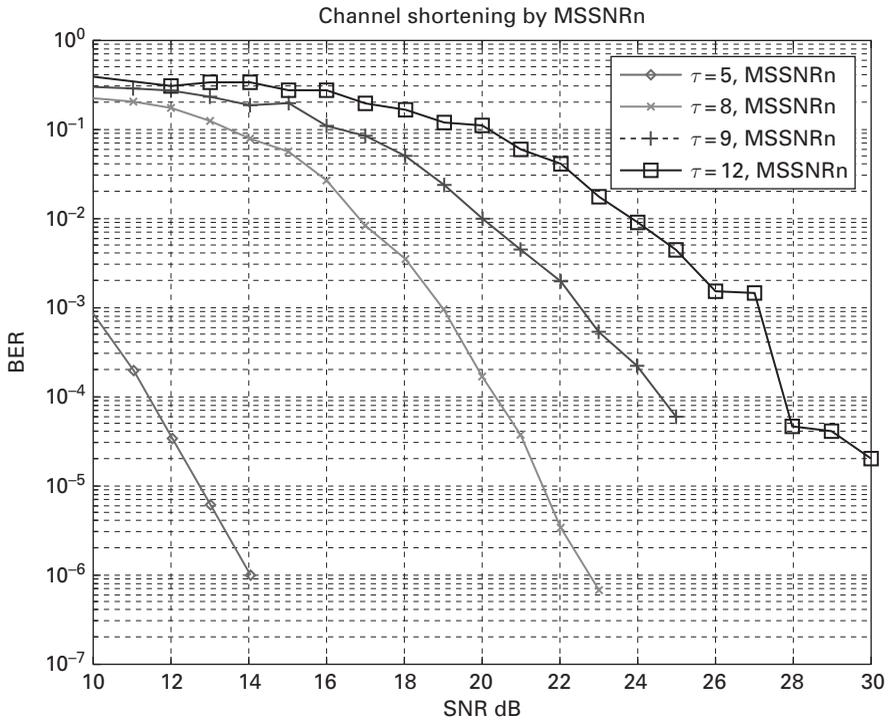


Figure 5.13 Error performance by MLSD with MSSNRn-TEQ. $\tau = 5$ case: $T = 40, \nu + 1 = 1$; $\tau = 8$ case: $T = 100, \nu + 1 = 10$; $\tau = 9$ case: $T = 100, \nu + 1 = 11$; $\tau = 12$ case: $T = 200, \nu + 1 = 16$.

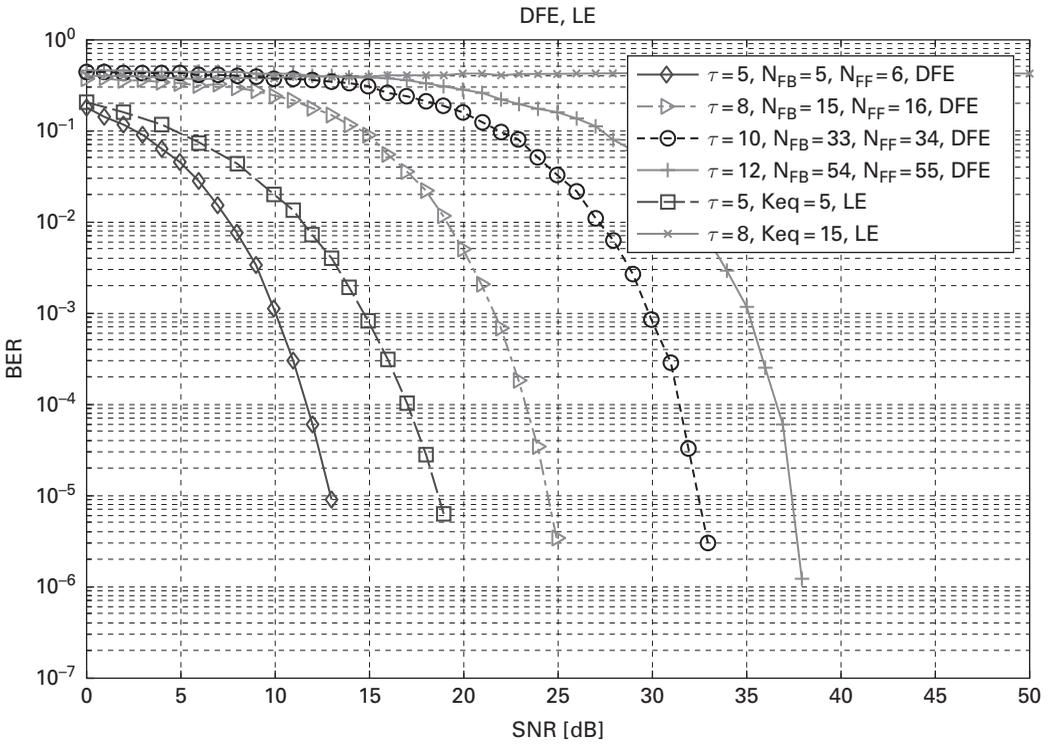


Figure 5.14 LMMSE equalization at $\tau = 5, 8$, and MMSE DFE equalization at $\tau = 5, 8, 10, 12$. We fixed N_{FB} (number of feedback taps) to be as long as the channel length, and set N_{FF} (number of feed-forward taps) correspondingly, such that performance of LE and DFE can be compared with equal tap numbers assumed.

We further compared these results with LMMSE and DFE BER curves shown in Figure 5.14. The figure shows that LMMSE does not work for τ values of 8 and higher. Also, one can see that unlike LE, one still can equalize reliably with DFE, even though a high SNR value is needed. However, if one makes a comparison at a BER of $1E-4$, both MSSNRn and MMSE-UEC outperform DFE curves. This enhancement comes at the cost of heavier computational intensity and memory requirements of the suggested transceiver architecture.

5.5 Summary and conclusions

To answer the fundamental question of feasibility of optical communications in a scattering medium, one has to accurately model atmospheric optical channel and estimate parameters such as angular, spatial, and temporal dispersions. As a collimated laser beam propagates through aerosols and cloud particles, it spreads after traveling some optical thickness value and approaches a steady-state condition of being nearly spatially diffuse. The rate of convergence to this steady-state determines the feasibility

of communications through various types of clouds and aerosols with different optical thickness values. Average distance between two successive scattering events is used to convert optical thickness of a specific scattering medium into its physical thickness and vice versa.

By directly applying the Markov chain model to angular distribution evolution of a laser beam in a scattering medium, and considering the fact that angular, spatial, and temporal distributions of energy are inter-related, the mixing rate of Monte-Carlo Markov Chain (MCMC) is found for different types of scattering media. Mixing time is introduced as the number of steps over which deviation from equilibrium state decreases by a factor e . This indicates the rate by which the energy distribution approaches being spatially isotropic.

Examining parameters of Table 5.2, we realize that mixing time is rather short for clouds and haze. That is, when coherent light is traveling through clouds, its spatial coherence rapidly degrades. However, it may take a much larger cloud length value for the coherence bandwidth of the medium to vanish. In the latter case, optical pulses broaden extensively in time and restoring them by signal processing might be difficult.

In this chapter, we also investigated the feasibility of applying TEQ combined with MLSD equalizer as a way to countermeasure severe channel ISI encountered in through-clouds FSO channels. We investigated the channel-shortening characteristics of two TEQ designs, in terms of SSNR_n and MSE versus TEQ length and TIR length. We also provided end-to-end mathematical formulations showing feasibility of electrical equalization on the FSO link. By Monte Carlo simulations, we conclude that, for low-to-mid range optical thickness values, the suggested DSP techniques can work relatively well to enhance BER performance. However, for mid-to-high range of optical thickness, we cannot rely solely on the TEQ, since the SCIR length distribution itself relies more on TIR length than on TEQ length.

Acknowledgment

A DARPA Grant sponsored by the US Air Force Research Laboratory/Wright-Patterson AFB Contract-FA8650-04-C-7114 and another DARPA Grant sponsored by the same under the Contract-FA8650-08-C-7850 and the Pennsylvania State University CICTR have supported this research. The author of this book chapter is grateful and acknowledges the contributions of his former PhD Students whose dissertations have all contributed to creating the knowledge embedded in this chapter. In particular works of Drs. B. Hamzeh, S. Lee, B. Wu, Z. Hajjarian and J. Fadlullah are acknowledged with utmost respect.

References

- [1] M. Kavehrad and B. Hamzeh, "Laser communication system using wavelet-based multi-rate signaling," *Proceedings of IEEE MILCOM*, vol. 1, pp. 398–403, Monterey-California, November 2004.

- [2] Z. Hajjarian, "MIMO optical communications and imaging through turbulent cloudy atmosphere," PhD dissertation, Pennsylvania State University, September 2009.
- [3] Z. Hajjarian, M. Kavehrad, and J. Fadlullah, "Analysis of wireless optical communications feasibility in presence of clouds using Markov chains," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 9, December 2009.
- [4] L. C. Andrews and R. L. Phillips, *Laser Beam Propagation through Random Media*, SPIE Press, second edition, 2005.
- [5] E. A. Bucher and R. M. Lerner, "Experiments on light pulse communication and propagation through atmospheric clouds," *Appl. Opt.* vol. 12, no. 10, pp. 2401–2414, October 1973.
- [6] E. A. Bucher, "Computer simulation of light pulse propagation for communication through thick clouds," *Appl. Opt.*, vol. 12, no. 10, pp. 2391–2400, October 1973.
- [7] N. S. Kopeika and J. Bordogna, "Background noise in optical communication systems," *Proc. IEEE*, vol. 58, no. 10, pp. 1571–1577, October 1970.
- [8] R. F. Lutomirski, A. P. Ciervo, and G. J. Hall, "Moments of multiple scattering," *Appl. Opt.*, vol. 34, no. 30, 7125–7136, October 1995.
- [9] G. N. Plass and G. W. Kattawar, "Monte Carlo calculations of light scattering from clouds," *Appl. Opt.*, vol. 7, no. 3, pp. 415–419, March 1968.
- [10] H. M. Gupta, "Space-time response of a cloud communication channel to an optical signal," *Opt. Quantum Electron.*, vol. 12, no. 6, pp 499–509, November 1980.
- [11] S. Arnon, D. Sadot, and N. S. Kopeika, "Simple mathematical models for temporal, spatial, angular characteristics of light propagating through the atmosphere for space optical communication: Monte Carlo simulations," *J. Mod. Opt.*, vol. 41, pp. 1955–1972, October 1994.
- [12] S. Arnon and N. S. Kopeika, "Free space optical communication: analysis of spatial widening of optical pulses for propagation through clouds," *Opt. Eng.*, vol. 34, pp. 511–516, February 1995.
- [13] K. Riley, D. Ebert, M. Kraus, J. Tessedorf, and C. Hansen, "Efficient rendering of atmospheric phenomena," *Proceedings of Eurographics Symposium on Rendering 2004*, pp. 375–386, June 2004.
- [14] F. G. Smith, *The Infrared & Electro-optical Systems Handbook: Atmospheric Propagation of Radiation*, Vol. 2, SPIE PRESS, 1993.
- [15] D. Deirmendjian, "Scattering and polarization properties of water clouds and hazes in the visible and infrared," *Appl. Opt.* vol. 3, no. 2, pp 187–196, February 1964.
- [16] A. Sawetprawichkul, P. Hsu and K. Mitra, "Parallel computing of three-dimensional Monte Carlo simulation of transient radiative transfer in participating media," American Institute of Aeronautics and Astronautics, 8th AIAA/ASME Joint Thermophysics and Heat Transfer Conference, St. Louis, Missouri, June 2002.
- [17] B. Y. Hamzeh, Multi-rate wireless optical communications in cloud obscured channels, Ph.D. Thesis, Pennsylvania State University, December 2005.
- [18] T. Binzoni, T. S. Leung, A. H. Gandjbakhche, D. Rüfenacht, and D. T. Delpy, "The use of the Henyey-Greenstein phase function in Monte Carlo simulations in biomedical optics," *Phys. Med. Biol.*, vol. 51, no. 17, pp. N313–N322, September 2006 .
- [19] T. Binzoni, T. S. Leung, A. H. Gandjbakhche, D. Rüfenacht, D. T. Delpy, "Comment on 'The use of the Henyey-Greenstein phase function in Monte Carlo simulations in biomedical optics'," *Phys. Med. Biol.*, vol. 51, no. 22, pp. L39–L41(1), November 2006.
- [20] L. Henyey and J. Greenstein, "Diffuse radiation in the galaxy," *Astrophys. Journal*, vol. 93, pp. 70–83, 1941.

-
- [21] E. Behrends, *Introduction to Markov Chains with Special Emphasis on Rapid Mixing*, Friedrick Vieweg & Son, 2001.
 - [22] S. Boyd, P Diaconis, and L. Xiao, "Fastest mixing Markov chain on a graph," *SIAM Review*, vol. 46, no. 4, pp. 667–689, December 2004.
 - [23] B. Wu, "Free-space optical communications through the scattering medium: analysis of signal characteristics," Ph.D. Thesis, Pennsylvania State University, December 2007.
 - [24] M. Kavehrad, "A countermeasure to improve outage performance of interference-limited microwave radio links," *Can. Elect. Comput. Eng. J.*, vol. 16, no. 1, pp. 13–18, January 1991.
 - [25] M. Kavehrad, B. Y. Hamzeh, "Laser communication system using wavelet-based multi-rate signaling," *Proceedings of IEEE MILCOM*, Monterey California, November 2004.
 - [26] X. Zhu and J. M. Kahn, "Free-space optical communication through atmospheric turbulence channels," *IEEE Trans. Comm.*, vol. 50, no. 8, pp. 1293–1300, August 2002.
 - [27] J. M. Kahn and X. Zhu, "Performance bounds for coded free-space optical communications through atmospheric turbulence channels," *IEEE Trans. Comm.*, vol. 51, no. 8, pp. 1233–1239, August 2003.
 - [28] H. Wu and M. Kavehrad, "Availability evaluation of ground-to-air hybrid FSO/RF links," *Int. J. Wireless Inf. Networks*, vol. 14, no. 1, March 2007.
 - [29] S. Lee and M. Kavehrad, "Free-space optical communications with channel shortening filter and viterbi equalizer," *Int. J. Wireless Inf. Networks*, vol. 16, no. 4, December 2009.
 - [30] P. J. W. Melsa, R. C. Younce, and C. E. Rohrs, "Impulse response shortening for discrete multitone transceivers," *IEEE Trans Comm*, vol. 44, no. 12, pp. 1662–1672, December 1996.
 - [31] R. Schur, J. Speidel, and R. Angerbauer, "Reduction of guard interval by impulse compression for DMT modulation on twisted pair cables," *Proc. IEEE Global Telecom. Conf.*, vol. 3, pp. 1632–1636, 2000.
 - [32] M. V. Bladel and M. Moenceclay, "Time-domain equalization for multicarrier communications," *Proc. IEEE Global Telecom. Conf.*, pp. 167–170, 1995.
 - [33] Forney, G., Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. 18, no. 3, pp. 363–378, May 1972..
 - [34] B. Hamzeh and M. Kavehrad, "Multi-rate RZ Communications for Dispersive Wireless Optical Channels," *Proceedings of SPIE Defense and Security Conference*, Orlando – Florida, April 2006.

6 Multiple-input multiple-output techniques for indoor optical wireless communications

Steve Hranilovic

The popular adage, “*a picture is worth a thousand words*” suggests that it is often easier to convey a large amount of information in parallel by way of an image rather than by serial transmission of words. Indeed, this is the underlying inspiration for a new class of indoor optical wireless (OW) channels. In these channels data are conveyed by coordinating the transmission of a series of transmitters and receivers. So called, multiple-input multiple-output (MIMO) OW systems employ a number of optical sources as transmitters and a collection of photodiodes as receivers. The arrangement of emitters and detectors varies depending on application and implementation technology.

The use of MIMO techniques in radio channels is a long-standing topic of interest. It has been shown that by increasing the number of antennae in a radio system that a *diversity* gain can be achieved, i.e., an improvement in the reliability of the system can be had. Alternatively, it has been demonstrated that multiple transmitting and receiving antennae can be used to provide a boost in rate, i.e. a *multiplexing* gain. In fact, a formal trade-off between diversity and multiplexing gain in MIMO radio channels has been established [1].

The OW MIMO systems discussed in this chapter differ fundamentally from earlier work on radio channels. In particular, only indoor OW channels are considered here and are free of multipath fading. In addition, the signalling constraints imposed by intensity-modulated/direct-detection (IM/DD) systems limits the direct application of theory from radio channels. Nonetheless, MIMO techniques have been applied to OW channels to yield improvements in reliability and to improve data rates. This chapter starts with a brief overview of the characteristics of indoor OW MIMO systems. Given that the application and available gains depend on channel architecture, the balance of the chapter considers the use of MIMO techniques in three main OW channel topologies: diffuse, spot-diffusing and point-to-point.

6.1 Indoor OW MIMO channel characteristics

6.1.1 Channel fundamentals

Optical wireless systems use optoelectronic components to convert signals from baseband to transmission wavelengths. Lasers and light-emitting diodes (LEDs) are typically used to convert an electrical current signal to a time-varying optical intensity. The transmitted signal of interest is an *optical intensity* which is the normalized power in units of Watts per steradian. The receiver accepts emissions inside of its field-of-view (FOV) and outputs a current proportional to the impinging optical power. These types of channels are termed intensity modulated/direct-detection (IM/DD) and are found in a vast majority of OW systems.

The quantity of interest at the transmitter is the instantaneous intensity, $I(t)$. Data are imposed on the instantaneous emitted optical power $x(t)$ which is the integration of $I(t)$ over the solid angle of the transmitter. At the receiver, a current $y(t)$ is output in proportion to the power on the detector surface. The received power is simply the integration of the irradiance, in Watts per area, over the detector surface.

The IM/DD channel imposes two primary constraints on all emitted signals. Since $x(t)$ represents a power it must be nonnegative, i.e.,

$$x(t) \geq 0. \quad (6.1)$$

International bodies impose safety limits on the permissible peak and average optical power to ensure skin and eye-safety. For modulated laser sources faster than 55 kHz, in most cases the average optical power constraint dominates the peak constraint [2]. A careful study of the eye-safety of many commercial OW devices can be found in [3]. As a result, a limit on the average emitted optical power must also be imposed for safety reasons, i.e.,

$$\lim_{u \rightarrow \infty} \frac{1}{2u} \int_{-u}^u x(t) dt \leq \mathcal{P} \quad (6.2)$$

where \mathcal{P} is eye-safety limitation of average optical power. These channel constraints must be considered in all signalling design for IM/DD OW systems.

The dominant source of noise in indoor OW systems is due to ambient lighting from fluorescent and incandescent sources. These devices produce emissions into the infrared which are quite detrimental to OW communications [4]. Fluorescent sources are modulated at tens to hundreds of kilohertz and have significant powers in the near-infrared band [5]. In addition, rooms with windows may allow sunlight to enter and further corrupt OW signalling. A conventional assumption is that the ambient light is of high intensity. This induces a Poisson shot noise with a large mean, which is well modeled by a Gaussian distribution [6]. As a result, in most indoor OW communication studies the noise is modeled as white, signal-independent, and Gaussian distributed [5].

6.1.2 Multiple transmitters

In indoor OW MIMO configurations, the arrangement of optical emitters into a single transmitter is an important feature. A straightforward method is to use a series of discrete

emitters to produce a time-varying optical intensity distribution in space. There are many options for this role including the use of vertical-cavity surface emitted lasers (VCSELs). In the area of optical interconnect, a 540-element array with 1080 VCSELs was built with a data rate of over 200 Mbps per pixel [7]. Specifically for OW communications, an array of integrated transceivers was constructed with a data rate of 155 Mbps per element using specially designed LEDs bonded directly to underlying CMOS drivers [8]. Recent studies have even investigated the integration of radio and optical transceivers together [9]. A nice overview of challenges in OW transceiver design can be found in [10].

If a large number of elements are required in close proximity, a spatial-light modulator (SLM) can be employed. In general terms, an SLM is able to produce a time-varying optical intensity image. An example of such a device is a liquid-crystal display which can have millions of emitters, however, each can be addressed at rates on the order of 100 Hz. Recent, micro-electrical-mechanical systems (MEMS) known as digital micro-mirror devices (DMD) integrate arrays of turnable mirrors with appropriate drive electronics. Commercially available DMD devices used in high-definition displays and projection equipment is capable of high-resolution output at switching periods as small as 8-16 μ sec [11, DLP® HDTV].

6.1.3 Receiver architectures

A key disadvantage of single-element OW receivers is that they are unable to reject multipath or ambient light interference. All signals in the FOV of a single element OW receiver are integrated over the surface of the detector. This is a key disadvantage of using such a receiver arrangement, however, single element receivers are inexpensive to implement.

Multiple element receivers have the advantage of sampling the spatial profile of the received optical intensity and rejecting portions which are corrupted by large amounts of interference. There are several options to implement such a receiver. One possibility is to use a series of discrete receivers and their optics and to orient them in different directions. These architectures are termed *angle diversity* receivers [12, 13]. Progress in this type of receiver is limited due to their large size and weight which is not appropriate for mobile devices. A more popular architecture is the use of a single imaging optical element (perhaps a lens) and an integrated array of receivers. These receivers are termed *imaging receivers* [14]. This architecture is significantly more compact than angle diversity receivers and permits the implementation of high-speed integrated transceivers [8]. Qualitatively, the lens does the operation of providing angular separation in the spatial field while the imaging array measures the intensity at each spatial point. Recent OW MIMO receivers have been constructed at rates of 5 Gbps with seven elements operating in the infrared band [15]. The high-speed imaging receiver is built in commercial CMOS technologies while the bonded photodetector array was produced in an InGaAs technology. In [16] an interesting multi-element receiver is presented which combines image acquisition and communications. In this way the same device can be used for imaging

to locate the transmitter and for data communications. Using these devices a 64×64 element receiver is made with 10 Mbps per channel with a total of 4 data outputs.

Imaging receivers have been considered and analyzed for a number of OW channel topologies [14, 17]. As the number of elements changes from 7 to 1000, the gain over a single-element receiver changes from 3.7 to 13 dB for a point-to-point link with $\text{FOV}=45^\circ$. For diffuse links and for spot-diffusing links the gains are on the order of 1–7 dB and 7–13 dB respectively [14]. It was also demonstrated that maximal ratio combining (MRC) and select best (SB) combining have nearly the same performance for a small number of elements < 20 , but this difference grows to about 2 dB gain as the number increases to 1000 [14].

Imaging receivers are also used in a variety of applications outside indoor OW communications. Arrays have been constructed for free-space optical communications [18, 19]. Another application area is for high-speed imaging. Imagers with frame rates near 10 kiloframes per second (kfps) with more than 10^5 elements have been constructed [20, 21]. Some other examples include a 312×260 imager working at Mfps rates for about 100 frames [22], and a smaller 12×12 imager operating at 4 Mfps for only 32 frames [23].

6.2 MIMO for diffuse OW channels

6.2.1 Diffuse channel model

In indoor settings, *diffuse* OW links rely on reflections of transmitted signals from objects in the room in order to establish a link. As a result, there is no strict requirement for a line-of-sight between transmitter and receiver [5, 24–26]. Although reflections of the optical signal from surfaces in the room may be either specular or diffuse, a vast majority are diffuse. This is due to the fact that the scale of surface roughness of typical surfaces in a room (e.g., walls, ceiling, etc.) is on the order of the wavelength of light. This generates diffuse reflections in which the angle of reflection is nearly independent of the incident angle [27].

The intensity distribution following a diffuse reflection can be well approximated using the Lambertian radiation

$$I_{\text{Lambertian}}(\theta, \phi) = \mathcal{P}_{\text{elem}} \frac{n+1}{2\pi} \cos^n(\theta),$$

where $I_{\text{Lambertian}}(\theta, \phi)$ is the optical intensity reflected at a polar angle $\theta \in [0, \frac{\pi}{2}]$ and azimuthal angle $\phi \in [0, 2\pi)$, as shown in Figure 6.1. The total reflected power by a surface element is denoted $\mathcal{P}_{\text{elem}}$ [W] and n is defined as the *mode number* of the Lambertian pattern [24, 28, 29]. Notice that the reflected Lambertian pattern is independent of ϕ due to symmetry in the azimuthal coordinate as in Figure 6.1.

Diffuse OW communication links do not direct their emissions but rather allow them to reflect diffusely from surfaces in the room. These unstructured emissions from the emitter bounce from floor, ceiling, walls, and room objects and provide coverage over

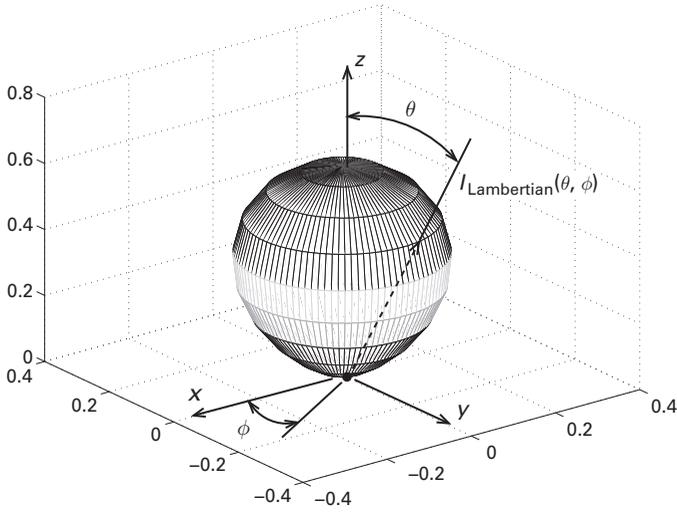


Figure 6.1 A Lambertian reflection pattern with $n = 3$ and normalized to unit power. The reflecting element is in the $x - y$ plane. (Reproduced with permission from [30].)

a wide area of the room. Thus, these links have been shown to be insensitive to both pointing and shadowing. Most importantly, this property allows a great degree of user mobility.

Indoor diffuse OW links are not impaired by multipath fading, as is the case with their radio counterparts. Fundamentally, this is due to the fact that the receiver photodiode is huge relative to the small wavelength of the light. In practice, the detector diameter is on the order of thousands of received wavelengths. Thus, a detector can have many variations in the received intensity which occur on the order of the wavelength of light. As a result of the large size of the detector relative to the wavelength, these local fades are effectively averaged out over the surface of the photodiode providing an inherent degree of spatial diversity against multipath fading [5].

Although multipath fading is not a major impairment, multipath induced *dispersion* is a major impairment in diffuse indoor links. The reception of multiple delayed and attenuated versions of the transmitted signal can be well modeled, as seen in Figure 6.2, as an effective lowpass channel. In addition, in order to collect sufficient power, receivers typically have wide FOV. In addition to collecting the transmitted signal, such receivers also collect a large amount of ambient light which, along with the large path loss, degrades the signal-to-noise ratio (SNR) of such links.

An accurate model for indoor diffuse OW channels is the linear baseband system

$$y(t) = Rx(t) * h(t) + z(t), \quad (6.3)$$

where R [A/W] is the *responsivity* of the photodiode, $h(t)$ is the impulse response, and $*$ denotes a continuous-time convolution operator. The channel noise, $z(t)$, is high-intensity shot noise generated by ambient light and is modeled as being signal independent and approximately Gaussian distributed [5, 24, 25, 29].

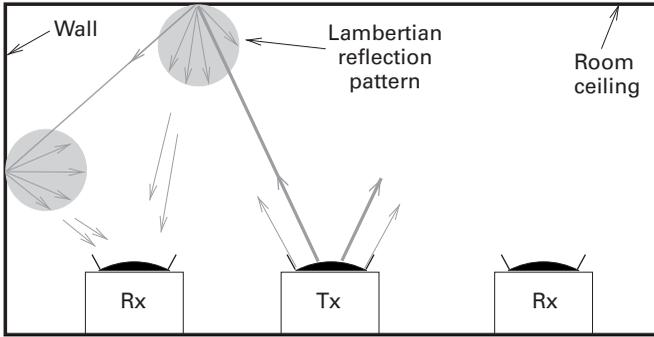


Figure 6.2 An illustration of an indoor diffuse OW link. The transmitter emits radiation over a wide angle and the receiver collects multipath components. (Reproduced with permission from [30].)

The channel impulse response is lowpass and has DC-gain defined as

$$H_0 = \int_{-\infty}^{\infty} h(t) dt, \quad (6.4)$$

which ranges from 10^{-7} to 10^{-5} [5, 31].

To quantify the degree of multipath dispersion, often the root-mean-square (RMS) delay spread of $h(t)$ is reported and is defined as,

$$\mathcal{D} = \sqrt{\frac{\int_{-\infty}^{\infty} (t - \varrho)^2 h^2(t) dt}{\int_{-\infty}^{\infty} h^2(t) dt}} \quad (6.5)$$

where

$$\varrho = \frac{\int_{-\infty}^{\infty} t h^2(t) dt}{\int_{-\infty}^{\infty} h^2(t) dt}$$

is the mean delay [5, 31]. In measurements of typical room configurations, this parameter varies in the range of 2 to 10 nanoseconds [5, 31].

There have been numerous techniques to quantify $h(t)$ for use in simulation studies. The most accurate are experimental measurements of test rooms [31–34], however, it is often difficult to generalize such results. Ray-tracing techniques [28, 35], related Monte Carlo-based ray-tracing simulations [36–38], and more recent photon-tracing simulations [27] provide good results at the expense of computing time. The use of functional modelling to estimate $h(t)$ [39, 40] provides results which are amenable to closed-form derivations, however, they are often quite approximate.

Two popular models for $h(t)$ are the exponential-decay model and the ceiling-bounce model [39]. In the exponential-decay model,

$$h(t) = \frac{H_0}{2\mathcal{D}} \exp\left(\frac{-t}{2\mathcal{D}}\right) U(t) \longleftrightarrow H(f) = \frac{H_0}{1 + j4\pi\mathcal{D}f}, \quad (6.6)$$

where $U(t)$ is the unit step and $H(f)$ is the continuous-time Fourier transform of $h(t)$. The ceiling bounce model is similar to the exponential-decay model with the tail of $h(t)$ dropping off as t^{-7} . The 3-dB cutoff of the channel can be easily related to the delay-spread in the case of functional models. For example, for the exponential model (6.6)

$$f_{3\text{dB}} = \frac{1}{4\pi\mathcal{D}}.$$

For this model, a \mathcal{D} in the range of 2–10 ns [5, 31], yields a channel bandwidth of approximately 10–40 MHz.

In any case, the impulse response depends on the specific room and on the positions and orientation of transmitters and receivers in the room. However, for a given room and configuration of transmitter and receiver, $h(t)$ is considered unchanging in time [5].

6.2.2 Previous work and possibilities

There has been relatively little work done on the use of MIMO techniques for diffuse channels. This channel is particularly challenging for MIMO systems in that the diffuse channel (with no line-of-sight) serves to remove the structure in the spatial distribution emitted by the transmitter through a series of reflections. This is due to the property of the diffuse Lambertian reflections, namely that the angle of reflection is nearly independent of the incident angle. Another way of considering the challenge is to note that the channel matrix of gains between transmitters and receivers has low rank since the impulse response typically does not vary much for small displacements such as those in elements of a transmitting or receiving array. However, having multiple copies of the same signal at the receiver is useful in providing a gain in SNR, i.e., a diversity gain.

There have been studies on the use of diversity receivers, such as imaging and angle diversity receivers, in diffuse settings [12, 14, 17]. As mentioned, the use of multiple narrow-FOV receivers yields the advantage of increasing the collected optical power while rejecting ambient light interference. However, for angle diversity systems which are able to sample signals arriving with vastly different path lengths, the processing algorithm must take into account the multipath spread between the branches.

An experimental OW MIMO diffuse system was reported in [41]. The popular 2×2 Alamouti space-time code was adapted to this channel. The channel gains were assumed known at the receiver by way of inserted pilot symbols. From the paper it seems as if the number of reflections is minimized to ensure a well-behaved channel matrix. Similar work was also performed at the University of Oxford, where a 2×9 MIMO configuration was established over a range of 1 m with a single reflection [42, 43]. No coding was employed and channel knowledge at the receiver was assumed. A pseudo-inverse of the channel matrix was applied at the receiver to recover data, but no detailed bit error rates (BERs) are reported under varying conditions.

In recent work [44], a novel indoor diffuse OW MIMO technique is introduced and termed *optical spatial modulation*. In this technique the room is populated with N_t transmitters distributed on the wall and ceiling while a mobile receiver has N_r receive elements. The key idea is that information is transmitted by selecting one of the N_t transmitters to be on while the others are off. That is, data are sent by the sequence of transmitters that are turned on. Given that the receiver has channel state information, a maximum likelihood receiver is derived for this case. However, the performance of this scheme depends greatly on the correlation between channel responses between each transmitter and the receiver array. The authors show that in some conditions this scheme works well when transmitter positions and orientations are optimized to ensure low correlation among the channel response from each transmitter to the receiver array.

Thus, previous work for OW MIMO diffuse channels focuses on diversity gains rather than multiplexing gains. The prospects for spatial multiplexing are more challenging since diffuse reflections tend to destroy spatial information of the incident beam. By limiting the number of reflections in the channel, diversity and multiplexing gains are possible. This is the basic idea behind spot-diffusing and point-to-point OW MIMO topologies which collect one or no reflected paths respectively.

6.3 Spot-diffusing OW MIMO systems

6.3.1 Channel architecture

The spot-diffusing configuration is one of the earliest examples of OW MIMO systems for indoor networking applications. It combines the advantages of mobility from diffuse links with the lower path loss and lower multipath distortion of line-of-sight links [35, 45–47].

Figure 6.3 presents a diagram of an indoor OW spot-diffusing system. The transmitter illuminates the ceiling with an array of *spots* using a series of low-divergence beams. Each spot on the ceiling can be thought of as a diffusely reflecting Lambertian source. The receiver is pointed toward the ceiling and images the spot(s) to detect the data. With proper system design, the receiver can be designed so that a small number of spots,

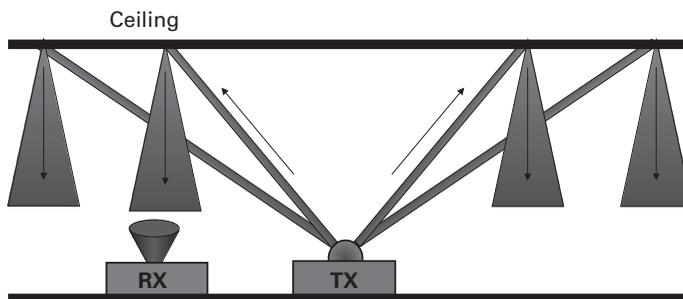


Figure 6.3 The spot-diffusing OW MIMO architecture.

ideally one, are imaged in any reception. Identical data are transmitted on all spots in the room and the arrangement and number of spots are chosen to ensure that for each position in the room that the receiver images at least one spot in its FOV. Spot-diffusing MIMO OW links were first introduced by Yun and Kavehrad [45] using a “fly-eye” receiver and are often termed *multi-spot diffusing* or *quasi-diffuse* links. Thus, spot-diffusing links can be considered as MIMO systems in which a spatial repetition of data is used to enable user mobility and to improve channel characteristics as noted below.

Notice that both diffuse links and spot-diffusing links both rely on diffuse reflections from room surfaces to establish a link from transmitter to receiver. However, in spot-diffusing links there is a well-defined *structure* to the emitted optical signals which is exploited to provide significant gain in the link. Firstly, spot-diffusing creates a series of spots using narrow divergence beams to the ceiling, while diffuse systems emit optical power in an unstructured manner. In this case, spot-diffusing links experience far smaller path loss than diffuse links as little power is lost from emitter to ceiling. In addition, diffuse emitters must have a wide divergence to cover points in the room which necessarily inherits a higher path loss.

The second major difference between diffuse and spot-diffusing links lies at the receiver. Due to the structured nature of the optical emissions on the ceiling, spot-diffusing systems can use narrow FOV receivers so long as a single spot is in the FOV. These narrow FOV receivers reject a larger portion of ambient light interference and hence reduce noise at the input of the receiver. In addition, the narrower FOV of the receiver rejects a large component of multipath distortion improving the bandwidth of such links. This is contrasted by the required wide FOV receivers for diffuse links which are limited by high ambient light noise and multipath distortion.

In previous computer simulation studies [46, 48], it was demonstrated that spot-diffusing links are not greatly impaired by multipath and have bandwidths in excess of 2 GHz whereas the bandwidth of diffuse links is limited to 50 MHz. Figure 6.4 plots the simulated impulse response for a spot diffusing channel in a prototypical room. Notice that the impulse response is dominated by the line-of-sight component and that the bandwidth is in excess of 2 GHz. Using an imaging receiver can also greatly improve the performance of spot-diffusing links by allowing for spatial rejection of multipath components and ambient light [17].

Spot-diffusing links permit a large degree of mobility and are robust against blockage. As long as there are sufficient number of spots on the ceiling and the FOV is large enough to catch at least one spot in every position, users are able to move about in a room while maintaining a high-speed link. It has been shown that spot-diffusing systems are robust against shadowing and that the likelihood of blockage is smaller than 2% [49].

6.3.2 Implementation challenges

Spot-diffusing links require a transmitter which is able to create a series of spots on the ceiling. In addition, in order to establish a link there must be joint design of the receiver FOV and the number of spots. Thus, for a given scenario, it is necessary to plan and design the spot-diffusing system specifically for the application.

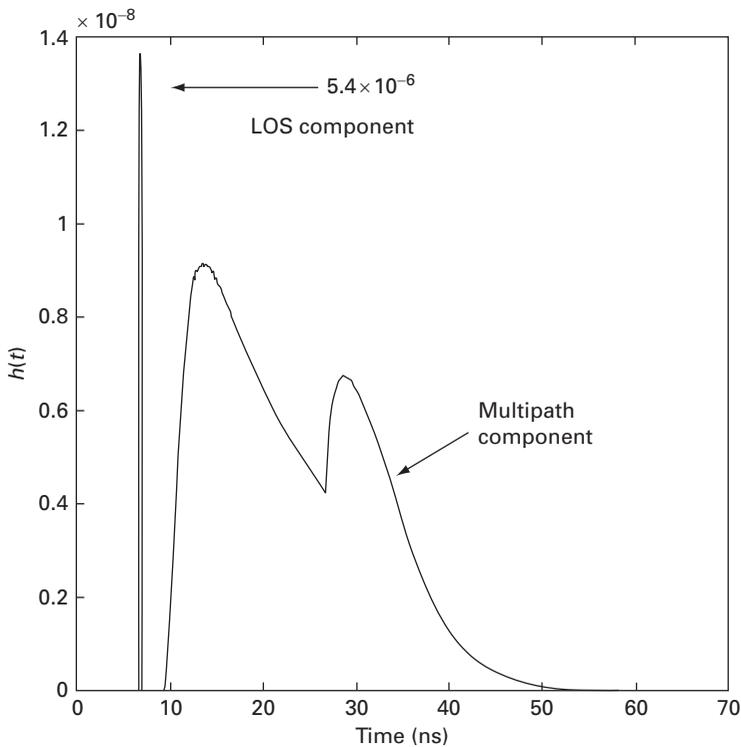


Figure 6.4 Simulated impulse response, $h(t)$, for a spot-diffusing system with 10×10 spots, in a $6 \times 6 \times 3$ m room. The receiver is single element with $\text{FOV}=11.2^\circ$ and $A_R = 1 \text{ cm}^2$. (Reproduced with permission from [50].)

There are a number of possible implementation techniques to implement the transmitter. The most direct approach is to use a series of emitters with narrow divergence oriented towards the ceiling. This approach results in an emitter which is bulky and which limits the number of spots. A preferable approach is to produce the series of spots using a single laser source and a computer generated hologram (CGH). Examples of this approach including the implementation of a CGH capable of producing an 8×8 beam array can be found in [47]. The CGH is a polymer which is fabricated to distort the wavefront of a point source (i.e., the laser) to produce the required beam pattern. This results in a compact and inexpensive multi-element transmitter. In contrast to radio channels, the creation of a complex beam pattern is quite simple in OW domain. In fact, CGH have also been suggested for larger arrays with 10×10 elements [51].

The number and position of spots on the ceiling must be optimized jointly with the selection of receiver FOV [47, 52]. For a single detector system, a small FOV is desirable to limit the amount of ambient light and multipath distortion. However, for a given number of spots, the FOV should be large enough so that at least one spot is always imaged by a receiver. Imaging receivers are particularly well suited to spot-diffusing links. They are simultaneously able to reject a large portion of ambient light and multipath

distortion. For this particular application, they may also be useful in selecting a single spot in the FOV to avoid any residual multipath distortion.

Alternate geometries for the arrangement of spots on the ceiling have also been considered. For example, line strip and diamond arrangements are less complex than previous full-array spot-diffusing systems but still provide a gain on the order of 4 dB electrical over diffuse systems [53, 54]. An experimental prototype spot-diffusing system has been demonstrated at 70 Mbps at a probability of error of 10^{-9} over a range of 4 m [12]. The specification for a spot-diffusing system operating at 100 Mbps has also been investigated and shown to be feasible over a range of 4.5 m [14]. A recent reference provides a nice background on implementation challenges in OW systems [10].

6.3.3 Dynamic spot diffusing architecture

The *dynamic spot-diffusing* (DSD) architecture is a special case of the general spot-diffusing model [55]. Figure 6.5 presents a block diagram of the DSD channel. In this architecture, the transmitter consists of one or a small number of spots which are translated across the ceiling in a closed path. The detector is a multi-element imaging receiver which performs select-best combining over all elements in the array. Data are received whenever the spot is in the FOV of the receiver and no data are received otherwise. The key difference between DSD and conventional spot-diffusing and diffuse channels is that the channel is time-varying, i.e., the spot motion introduces fading into the link. This added fading in fact eases transmitter implementation, allows for adaptive transmission, and permits larger transmit powers.

In conventional spot-diffusing techniques an array of spots is required over the ceiling. The DSD system has far fewer spots and hence has lower multipath distortion. As a result, high-speed modulators can be used to send data on the spot. In addition, the motion of the spot relaxes eye-safety constraints and permits the instantaneous

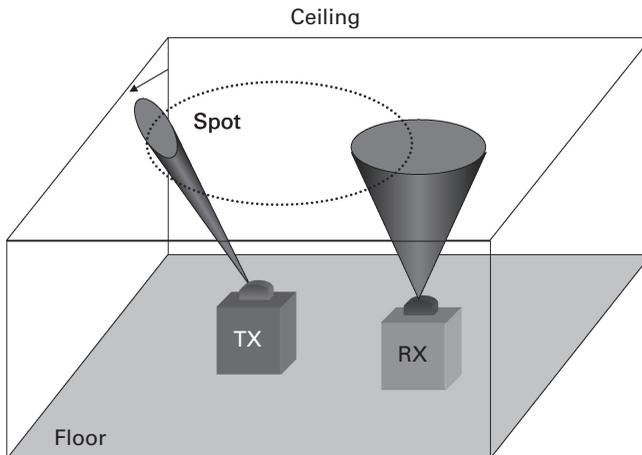


Figure 6.5 A dynamic spot-diffusing link with one rotating spot on ceiling and a imaging receiver.

emission of larger optical powers. The spot path can be adapted based on user feedback and, unlike previous spot-diffusing techniques, allows for flexibility in defining the distribution of optical power over the ceiling.

The DSD channel alternates between two states: a high data rate, high SNR channel when the spot is in the FOV of the receiver and a channel with near zero capacity. Rateless codes are applied to the DSD channel in order to be able to permit reliable communication. Examples of these codes are Luby transform (LT) codes and Raptor codes and are quite robust under a variety of channel conditions. User mobility is permitted in this channel topology and the rateless codes seamlessly adapt to the capacity of the channel at the given location. Simulations have demonstrated that rates on the order of 440 Mbps are achievable in a room using a previously reported receiver, eye-safe power limits, and two spots modulated at 1 Gbps. In fact, it is suggested that the bandwidth of the DSD channel is sufficient to support even higher data rates approaching 10 Gbps when appropriate optoelectronics become available [50, 55].

6.4 Point-to-Point OW MIMO communications

The most common topology for OW MIMO link is a point-to-point configuration where a line of sight exists between transmitter and receiver. This configuration exploits the spatial degrees of freedom supplied by arrays of emitters and receivers to achieve a gain in spectral efficiency via multiplexing. The receiver is typically an array of photodetectors which samples the spatial distribution of optical intensity at the receiver plane.

Whereas spot-diffusing links transmit identical data in all spatial directions to permit user mobility, point-to-point OW MIMO systems use spatio-temporal coding in order to convey more data over the channel. Thus, the signalling problem in point-to-point OW MIMO channels is equivalent to finding a set of time-varying images which are able to convey a large amount of data in spite of channel impairments.

6.4.1 Channel topology and applications

There are a huge number of areas in which point-to-point OW MIMO links are applied. With the availability of high-speed SLMs and detectors many new avenues of application are being opened. A small list of applications include holographic data storage systems [56–58], optical interconnects [59, 60], two-dimensional (2D) bar codes [61, 62], and indoor OW MIMO communication systems [5, 63–65].

6.4.1.1 Holographic data storage

Holographic data storage systems store data *volumetrically*, i.e., as a series of pages in a recording medium [66, 67]. This is in contrast to wide-spread magnetic storage systems which store data areally on a medium. The theoretical storage density of some holographic systems has been estimated to exceed 1 Tb/cm^3 [68] and are expected to have data transfer rates exceeding 10 Gbps [69, 70].

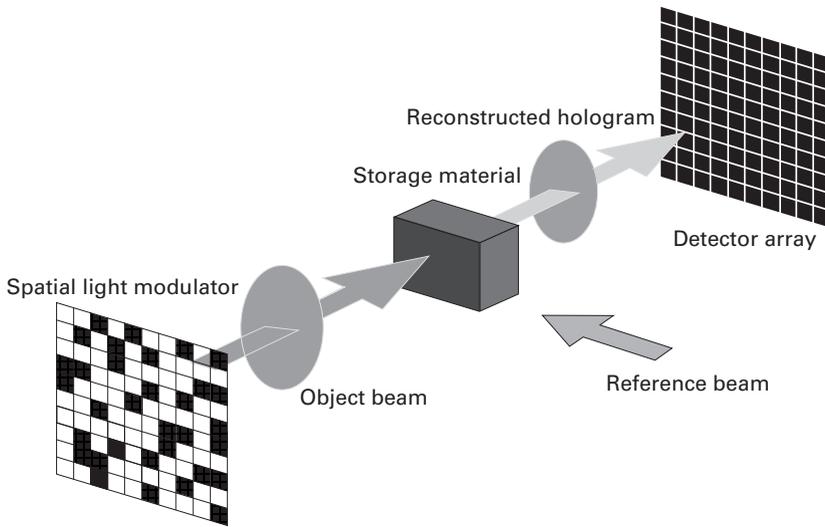


Figure 6.6 A holographic data storage system in an example of an optical wireless MIMO system. (Reproduced with permission from [72].)

Figure 6.6 presents a diagram of a typical holographic storage system. Binary data bits are multiplexed spatially into 2D images. The coherent interference pattern between the data image and a reference beam is stored in a medium such as lithium niobate. To recover the data image, the reference beam is used to illuminate the correct orientation of the medium and the resulting field is projected onto an array of detectors [71].

The main impairments in holographic storage systems arise due to the many sources of noise, misalignment, and spatial inter-symbol interference (ISI) [56, 58, 73]. To combat noise and the spatial bandwidth limitation of the holographic system, spatial equalization techniques have been developed. Developments include an adaptation of the Viterbi algorithm to process images [74] and the use of decision feedback techniques [75]. These techniques are rather complex considering the high-data throughputs of such systems. Constrained coding has been applied to these systems as well. These codes take as input the data bits and output images in which certain error producing output patterns are eliminated. In [76], a constrained code of rate $3/4$ was applied to a 2D holographic data storage channel and achieved a BER on the order of 10^{-3} with a coding gain of 5 dB over simple thresholding. Misalignment between transmitter and receiver arrays can lead to a large number of output errors. Oversampling at the receive array is one method considered to mitigate the impact of misalignment at the expense of more complex receiver arrays [77, 78].

6.4.1.2 Optical interconnects

Optical interconnection of printed circuit boards and chips is seen as a solution to the bottleneck of electrical signalling in many electronic systems. Copper interconnect, although appropriate for short links on the order of centimeters, is prone to attenuation,

crosstalk, and EMI issues [79] and is generally limited to about 15 Gbps for moderate length links [80]. Interconnects with rates as high as 20 Gbps have been reported [79] and there is much interest in active optical cables which are envisioned to deliver rates approaching a Tbps. Wireless optical interconnects have also been areas of interest in which 2D arrangements of emitters and detectors are used to produce multiple parallel channel for chip-to-chip, board-to-board, board-to-backplane applications.

A 4×4 OW interconnect has been demonstrated with a total data rate in excess of 1.7 Gbps [81]. This work was extended to a 256-channel bidirectional wireless optical interconnect between circuit boards [82]. These interconnect systems rely on tight spatial alignment between each transmitter element and each receiver. As such, they are often termed *pixel matched* systems. In fact, even after an initial alignment, dynamic forces in buildings can cause a relative motion between transmitter and receiver which causes misalignment. Some work has also been done to develop active alignment systems for wireless optical interconnects which are able to compensate for misalignment errors [83].

6.4.1.3 Two-dimensional barcodes

Conventional UPC barcodes are ubiquitous to label consumer and industrial products. There has recently been growing interest to include a greater amount of data into each barcode. Two-dimensional barcodes encode large amounts of data in the form of an image. There exist many standards for 2D barcodes. A popular format are QR Codes^{TM1} which can store over 7000 numerals in a 2D bar code while a conventional 1D barcode is limited to about 20. Figure 6.7 shows an example of such a code encoding a long text sequence. Notice that these barcodes are also MIMO OW communication links which communicate in *time* rather than *space*, i.e., they convey information from one time to another rather than one place to another. Many smart phones can read 2D barcodes using their built-in cameras and appropriate software. This functionality has been used to exchange identity information and to transmit URLs to readers of newspapers to point them to further content. Recent work has centered on improving the data density of such barcodes in spite of noise, misalignment, and blurring errors. Multi-level and multi-color barcodes have also been considered to improve data density [61, 62]. In recent work, advanced error correcting codes have been applied to 2D barcode applications [84, 85].

6.4.1.4 Wireless optical MIMO communications

A more straightforward application of OW MIMO systems is for indoor communications. Figure 6.8 depicts a possible embodiment of such a communication link. As mentioned above the transmitter is an SLM which encodes data into a time-varying optical intensity distribution, i.e., a sequence of images. The receiver samples the optical intensity field at its aperture and outputs measurements. As in all indoor OW systems, eye-safety standards must be satisfied for these types of links.

¹ QR Code is a registered trademark of Denso Wave Inc.



Figure 6.7 A 2D barcode is another example of a 2D OW MIMO storage channel. This QR CodeTM encodes the quotation: “Wireless communications is indeed one of the most active and fast growing sectors of the telecommunication industry. It has notably evolved since Guglielmo Marconi transmitted Morse-coded signals with his radio apparatus in the late 1800s.” (Generated online from [86].) (Reproduced with permission from [30].)

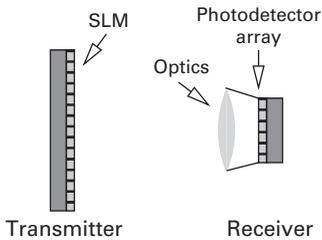


Figure 6.8 Optical Wireless MIMO point-to-point link. (Reproduced with permission from [72].)

The concept of employing multiple transmitters and receivers in this manner for indoor OW communications was first presented in [64] where a prototype system was implemented. The transmitter consisted of a 512×512 pixel liquid crystal display with 256 intensity levels. The receiver was a CCD imager with 154×154 pixels and operating at 60 frames per second. Data were multiplexed using spatial-discrete multitone (SDMT), to be described in Section 6.4.3 [29, Chapter 7]. Applying low-density parity check codes coupled with multi-stage coding and multi-stage decoding, they were able to obtain a data rate of approximately 17 kbits per frame, which amounted to 76% of estimated channel capacity [64].

In order to simplify the transmitter, binary-level SLMs were considered for point-to-point OW MIMO links. Halftoned-SDMT (HSDMT) was introduced as a means to create binary-level images which were robust to channel impairments [65]. Binary-level transmitters are less complex and less sensitive to channel nonlinearities. An additional

benefit of binary-level transmitters, mentioned in Section 6.1.2, is that they are able to be modulated at high-rates. Possible SLMs are arrays of lasers or even DMD arrays. In [87] a capacity estimate for HSDMT systems was obtained using channel measurements. Predicted rates using commercial video processing hardware were on the order of 450 Mbps for a 1 m link using 0.5 megapixel arrays at 7000 frames per second. Section 6.4.5 presents more detailed results of these experiments. Techniques developed to produce DMT signals compatible with optical intensity channels have also been applied to 2D OW point-to-point MIMO channels. In [88] the asymmetrically clipped optical (ACO)-OFDM technique is adapted for 2D channels to produce images which are nonnegative without the need for directly adding a bias to each pixel at the cost of losing half of the in-band spatial frequency bins.

Spatial multiplexing has also been considered in indoor environments using a mixture of a line arrangement of transmitters and an array of receivers [89]. The transmitters were considered to be a linear array of lasers and the receiver had an array of photodetecting elements given a coupled IM/DD channel between each pair of transmitters and receivers. Each receive photodiode receives a linear combination of all transmit intensities. As a result, this OW MIMO system was well modeled as a channel with spatial inter-channel interference. A linear equalizer was formulated to combat the interference between channels and to decouple them. It should be noted that this approach is quite computationally intensive, especially as the number of elements grows, and requires the receiver to know the channel characteristics. A similar technique was also employed later in which the MIMO channel was measured a priori and inversion of the MIMO channel was done at the receiver [42, 43].

In recent times there has been increased excitement over point-to-point OW MIMO on visible-light communication (VLC) channels. Although these channels are substantially similar to infrared channels, the primary difference is in the device bandwidth constraint (rather than multipath distortion) and the positioning of lighting elements. In [90] a comprehensive introduction to point-to-point MIMO for VLC channels is presented. The authors treat a collection of lighting fixtures on the ceiling as a MIMO emitter and use both imaging and non-imaging receivers. Again, complete channel knowledge is assumed at the receiver which simply inverts the channel matrix. In [91] a 2×9 point-to-point OW VLC MIMO system is presented which is able to send two independent OFDM signals. Bit error rates of 10^{-5} are measured for 100 Mbps links over a range of 65 cm.

6.4.2 Channel model

Figure 6.8 presents a diagram of a point-to-point OW MIMO system. The transmitter is an SLM which creates a 2D optical intensity image while the receiver is an array of photodetectors and data are sent by transmitting a time-varying image.

An assumption in the channel model is that imaging optics are employed at the receiver and that the axes of the transmitter and receiver coincide. In this case the received image $y(u, v)$ is an orthographic projection of the transmitted frame $x(u, v)$.

The images at the receiver need not be strictly aligned to the receive pixels and may have undergone some translation or rotation.

A good approximation for this system, from scalar diffraction theory, is to model the input–output relation as a spatial lowpass filter which is space invariant and characterized by a point-spread function (PSF) [92, Chapter 6]. In addition, the shape of both transmitter and receiver pixels can also be incorporated into the lowpass channel model [64, 93]. The channel model is thus,

$$y(u, v) = x(u, v) \otimes h(u, v) + z(u, v), \quad (6.7)$$

where $h(u, v)$ is the system PSF, \otimes is the 2D linear convolution, and $z(u, v)$ is spatially and temporally white channel noise [64, 93]. Notice that this channel model is significantly different than conventional models for MIMO systems in radio systems or in some other work in OW communications [42, 43]. In radio systems a more general model is adopted in which channel gains between each transmit and receive element are collected into a *channel matrix*. In this light, the channel matrix of the point-to-point OW MIMO channel has far more structure than in the general case. Spatial invariance and the notion of PSF from scalar diffraction theory are the sources of the great structure in channel gains. Thus, it is instructive to exploit this inherent structure in the spatial distribution of the received intensity in channel modelling and system design.

Notice that in the model (6.7) the magnification is set to unity. In general this is not a requirement but is done here for simplicity of description. In fact, if the magnification of the system is chosen so that the received image is an integer number of receiver pixels system design and modeling are unaffected [64]. In cases in which this condition on magnification is not met, spatial equalization is needed to ensure good system performance free of inter-channel interference, and will be discussed in the next section [72, 94]. Similar models to (6.7) have been applied in holographic data storage [95] and areal optical recording [96], and in previous work in OW communication [64, 93].

The transmitted image, $x(u, v)$, must satisfy non-negativity and average amplitude constraints, i.e.,

$$x(u, v) \geq 0, \quad (6.8)$$

and

$$\mathcal{P}_t = \frac{1}{|\mathcal{A}_t|} \iint_{\mathcal{A}_t} x(u, v) dudv \leq \mathcal{P}, \quad (6.9)$$

where \mathcal{A}_t is the area of the transmit aperture. Notice that it is assumed that the average amplitude constraint holds for every frame, not just for the average over all frames which is a more restrictive constraint.

6.4.3 Spatial discrete multitone modulation

A popular modulation technique for frequency selective channels is discrete multitone modulation (DMT) [97], [98, Section 12.2.2]. Spatial discrete multitone (SDMT) modulation was developed as a generalization of DMT to two-dimensional spatially dispersive channels such as the point-to-point OW MIMO channel [64, 93], [29, Section 7.5].

A key feature of SDMT is that it does not require strict spatial alignment between transmitter and receiver, but is insensitive to spatial sampling phase (i.e. position) of the receiver. Consider defining an SDMT symbol, $x(n_1, n_2)$, to be transmitted as an image of $N_1 \times N_2$ pixels. Define $X(k_1, k_2)$ as the discrete Fourier transform (DFT) of $x(n_1, n_2)$. Data can be loaded into the image in spatial frequency domain directly ensuring that Hermitian symmetry, $X(k_1, k_2) = X^*(N_1 - k_1, N_2 - k_2)$, is present so that the output amplitudes are real. The transmitted image is then

$$x(n_1, n_2) = \mathcal{F}^{-1}\{X(k_1, k_2)\},$$

where \mathcal{F}^{-1} denotes the inverse DFT.

Notice from (6.7), that the received image is the linear convolution of the transmit image by the PSF. In the same spirit as DMT, a cyclic extension is appended around each transmit SDMT image which greatly simplifies equalization at the receiver [64, 97]. The size of the cyclic extension should be at least half of the channel memory to ensure the periodicity assumption of the DFT is satisfied.

With these assumptions, the channel model (6.7) can be written in the discrete spatial frequency domain (k_1, k_2) as

$$Y(k_1, k_2) = H_0 X(k_1, k_2) H(k_1, k_2) + Z(k_1, k_2) \quad (6.10)$$

where $H_0 H(k_1, k_2)$ is the optical transfer function (OTF) of the system (i.e. the DFT of the discrete space PSF) and H_0 is the spatial DC gain of the channel. Both $Y(k_1, k_2)$ and $Z(k_1, k_2)$ are the DFTs of the image at the receiver and the noise. Notice that since images are assumed to be rectangular that DFT and IDFT are separable and have efficient implementations. This is not a requirement of the channel model and other more dense arrangements of pixels are indeed possible, e.g., hexagonal.

Figure 6.9 shows an example of an SDMT image with its cyclic extension. At the receiver, the cyclic extension is clipped and the DFT is performed to allow detection

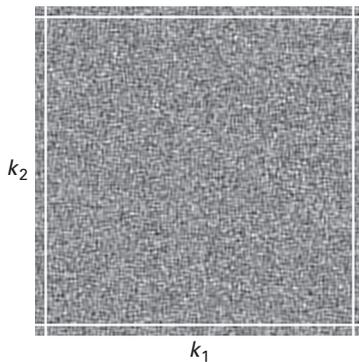


Figure 6.9 A typical SDMT symbol with 256×256 pixels and a cyclic extension of width 8 pixels around the image. Note that the white border is only for clarification. (Reproduced with permission from [30].)

in the spatial frequency domain. Equalization is done in each spatial frequency bin and consists of a single complex multiplication by $\frac{1}{H_0 H(k_1, k_2)}$ [64, 97]. In other words,

$$\frac{Y(k_1, k_2)}{H_0 H(k_1, k_2)} = X(k_1, k_2) + \frac{Z(k_1, k_2)}{H_0 H(k_1, k_2)}.$$

Thus, using the IDFT/DFT and loading data in spatial frequency, the point-to-point OW MIMO channel can be considered as a series of parallel Gaussian noise channels in spatial frequency domain. The capacity of the system, ignoring the amplitude constraints, is achieved by constructing images by performing the water pouring algorithm over spatial frequency channels [99, Chapter 10], [64].

In earlier work, the water pouring algorithm was applied directly without considering non-negativity or average amplitude constraints and biasing and clipping were done to ensure that the resulting image was transmittable on the OW channel [64, 93]. Due to the fact that the channel constraints do not map nicely into spatial frequency domain, the use of the water pouring algorithm is not straightforward. Another issue with the implementation of SDMT is that arrays of emitters are required that are able to produce a continuous range of intensities at high dynamic range and high rate.

6.4.4 Spatial halftoning with binary-level emitters

The use of SDMT requires a complex transmit array which must output continuous-level amplitudes. There exist many SLMs which can output binary-level images at high frame rates and a large number of transmitting elements. Examples include DMD arrays and arrays of lasers. That is, these SLMs have the ability to generate images which contain high-frequency spatial content. In contrast in many OW MIMO point-to-point links there exists a narrow spatial bandwidth which effectively filters out the high-frequency spatial modes due to blurring. Therefore, in many scenarios there are excess degrees of freedom at the transmitter to produce images with binary levels and high spatial frequency content but the channel is unable to support the transmission of such modes. The high-spatial frequency modes are referred to as the *out-of-band* region of the channel while the other spatial frequencies are termed the *in-band* region and are illustrated in Figure 6.10 for a Gaussian OTF.

In order to trade off the excess spatial degrees of freedom at the transmitter for the ability to generate binary-level images *error diffusion halftoning* was applied to SDMT to yield *halftoned* SDMT (HSDMT) [65]. Halftoning was originally proposed in 1975 and is a special instance of $\Delta\Sigma$ modulation [100, 101]. However, in this application, the goal is not to produce a binary-level image which is perceptually close to a continuous-level image, but to construct a high-rate communication system. Figure 6.11 presents a block diagram of an HSDMT system. An SDMT signal is created where data are modulated into the low spatial frequency bins which are able to propagate through the channel. An IFFT is performed to yield a continuous-level image which is passed through a halftoning system in the dashed box. Define q as the quantizer error and $\tilde{q} = v - x$ as the closed loop *quantization noise*, where x is the continuous-level SDMT image and v is the binary-level output image. In spatial frequency domain,

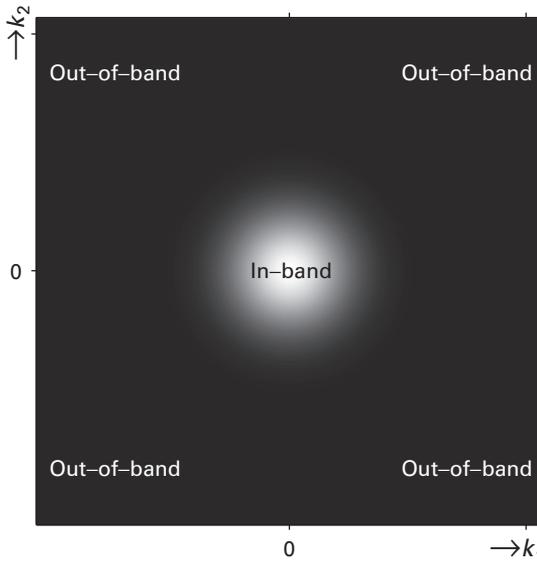


Figure 6.10 An example of an OTF with in-band and out-of-band regions indicated. In this example, the OTF is circularly symmetric which is typical in many imaging systems. (Reproduced with permission from [30].)

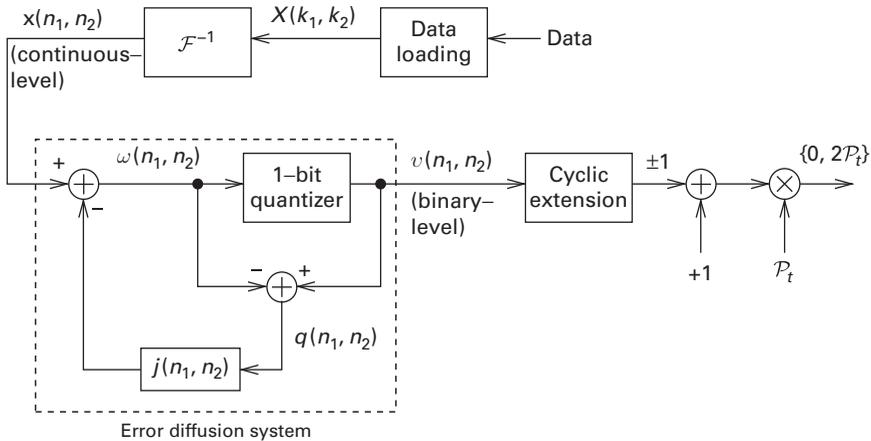


Figure 6.11 A block diagram of half-toned-SDMT. (Reproduced with permission from [30].)

$$\tilde{Q}(k_1, k_2) = Q(k_1, k_2) [1 - J(k_1, k_2)], \tag{6.11}$$

where $Q(k_1, k_2)$, $\tilde{Q}(k_1, k_2)$, and $J(k_1, k_2)$ are DFTs of the errors q , \tilde{q} , and the feedback filter j . It follows that the power spectral density (PSD) of the quantization noise, \tilde{q} , is

$$\Phi_{\tilde{q}}(k_1, k_2) = \Phi_q(k_1, k_2) |1 - J(k_1, k_2)|^2, \tag{6.12}$$

where $\Phi_q(k_1, k_2)$ is the PSD of the quantizer error. It is clear that the power spectrum of the quantizer error is shaped by the transfer function $|1 - J(k_1, k_2)|^2$, which is termed the noise shaping function. The feedback filter must be designed to shape the quantization to the out-of-band spatial frequency bins which are effectively filtered-out by the channel $H(k_1, k_2)$.

The received image, y , is a continuous-tone image which contains the lowpass transmitted data bins as well as some remaining quantization noise. That is,

$$Y(k_1, k_2) = \mathcal{P}_t H_0 \left[X(k_1, k_2) + \tilde{Q}(k_1, k_2) + N_1 N_2 \delta(k_1, k_2) \right] H(k_1, k_2) + Z(k_1, k_2). \quad (6.13)$$

After equalization in each of the in-band spatial frequency bins

$$\frac{Y(k_1, k_2)}{\mathcal{P}_t H_0 H(k_1, k_2)} = \left[X(k_1, k_2) + N_1 N_2 \delta(k_1, k_2) \right] + \tilde{Q}(k_1, k_2) + \tilde{Z}(k_1, k_2),$$

where

$$\tilde{Z}(k_1, k_2) = \frac{Z(k_1, k_2)}{\mathcal{P}_t H_0 H(k_1, k_2)}$$

is the remaining channel noise. Thus, the received image is a biased version of the original corrupted by channel noise as well as residual quantization noise.

6.4.5 Experiments and results

In order to quantify the performance of SDMT systems and contrast them versus HSDMT, an experimental link was constructed and channel measurements performed. Figure 6.12 presents a picture of the experimental link.

The transmitter SLM is a commercial DLP chip coupled to projection optics [11]. This setup permitted frames to be output at rates of up to 10 kfps and a maximum

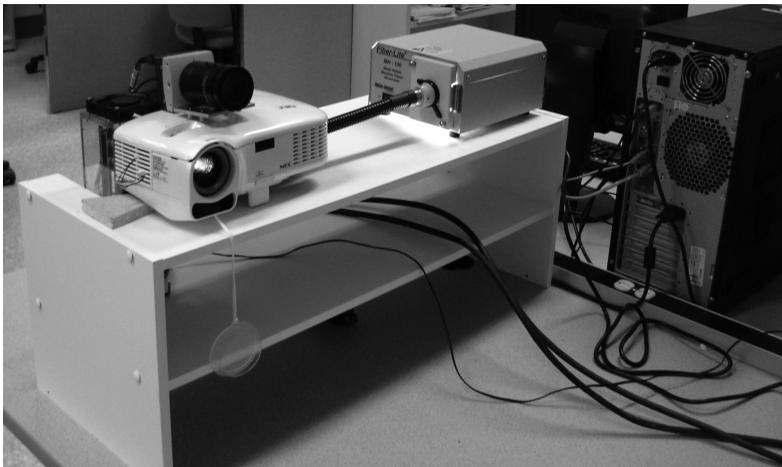


Figure 6.12 Point-to-point OW MIMO link prototype setup.

size of 1024×768 pixels. The receiver is a CMOS digital camera [102, MC1310] connected to a frame grabber [103, Odyssey XCL]. An image was projected onto a wall 1.1 m away and the reflected image detected on the receiver. The camera was positioned to approximately align the optical axes to minimize projective distortion. The OTF and channel noise were measured for this link and used to construct a comprehensive channel model [30, 72, 87]. The feedback filter for the halftoning algorithm was chosen carefully to maximize the capacity of the underlying channel [65, 87] and was selected to be

$$j(n_1, n_2) = \begin{bmatrix} 0 & 0.9 \\ 0.9 & -0.8 \end{bmatrix}.$$

Figure 6.13 presents estimates of the capacity in Mbps for an HSDMT system where transmit and receive frames are assumed to have 750×750 pixels each for various frame rates.

An upper bound on the performance of any HSDMT system can be found by setting $\tilde{Q} = 0$ for all spatial frequency bins in (6.13). This is referred to as the continuous upper bound since it ignores the amplitude constraints and sends continuous level signals on the channel. For the other cases, it was assumed that the quantization noise was white and signal independent. This was justified by setting the input power based on a theoretical bound and confirming the independence assumption after extensive simulation [30]. The capacity was then estimated by considering parallel Gaussian channels with the optimum water pouring power allocation. Simulations were also carried out for cases where multi-level quantizers were available with similarly designed systems [72, 104].

Notice that in all cases the capacity of the link increases with frame rate up to about 7000 fps and then decreases slightly. As frame rate increases, the number of bits sent per frame decreases, however, this is balanced by the increase in number of frames per

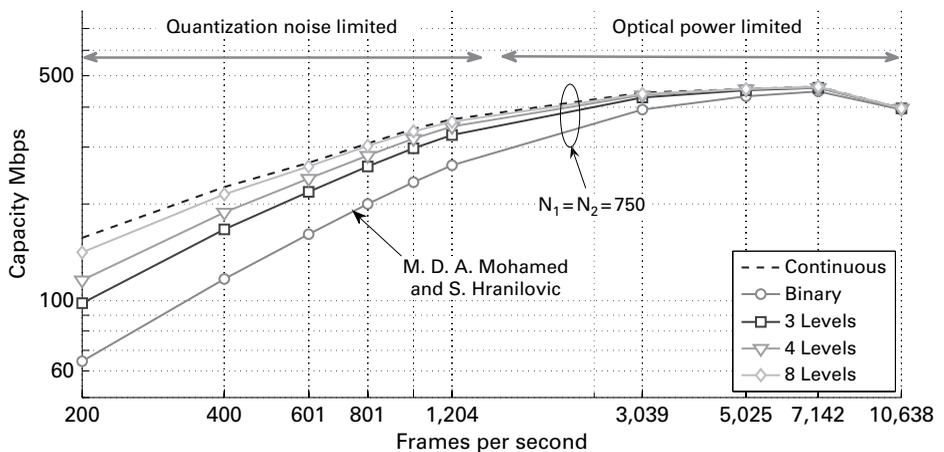


Figure 6.13 Capacity estimate of HSDMT channel versus frame rate. Continuous SDMT transmitter is the idealistic case without non-negativity constraint used only for comparison. The presented results are based on the measured channel parameters. (Reproduced with permission from [72].)

second. Asymptotically, in the case of white noise these two effects should be perfectly balanced, however, it is apparent that the noise in the experimental link was not white. Notice also that at low fps, there is a large discrepancy between the capacity available using the idealistic continuous model and the halftoned schemes. In this range of frame-rates the channel is termed *quantization noise limited* since the water pouring spectrum is dominated by the quantization noise introduced by halftoning. Notice also that in this regime quantizers with multiple levels are useful in approaching the continuous bound on capacity [72, 104]. At higher frame rates, the channel is termed *optical power limited* since the dominant source of noise is due to ambient light. In this regime, even binary signalling can approach the continuous bound. Therefore, using binary-level halftoning with high frame rate provides high data rates. In a first trial experiment using no coding, we have been able to achieve rates of 250 Mbps with binary halftoning with 768×1024 frames over a distance of 0.55 m in the presence of a significant magnification error of 0.89 using techniques presented in [94].

6.5 Future directions

The use of multiple transmitters and receivers in indoor OW channels has been shown to improve data rates and reliability in a wide variety of applications. A diversity gain can be had simply by using a receiver with multiple elements. The key feature of such a receiver, be it angle diversity or imaging, is that it is able to spatially sample the received optical intensity image. This permits the use of combining techniques to improve the received SNR. In addition, in many point-to-point links, a multiplexing gain in the rate is available by exploiting spatial degrees of freedom. Although the channel has dispersion, the spatial dispersion is well structured and nicely represented by a linear space model. The key feature in these systems is that the multi-element transmitter can impose structure on the spatial distribution of power. Similarly, a multi-element receiver is able to provide a gain, either in SNR or in rate, by exploiting any underlying *structure* in the received optical intensity distribution. This is useful in cases where multiple or even single emitters are available.

There is a plethora of exciting future applications for MIMO techniques in OW systems. Among the applications are visible light communication (VLC) channels. Such channels are quickly becoming of commercial interest due to their use of illumination devices for a dual role as communication systems. The white phosphorescent LEDs typically used in such VLC applications have very limited bandwidth, at most 20 MHz. The use of MIMO techniques to improve the spectral efficiency of such links is currently a topic of intense interest in the community. Many VLC applications are naturally MIMO problems since each illumination device is composed of numerous white LED emitters. In addition, white LEDs are currently used in backlighting applications for LCD displays and their use for communications has yet to be deeply investigated. In automotive applications, arrays of white LEDs and receivers are envisioned for car-to-car communications primarily for collision avoidance and navigation applications. Another interesting avenue of study for indoor OW MIMO is in the area of areal storage

and bar codes. These storage channels can be simply viewed as a communications system in time. There are many applications for such 2D storage and transmission channels which would benefit from further study.

Ongoing challenges in the implementation of high-speed optical transmit arrays and integrated imaging receivers are vital directions of research required in order to allow system designers to reap the rewards of MIMO techniques. Although progress has been made in developing integrated emitters and holographic-based filtering techniques, added work is required to implement such systems. Encouragingly, there is much work in the development of high-speed detector arrays and circuitry which are packaged in a compact form factor.

Certainly, there are many hurdles to overcome in the implementation of OW MIMO systems. However, the potential gains in SNR and rate are exciting and justify redoubled efforts to make such approaches widespread. Regardless of the application, the use of MIMO OW techniques improves system performance to such a degree that indeed with OW MIMO techniques “a picture can be worth *far more than* a thousand words”!

References

- [1] L. Zheng and D.N.C. Tse, “Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels,” *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
- [2] *IEC 60825-1, Safety of Laser Products - Part 1: Equipment Classification, Requirements, and User’s Guide*, International Electrotechnical Commission (IEC), 2001, edition 1.2.
- [3] A. C. Boucouvalas, “IEC 825-1 eye safety classification of some consumer electronic products,” in *Proceedings of the IEE Colloquium on Optical Free Space Communication Links*, IEE Savoy Place, London, 19 Feb. 1996, pp. 13/1–13/6.
- [4] T.-S. Chu and M. Gans, “High speed infrared local wireless communication,” *IEEE Communications Magazine*, vol. 25, no. 8, pp. 4–10, Aug. 1987.
- [5] J. M. Kahn and J. R. Barry, “Wireless infrared communications,” *Proceedings of the IEEE*, vol. 85, no. 2, pp. 263–298, Feb. 1997.
- [6] W. Hubbard, “The approximation of a Poisson distribution by a Gaussian distribution,” *Proceedings of the IEEE*, vol. 58, no. 9, pp. 1374–1375, Sept. 1970.
- [7] M. B. Venditti, E. Laprise, J. Faucher *et al.*, “Design and test of an optoelectronic-VLSI chip with 540-element receiver-transmitter arrays using differential optical signaling,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 9, no. 2, pp. 361–379, Mar./Apr. 2003.
- [8] D. C. O’Brien, G. E. Faulkner, E. B. Zyambo *et al.*, “Integrated transceivers for optical wireless communications,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 11, no. 1, pp. 173–183, Jan./Feb. 2005.
- [9] A. Boryssenko, J. Liao, J. Zeng, V. Joyner, and Z. R. Huang, “Studies on RF-optical dual mode wireless communication modules,” in *2009 IEEE International Microwave Symposium*, June 7–12, 2009, pp. 805–808.
- [10] D. O’Brien, H. L. Minh, G. Faulkner *et al.*, “High data-rate infra-red optical wireless communications: Implementation challenges,” in *2010 IEEE Globecom Workshop on Optical Wireless Communications*, pp. 1047–1051, Dec. 2010.

- [11] Texas Instruments Incorporated, Digital Light Processing. URL: www.dlp.com.
- [12] J. B. Carruthers and J. M. Kahn, "Angle diversity for nondirected wireless infrared communication," *IEEE Transactions on Communications*, vol. 48, no. 6, pp. 960–969, June 2000.
- [13] A. Tavares, R. Valadas, R. L. Aguiar, and A. O. Duarte, "Angle diversity and rate-adaptive transmission for indoor wireless optical communications," *IEEE Communications Magazine*, pp. 64–73, Mar. 2003.
- [14] J. M. Kahn, R. You, P. Djahani *et al.*, "Imaging diversity receivers for high-speed infrared wireless communication," *IEEE Communications Magazine*, pp. 88–94, Dec. 1998.
- [15] J. Zeng, V. Joyner, J. Liao, S. Deng, and Z. Huang, "A 5Gb/s 7-channel current-mode imaging receiver front-end for free-space optical MIMO," in *Proc. IEEE Midwest Symposium on Circuits and Systems (MWSCAS)*, Aug. 2009, pp. 148–151.
- [16] K. Kagawa, J. Ohta, and J. Tanida, "Dynamic reconfiguration of differential pixel output for CMOS imager dedicated to WDM-SDM indoor optical wireless LAN," *IEEE Photonics Technology Letters*, vol. 48, no. 18, pp. 1308–10, Sept. 2009.
- [17] P. Djahani and J. M. Kahn, "Analysis of infrared wireless links employing multibeam transmitters and imaging diversity receivers," *IEEE Transactions on Communications*, vol. 48, no. 12, pp. 2077–2088, Dec. 2000.
- [18] B. S. Leibowitz, B. E. Boser, and K. S. J. Pister, "CMOS smart pixel for free-space optical communication," in *Proceedings of the SPIE*, vol. 4306, Jan. 2001, pp. 308–318.
- [19] D. C. O'Brien, G. E. Gaulkner, K. Jim *et al.*, "High-speed integrated transceivers for optical wireless," *IEEE Communications Magazine*, vol. 41, no. 3, pp. 58–62, 2003.
- [20] S. Kleinfelder, S. Lim, X. Liu, and A. El Gamal, "A 10000 frames/s CMOS digital pixel sensor," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 12, pp. 2049–2059, Dec. 2001.
- [21] S. Kleinfelder, S. Lim, X. Liu, and A. E. Gamal, "A 10 kframe/s 0.18 μm CMOS digital pixel sensor with pixel-level memory," in *Proceedings of the IEEE International Solid-State Circuits Conference*, San Francisco, CA, USA, 4–8 Feb. 2001, pp. 88–89.
- [22] T. Etoh, D. Poggemann, G. Kreider *et al.*, "An image sensor which captures 100 consecutive frames at 1000000 frames/s," *IEEE Transactions on Electron Devices*, vol. 50, no. 1, pp. 144–151, Jan. 2003.
- [23] S. Kleinfelder, Y. Chen, K. Kwiatkowski, and A. Shah, "High-speed CMOS image sensor circuits with in situ frame storage," *IEEE Transactions on Nuclear Science*, vol. 51, no. 4, pp. 1648–1656, Aug. 2004.
- [24] F. Gfeller and U. Bapst, "Wireless in-house data communication via diffuse infrared radiation," *Proceedings of the IEEE*, vol. 67, no. 11, pp. 1474–1486, Nov. 1979.
- [25] J. M. Kahn, J. R. Barry, M. D. Audeh, J. B. Carruthers, W. J. Krause, and G. W. Marsh, "Non-directed infrared links for high-capacity wireless LANs," *IEEE Personal Communications*, vol. 1, no. 2, pp. 12–25, 1994.
- [26] M. Kotzin and A. van den Heuvel, "A duplex infra-red system for in-building communications," in *Proceedings of the 36th IEEE Vehicular Technology Conference*, vol. 36, May 1986, pp. 179–185.
- [27] H.-S. Lee, "A photon modeling method for the characterization of indoor optical wireless communication," *Progress In Electromagnetics Research, PIER* 92, pp. 121–136, 2009.
- [28] J. R. Barry, J. M. Kahn, W. J. Krause, E. A. Lee, and D. G. Messerschmitt, "Simulation of multipath impulse response for indoor wireless optical channels," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 3, pp. 367–379, Apr. 1993.

-
- [29] S. Hranilovic, *Wireless Optical Communication Systems*. Springer, 2004.
- [30] M. D. A. Mohamed, "Modulation and detection in wireless optical channels using temporal and spatial degrees of freedom," Ph.D. dissertation, Department of Electrical and Computer Engineering, McMaster University, 2010.
- [31] J. M. Kahn, W. J. Krause, and J. B. Carruthers, "Experimental characterization of non-directed indoor infrared channels," *IEEE Transactions on Communications*, vol. 43, no. 234, pp. 1613–1623, Feb./Mar./Apr. 1995.
- [32] H. Hashemi, G. Yun, M. Kavehrad, F. Behbahani, and P. A. Galko, "Indoor propagation measurements at infrared frequencies for wireless local area networks applications," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 3, pp. 562–576, Aug. 1994.
- [33] H. Hashemi, G. Yun, M. Kavehrad, and F. Behbahani, "Frequency response measurements of the wireless indoor channel at infrared frequencies," in *Proceedings of the IEEE International Conference on Communications*, vol. 3, May 1994, pp. 1511–1515.
- [34] M. R. Pakravan and M. Kavehrad, "Indoor wireless infrared channel characterization by measurements," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 4, pp. 1053–1073, July 2001.
- [35] Y. A. Alqudah and M. Kavehrad, "MIMO characterization of indoor wireless optical link using a diffuse-transmission configuration," *IEEE Transactions on Communications*, vol. 51, no. 9, pp. 1554–1560, Sept. 2003.
- [36] F. J. López-Hernández, R. Pérez-Jiménez, and A. Santamaría, "Ray-tracing algorithms for fast calculation of the channel impulse response on diffuse IR wireless indoor channels," *Optical Engineering*, vol. 39, no. 10, pp. 2775–2780, 2000.
- [37] O. González, S. Rodríguez, R. Pérez-Jiménez, B. R. Mendoza, and A. Ayala, "Error analysis of the simulated impulse response on indoor wireless optical channels using a Monte Carlo-based ray-tracing algorithm," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 124–130, Jan. 2005.
- [38] Y. Cocheril and R. Vauzelle, "A new ray-tracing based wave propagation model including rough surfaces scattering," *Progress In Electromagnetics Research, PIER 75*, pp. 357–381, 2007.
- [39] J. B. Carruthers and J. M. Kahn, "Modeling of nondirected wireless infrared channels," *IEEE Transactions on Communications*, vol. 45, no. 10, pp. 1260–1268, Oct. 1997.
- [40] V. Jungnickel, V. Pohl, S. Nönnig, and C. von Helmolt, "A physical model of the wireless infrared communication channel," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 3, pp. 631–640, Apr. 2002.
- [41] M. Garfield, C. Liang, T. P. Kurzweg, and K. R. Dandekar, "MIMO space-time coding for diffuse optical communication," *Microwave and Optical Technology Letters*, vol. 48, no. 6, pp. 1108–1110, June 2006.
- [42] D. O'Brien, "Multi-input multi-output (MIMO) indoor optical wireless communications," in *43rd Asilomar Conference on Signals, Systems and Computers*, Nov. 1–4, 2009, pp. 1636–1639.
- [43] D. O'Brien, "Optical multi-input multi-output systems for short-range free-space data transmission," in *7th Conference on Communication Systems, Networks and Digital Signal Processing (CSNDSP 2010)*, Jul. 21–23, 2010, pp. 566–570.
- [44] R. Mesleh, R. Mehmood, H. Elgala, and H. Haas, "Indoor MIMO optical wireless communication using spatial modulation," in *Proc. 2010 International Conference on Communications*, May 23–27, 2010.

- [45] G. Yun and M. Kavehrad, "Spot-diffusing and fly-eye receivers for indoor infrared wireless communications," *In Proc. of IEEE International Conference on Selected Topics in Wireless Communications*, pp. 262–265, 1992.
- [46] S. T. Jivkova, B. A. Hristov, and M. Kavehrad, "Power-efficient multispot-diffuse multiple-input-multiple-output approach to broad-band optical wireless communications," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 3, pp. 882–889, May 2004.
- [47] S. T. Jivkova and M. Kavehrad, "Multispot diffusing configuration for wireless infrared access," *IEEE Transactions on Communications*, vol. 48, no. 6, pp. 970–978, June 2000.
- [48] S. Jivkova and M. Kavehrad, "Receiver designs and channel characterization for multi-spot high-bit-rate wireless infrared communications," *IEEE Transactions on Communications*, vol. 49, no. 12, pp. 2145–2153, Dec. 2001.
- [49] S. Jivkova and M. Kavehrad, "Shadowing and blockage in indoor optical wireless communications," *In IEEE Proceedings Global Telecommunications Conference*, pp. 3269–3273, 2003.
- [50] F. Khozeimeh, "Dynamic spot diffusing channel – a novel configuration for indoor optical wireless communications," Master's thesis, Department of Electrical and Computer Engineering, McMaster University, 2006.
- [51] M. Kavehrad and S. Jivkova, "Indoor broadband optical wireless communications: optical subsystems design and their impact on channel characteristics," *IEEE Wireless Communications Magazine*, pp. 30–35, Apr. 2003.
- [52] S. Jivkova and M. Kavehrad, "Multi-spot diffusing configuration for wireless infrared access: Joint optimization of multi-beam transmitter and angle diversity receiver," *In Proc. Conference Optical Wireless Communications II, SPIE.*, vol. 3850, pp. 72–77, Sept. 1999.
- [53] A. G. Al-Ghamdi and J. M. H. Elmirghani, "Line strip spot-diffusing transmitter configuration for optical wireless systems influenced by background noise and multipath dispersion," *IEEE Transactions Communications*, vol. 52, no. 1, pp. 37–45, Jan. 2004.
- [54] A. G. Al-Ghamdi and J. M. H. Elmirghani, "Multiple spot diffusing geometries for indoor optical wireless communication systems," *International Journal of Communication Systems*, vol. 16, pp. 909–922, 2003.
- [55] F. Khozeimeh and S. Hranilovic, "Dynamic spot diffusing configuration for indoor optical wireless access," *IEEE Transactions on Communications*, vol. 57, no. 6, pp. 1765–1775, June 2009.
- [56] R. M. Shelby, J. Hoffnagle, G. W. Burr *et al.*, "Pixel-matched holographic data storage with megabit pages," *Optics Letters*, vol. 22, no. 19, pp. 1509–1511, October 1997.
- [57] G. W. Burr, J. Ashley, H. Coufal *et al.*, "Modulation coding for pixel-matched holographic data storage," *Optics Letters*, vol. 22, no. 9, pp. 639–641, May 1997.
- [58] S. Nabavi and B. V. K. V. Kumar, "Detection methods for holographic data storage," in *Proceedings of the IEEE Optical Data Storage*, Montreal, QC, Canada, April 2006, pp. 156–158.
- [59] D. V. Plant and A. G. Kirk, "Optical interconnects at the chip and board level: Challenges and solutions," *Proceedings of the IEEE*, vol. 88, no. 6, pp. 806–818, 2000.
- [60] J. Bristow, J. Lehman, M. Hibbs-Brenner, and Y. Liu, "Applying optical interconnects to electronic systems: Promise vs. practicality," *Proceedings of the Fourth International Conference on Massively Parallel Processing Using Optical Interconnections.*, pp. 54–59, 22–24 June 1997.

- [61] R. Villn, S. Voloshynskiy, O. Koval, and T. Pun, "Multilevel 2-D bar codes: Towards high capacity storage modules for multimedia security and management," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 4, pp. 405–420, Dec. 2006.
- [62] W. Claycomb and D. Shin, "Using a two dimensional colored barcode solution for authentication in pervasive computing," in *Proceeding of IEEE ACS/ IEEE International Conference on Pervasive Services*, Lyon, France, June 26–29 2006, pp. 173–180.
- [63] D. C. O'Brien, S. Quasem, S. Zikic, and G. E. Faulkner, "Multiple input multiple output systems for optical wireless: Challenges and possibilities," *Proceeding of SPIE*, vol. 6304, 2006.
- [64] S. Hranilovic and F. R. Kschischang, "A pixelated MIMO wireless optical communication system," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 12, no. 4, pp. 859–874, July/Aug. 2006.
- [65] M. D. A. Mohamed and S. Hranilovic, "Two-dimensional binary half-toned optical intensity channels," *IET Communications*, vol. 2, no. 1, pp. 11–17, January 2008.
- [66] H. Coufal, "Holographic data storage or reliability; a status report," in *Optical Data Storage Topical Meeting, 1997. ODS. Conference Digest*, 7–9 Apr. 1997, pp. 46–47.
- [67] "InPhase Technologies," www.inphase-technologies.com, 2007.
- [68] J. F. Heanue, M. Bashaw, and L. Hesselink, "Volume holographic storage and retrieval of digital data," *Science*, vol. 265, no. 5173, pp. 749–752, Aug. 1994.
- [69] S. S. Orlov, W. Phillips, E. Bjornson, L. Hesselink, and R. Okas, "10 gigabit/second sustained optical data transfer rate from a holographic disk digital data storage system," *Presented at the OSA Annual Meeting 2000, Paper MK3.*, 2000.
- [70] S. S. Orlov, "Volume holographic data storage," *Communications of the ACM*, vol. 43, no. 11, pp. 46–54, 2000.
- [71] J. Ashley, M. Bernal, G. W. Burr *et al.*, "Holographic data storage," *IBM Journal of Research and Development*, vol. 44, no. 3, pp. 341–368, May 2000.
- [72] A. Dabbo, "Design of wireless optical MIMO links," Master's thesis, Department of Electrical and Computer Engineering, McMaster University, 2009.
- [73] E. Bisailon, D. F. Brosseau, T. Yamamoto *et al.*, "Free-space optical link with spatial redundancy for misalignment tolerance," *IEEE Photonic Technology Letters*, vol. 14, no. 4, pp. 859–874, 2002.
- [74] J. F. Heanue, K. Gurkan, and L. Hesselink, "Signal detection for page-access optical memories with inter-symbol interference," *Journal of Applied Optics*, vol. 35, no. 14, pp. 2431–2438, 10 May 1996.
- [75] J. K. Nelson, A. C. Singer, and U. Madhow, "Multi-directional decision feedback for 2D equalization," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing.*, vol. 4, pp. iv (921–924), 17–21 May 2004.
- [76] C. Y. Chen and T. D. Chiueh, "A low-complexity high-performance modulation code for holographic data storage," in *14th IEEE International Conference on Electronics, Circuits and Systems*, 11–14 December 2007, pp. 788–791.
- [77] M. Ayres, A. Hoskins, and K. Curtis, "Image oversampling for page-oriented optical data storage," *Applied Optics*, vol. 45, no. 11, pp. 2459 – 2464, 10 Apr. 2006.
- [78] H. Ohtoh and M. Yamamoto, "Data position detection analysis for 1.X time oversampled image in holographic data storage," in *Conference on Lasers and Electro-Optics*, Pacific Rim, 26–31 August 2007, pp. 1–2.
- [79] D. A. B. Miller, "Rationale and challenges for optical interconnects to electronic chips," *Proceedings of IEEE*, vol. 88, no. 6, pp. 728–749, June 2000.

- [80] D. Lammers, "Intel prepares for optical-interconnect future," *EE Times*, 3 August 2004.
- [81] D. V. Plant, B. Robertson, H. S. Hinton *et al.*, "4 × 4 vertical-cavity surface-emitting laser (VCSEL) and metal–semiconductor–metal (MSM) optical backplane demonstrator system," *Applied Optics*, vol. 35, no. 32, pp. 6365–6368, 1996.
- [82] D. V. Plant, M. B. Venditti, E. Laprise *et al.*, "256-channel bidirectional optical interconnect using VCSELs and photodiodes on CMOS," *Journal of Lightwave Technology*, vol. 19, no. 8, pp. 1093–1103, Aug. 2001.
- [83] M. Naruse, S. Yamamoto, and M. Ishikawa, "Real-time active alignment demonstration for free-space optical interconnections," *IEEE Photonic Technology Letters*, vol. 13, no. 11, pp. 1257–1259, Nov. 2001.
- [84] S. Lyons, "Two-dimensional barcodes for mobile phones," Master's thesis, Department of Electrical and Computer Engineering, University of Toronto, 2009.
- [85] S. Lyons and F. Kschischang, "Two-dimensional barcodes for mobile phones," in *25th Biennial Symposium on Communication*, May 12–14, 2010, pp. 344–347.
- [86] www.terryburton.co.uk/barcodewriter/generator/, (last accessed: Sept. 12, 2010).
- [87] M. D. A. Mohamed, A. Dabbo, and S. Hranilovic, "MIMO optical wireless channel using half-toning," in *IEEE International Conference on Communications ICC*, Beijing, China, May 2008, pp. 1306–1364.
- [88] M. R. H. Mondal, K. Panta, and J. Armstrong, "Performance of two dimensional asymmetrically clipped optical OFDM," in *2010 IEEE Globecom Workshop on Optical Wireless Communications*, pp. 995–999, Dec. 2010.
- [89] D. Takase and T. Ohtsuki, "Spatial multiplexing in optical wireless MIMO communications over indoor environment," *The Institute of Electronic, Information, and Communication Engineers (IEICE) Transactions*, vol. E89-B, no. 4, pp. 1364–1371, Apr. 2006.
- [90] L. Zeng, D. C. O'Brien, H. L. Minh *et al.*, "High data rate multiple input multiple output (MIMO) optical wireless communications using white LED lighting," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 9, pp. 1654–1662, Dec. 2009.
- [91] A. H. Azhar, T.-A. Tran, and D. O'Brien, "Demonstration of high-speed data transmission using MIMO-OFDM visible light communications," in *2010 IEEE Globecom Workshop on Optical Wireless Communications*, pp. 1052–1054, Dec. 2010.
- [92] J. Goodman, *Introduction to Fourier Optics*, 2nd edn. McGraw-Hill, 1996.
- [93] S. Hranilovic and F. R. Kschischang, "Short-range wireless optical communication using pixelated transmitters and imaging receivers," in *Proceedings of the IEEE International Conference on Communications*, vol. 2, Paris, France, 20–24 June 2004, pp. 891–895.
- [94] A. Dabbo and S. Hranilovic, "Receiver design for wireless optical MIMO channels with magnification," in *Proceedings of the 10th International Conference on Telecommunications, ConTEL 2009*, Zagreb, Croatia, 8–10 June 2009, pp. 51–58.
- [95] S. Nabavi and B. V. K. V. Kumar, "Detection methods for holographic data storage," in *Proceedings of the IEEE Optical Data Storage Topical Meeting*, Montreal, QC, Canada, 23–26 Apr. 2006, pp. 156–158.
- [96] P. H. Siegel, "Information-theoretic limits of two-dimensional optical recording channels," in *Proceedings of the IEEE Optical Data Storage Topical Meeting*, Montreal, QC, Canada, 23–26 Apr. 2006, pp. 165–167.
- [97] A. Peled and A. Ruiz, "Frequency domain data transmission using reduced computational complexity algorithms," in *Proceedings of the IEEE International Conference on*

-
- Acoustics, Speech, and Signal Processing*, vol. 5, Denver, CO, USA, 9–11 Apr. 1980, pp. 964–967.
- [98] J. G. Proakis, *Digital Communications*, 4th edn. McGraw-Hill, 2001.
- [99] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
- [100] R. W. Floyd and L. Steinberg, “An adaptive algorithm for spatial gray scale,” in *SID Symposium Digest, Society for Information Display*, 1975, pp. 36–37.
- [101] D. Anastassiou, “Error diffusion coding for A/D conversion,” *IEEE Transactions on Circuits and Systems*, vol. 36, no. 9, pp. 1175–1186, Sept. 1989.
- [102] Mikrotron GmbH. URL: www.mikrotron.de.
- [103] Matrox Imaging. URL: www.matrox.com.
- [104] A. Dabbo and S. Hranilovic, “Multilevel error diffusion for wireless optical MIMO channels,” in *Proceedings of the 24th Biennial Symposium on Communications*, Kingston, ON, Canada, 24–26 June 2008, pp. 208–211.

7 Channel capacity

Amos Lapidoth, Stefan M. Moser, and Michèle Wigger

7.1 Introduction and channel models

Optical communications take place at high frequencies at which it is sometimes difficult to modulate the phase and frequency of the transmitted signal. Instead, some optical systems modulate the intensity of the signal. Such systems are consequently fundamentally different from standard (e.g., mobile) wireless systems. For example, the channel input cannot be negative because it corresponds to the instantaneous light intensity. Also, since the light intensity, and thus the input, is proportional to the instantaneous optical power, the average-power constraints apply *linearly* to the input and not, as is usual in standard radio communications, to its square.

In this chapter we focus on communication systems that employ *pulse amplitude modulation (PAM)*, which in the case of optical communication is called *pulse intensity modulation*. In such systems the transmitter modulates the information bits onto continuous-time pulses of duration T , and the receiver preprocesses the incoming continuous-time signal by integrating it over nonoverlapping intervals of length T . Such continuous-time systems can be modeled as *discrete-time channels* where the (discrete) time k input and output correspond to the integrals of the continuous-time transmitted and received signals (i.e., optical intensities) from kT to $(k + 1)T$. Note that for such discrete-time systems, the achieved data rate is not measured in bits (or nats) per second, but in *bits (or nats) per channel use*.

We discuss three different discrete-time, pulse intensity modulated, optical channel models: the *discrete-time Poisson channel*, the *free-space optical intensity channel*, and the *optical intensity channel with input-dependent Gaussian noise*.

7.1.1 Discrete-time Poisson channel

In the continuous-time Poisson channel, the channel output is the sum of two independent stochastic processes. Conditional on the input, the first is a *conditional Poisson process* – also known as a *doubly stochastic Poisson process* – whose rate function is equal to the input. The second is a homogeneous Poisson process that accounts for spurious counts resulting from background radiation and receiver noise. If we employ

pulse intensity modulation where the input is constant over length- T time intervals of the form $[kT, (k+1)T)$ for integers k , then the number of counts registered in the intervals $\{[kT, (k+1)T)\}$ form a sufficient statistic for guessing the input waveform based on the channel output. This leads to the discrete-time Poisson channel.

In the discrete-time Poisson channel, we model the input as a nonnegative sequence $\{x_k\}$, where x_k corresponds to the integral of the continuous-time input over the interval $[kT, (k+1)T)$. Thus, since we envision pulse intensity modulation, x_k is the product of T by the instantaneous power of the transmitted signal during the interval $[kT, (k+1)T)$. The output is a sequence $\{Y_k\}$, where Y_k denotes the number of counts registered during the interval $[kT, (k+1)T)$. We assume that, conditional on the transmitted waveform, Y_k has a Poisson distribution of mean $x_k + \lambda$, where λ is the expected number of spurious counts during the interval $[kT, (k+1)T)$. We sometimes refer to λ as *dark current*. The discrete-time Poisson channel is thus a memoryless discrete-time channel whose output Y takes value in the set of nonnegative integers \mathbb{Z}_0^+ and whose input X takes value in the set of nonnegative real numbers \mathbb{R}_0^+ . Conditional on the nonnegative input x , the output is Poisson with mean $x + \lambda$:

$$W(y|x) = e^{-(x+\lambda)} \frac{(x+\lambda)^y}{y!}, \quad y \in \mathbb{Z}_0^+, x \in \mathbb{R}_0^+. \quad (7.1)$$

For safety reasons and practical implementation considerations, the optical average power and the maximal instantaneous power must be constrained. We therefore impose the constraints

$$E[X] \leq \mathcal{E}, \quad (7.2)$$

$$\Pr[X > \mathcal{A}] = 0, \quad (7.3)$$

for some fixed parameters $\mathcal{E}, \mathcal{A} > 0$, and refer to \mathcal{E} as the allowed average power and to \mathcal{A} as the allowed peak power.¹ Note again that the average-power constraint is on the expectation of the channel input and not on its square.

We denote the ratio of the allowed average power to the allowed peak power by α ,

$$\alpha \triangleq \frac{\mathcal{E}}{\mathcal{A}}, \quad (7.4)$$

where $0 < \alpha \leq 1$. Note that for $\alpha = 1$ the average-power constraint is inactive in the sense that it has no influence on the system and is automatically satisfied whenever the peak-power constraint is satisfied. Thus, $\alpha = 1$ corresponds to the case with only a peak-power constraint. Similarly, $\alpha \ll 1$ corresponds to a stringent average-power constraint and only a mild peak-power constraint.

In our discussion of capacity, we shall pay particular attention to the low-power asymptotic regime. Since the input power of the discrete-time Poisson channel

¹ A peak-power constraint \mathcal{A} and an average-power constraint \mathcal{E} in the discrete-time model correspond to a peak-power constraint \mathcal{A}/T and an average-power constraint \mathcal{E}/T in the corresponding continuous-time model. Also, “power in the continuous-time model” refers to the “average number of photons emitted per second.” To express this in energy per second we must multiply by the energy of each photon $\hbar\omega$, where ω (in radians per second) is the transmitter’s frequency. We can also divide by the detector’s quantum efficiency, if it is not 1.

corresponds to the product of the light intensity and the pulse duration in the continuous-time model, the low-power regime where $\mathcal{E} \downarrow 0$ or $\mathcal{A} \downarrow 0$ can come about in two different ways: either from the continuous-time scenario with pulses of very short duration $T \ll 1$,² or from a continuous-time scenario with fixed pulse duration and very low transmit power. In the former case the number of spurious counts in each (short) pulse interval is also very small, so the dark current λ is also very small, whereas in the latter scenario the number of spurious counts does not decrease with the power. In the same way, the high-power regime, which will also be of interest to us, can come about from two different scenarios depending on whether the pulse duration is held fixed or not.

7.1.2 Free-space optical intensity channel

In our second model we account for three sources of noise: thermal noise of the post-detection circuitry following the photo detection, which is well-modeled by a Gaussian distribution; relative-intensity noise, which models random intensity fluctuations inherent to low-cost laser sources and which can also be assumed to be Gaussian [13]; and shot noise caused by the ambient light. The shot noise has an appreciable impact only at large intensities where its distribution tends to be Gaussian and in a first approximation can be assumed to be independent of the signal. At low intensity the thermal noise is the limiting factor. The sum of these three noise sources can be well-modeled by independent and memoryless additive Gaussian noise,

$$Y = x + Z, \quad (7.5)$$

where Y denotes the channel output; where x denotes the channel input, which is proportional to the optical intensity and therefore cannot be negative,

$$x \in \mathbb{R}_0^+; \quad (7.6)$$

and where the additive noise is zero-mean Gaussian with variance σ^2 . It is important to note that, unlike the input, the output Y may be negative since the noise introduced at the receiver can be negative. The input is again constrained in both its average power (7.2) and its peak power (7.3), with α denoting the average-to-peak power ratio (7.4).

This model is sometimes used to describe optical communication through air and is therefore known as the *free-space optical intensity channel*. Another name is *optical direct-detection channel with Gaussian post-detection noise*. See [13], [16] for more details.

For a channel model that also accounts for the random components of the ambient light and for the effects caused by an optical pre-amplifier at the receiver, see [18] and [17].

² This corresponds to the wide-band regime where by “wide-band” we mean that the communication bandwidth, i.e., the reciprocal of the pulse duration T , is large enough so that $\eta PT / \hbar\omega \ll 1$, but this bandwidth is still much smaller than the optical center frequency ω . Once the bandwidth becomes comparable to the optical center frequency, photon-flux is no longer proportional to input power, and therefore our channel model becomes inadequate.

7.1.3 Optical intensity channel with input-dependent Gaussian noise

In our third model we still assume that the signal at the receiver is impaired by additive Gaussian noise, however, we now assume that the variance of the noise depends on the channel input. Thereby we try to better reflect the implicit dependence of the random distortion in an optical channel on the signal itself. This model thus combines some of the features of the previous two models: the assumption of Gaussian noise of the free-space optical intensity channel and the double-stochastic (signal-dependent) nature of the discrete-time Poisson channel.

The model is called the *optical intensity channel with input-dependent Gaussian noise*, and its conditional channel law is

$$W(y|x) = \frac{1}{\sqrt{2\pi(\sigma^2 + x)}} e^{-\frac{(y-x)^2}{2(\sigma^2 + x)}}, \quad y \in \mathbb{R}, x \geq 0. \quad (7.7)$$

Thus, the channel is described by

$$Y = x + \sqrt{x}Z_1 + Z_0, \quad (7.8)$$

where Y denotes the channel output, $x \geq 0$ denotes the channel input, Z_0 is a zero-mean, variance- σ^2 Gaussian random variable describing the input-independent noise, and Z_1 is a zero-mean, unit-variance Gaussian random variable describing the input-dependent noise. Here Z_0 and Z_1 are independent. Note that there is no loss in generality in assuming that Z_1 is of unit variance. In fact, by appropriately scaling the channel input and output, every channel of the form (7.8) can be transformed into a channel where the input-dependent noise part has unit variance.

As before, we consider both an average-power constraint (7.2) and a peak-power constraint (7.3), and we denote the average-to-peak power ratio by α (7.4).

7.1.4 Channel capacity

Loosely speaking, the capacity of a channel is the highest rate at which one can communicate reliably over the channel. For all rates below channel capacity it is possible to achieve arbitrarily small probability of error by using channel codes of sufficiently large block-length, whereas for all rates above capacity this is impossible. The channel capacity thus describes the quality of the channel and provides an important benchmark for practical systems. For discrete-time channels, the capacity is typically measured in nats per channel use.

The capacity depends on the channel model and on the given input constraints. We denote the capacity with allowed peak power \mathcal{A} and allowed average power \mathcal{E} by $\mathcal{C}(\mathcal{A}, \mathcal{E})$. Shannon proved [29] that

$$\mathcal{C}(\mathcal{A}, \mathcal{E}) = \sup_Q I(Q, W), \quad (7.9)$$

where $I(Q, W) = I(X; Y)$ denotes the mutual information [4] between the input X and the output Y of a channel of law $W(\cdot|\cdot)$ when the input X is distributed according to Q ; and where the supremum is over all probability distributions Q that satisfy the given input constraints. For our channel models the supremum is thus over all probability distributions Q on \mathbb{R}_0^+ that satisfy the constraints (7.2) and (7.3).

When only an average-power constraint is imposed, capacity is denoted by $\mathcal{C}(\mathcal{E})$ and is given as in (7.9), but where the supremum is taken over all laws Q on $X \geq 0$ satisfying (7.2) only.

Unfortunately, it is analytically and numerically very difficult to evaluate (7.9). We shall therefore only present upper and lower bounds on channel capacity and asymptotic expansions.

To distinguish between the capacities of the three considered channel models we add a subscript to the capacity indicating the channel. Thus, the capacity of the free-space optical intensity channel under a peak-power constraint \mathcal{A} and an average-power constraint \mathcal{E} is denoted by $\mathcal{C}_{\text{FreeSpace}}(\mathcal{A}, \mathcal{E})$, and the capacity of this channel under only an average-power constraint by $\mathcal{C}_{\text{FreeSpace}}(\mathcal{E})$. Similarly, the capacities of the optical intensity channel with input-dependent Gaussian noise under an average- and a peak-power constraint and under only an average-power constraint are denoted by $\mathcal{C}_{\text{DepNoise}}(\mathcal{A}, \mathcal{E})$ and $\mathcal{C}_{\text{DepNoise}}(\mathcal{E})$, respectively.

For the discrete-time Poisson channel we make the dependence of the capacity on the dark current λ explicit by writing $\mathcal{C}_{\text{Poisson}}(\lambda, \mathcal{A}, \mathcal{E})$ in the case of both an average- and a peak-power constraint, and $\mathcal{C}_{\text{Poisson}}(\lambda, \mathcal{E})$ in the case of an average-power constraint only.

Note that in the following the logarithmic function $\log(\cdot)$ stands for the natural logarithmic function, and all results are expressed in the unit of *nats*. To display the results in *bits*, all expressions have to be divided by $\log 2$.

7.2 Capacity results

For each of the above channels the capacity can be computed numerically using the Karush–Kuhn–Tucker conditions [11] and the fact that the capacity-achieving input distributions are discrete with a finite number of mass points [27], [3]. However, an analytic expression for the capacity with general peak- and average-power constraints seems out of reach.

Our main results in this chapter are analytic expressions for the asymptotic behavior of the channel capacity in the asymptotic high-power regime where both the average-power and the peak-power constraints are very large and in the asymptotic low-power regime where at least one of the power constraints is very small. In the asymptotic high-power regime our goal is to determine the capacity up to a gap that tends to 0 as the powers tend to infinity, and in the asymptotic low-power regime our goal is to determine the capacity up to a gap that tends to zero faster than capacity. We also present bounds on the capacity for arbitrary (nonasymptotic) average- and peak-power constraints.

7.2.1 Discrete-time Poisson channel

While the capacity of the discrete-time Poisson channel is unknown, the capacity of the continuous-time Poisson channel where the input signal is not restricted to be pulse intensity modulated has been derived exactly: the case with a peak-power constraint only was solved by Kabanov [15]; the more general case of peak- and average-power constraints was treated by Davis [6]; Wyner [32] found the reliability function of the channel; Frey [10], [9] studied the capacity of the continuous-time Poisson channel under an L -norm constraint; and Shamai and Lapidot [28] investigated the capacity under spectral constraints. The capacity of (single- or multiple-antenna) Poisson fading channels has been studied, e.g., in [2], [12].

In the following subsections we present results on the asymptotic high- and low-power capacity of the discrete-time Poisson channel. In the high-power regime, we present results on the scenario with both an average- and a peak-power constraint; on the scenario with only a peak-power constraint; and on the scenario with only an average-power constraint. In the scenario with both average-power and peak-power constraints the two constraints are assumed to grow proportionally, i.e., with a constant average-to-peak power ratio α (see (7.4)). For all three scenarios the precise asymptotic high-power capacity is known when the dark current is held fixed (i.e., when it does not tend to infinity with the power). When the dark current grows proportionally with the power, then only the case where no peak-power constraint is imposed has been studied. For this setup upper and lower bounds on the asymptotic high-power capacity have been proposed, but they do not coincide.

The asymptotic low-power capacity is discussed for a scenario where the dark current λ is fixed and positive, and for a scenario where it tends to 0 proportionally to the average-power constraint \mathcal{E} , i.e., with a constant ratio

$$c \triangleq \frac{\lambda}{\mathcal{E}}. \quad (7.10)$$

Both scenarios are examined with a fixed peak-power constraint (i.e., the peak-power constraint does not tend to 0) or without peak-power constraint.

7.2.1.1 High-power results

Before presenting the theorem, we define for given $\alpha \in (0, \frac{1}{3})$ the parameter ν^* as the unique solution to the equation

$$\alpha = \frac{1}{2\nu} - \frac{e^{-\nu}}{\sqrt{\nu}\sqrt{\pi}\operatorname{erf}(\sqrt{\nu})}. \quad (7.11)$$

Here $\operatorname{erf}(\cdot)$ denotes the error function defined as

$$\operatorname{erf}(\xi) \triangleq \frac{2}{\sqrt{\pi}} \int_0^\xi e^{-t^2} dt, \quad \forall \xi \in \mathbb{R}. \quad (7.12)$$

Note that the function $\nu \mapsto \frac{1}{2\nu} - \frac{e^{-\nu}}{\sqrt{\nu}\sqrt{\pi}\operatorname{erf}(\sqrt{\nu})}$ is monotonically decreasing in $[0, \infty)$ and tends to $\frac{1}{3}$ for $\nu \downarrow 0$ and to 0 for $\nu \uparrow \infty$. Hence, for $\alpha \in (0, \frac{1}{3})$ a solution always exists and is unique.

THEOREM 7.1 (Constant dark current, [21]) *If the channel is peak-power constrained to \mathcal{A} and average-power constrained to $\mathcal{E} = \alpha\mathcal{A}$ for some $\alpha \in (0, \frac{1}{3})$, then*

$$\mathcal{C}_{\text{Poisson}}(\lambda, \mathcal{A}, \alpha\mathcal{A}) = \frac{1}{2} \log \left(\frac{2\mathcal{A}}{\pi e} \right) - (1 - \alpha)v^* - \log(1 - 2\alpha v^*) + o_{\mathcal{A}}(1), \quad (7.13)$$

where $o_{\mathcal{A}}(1)$ is a term that tends to 0 as the power \mathcal{A} tends to infinity and where v^* is the unique solution to (7.11).

If the channel is peak-power constrained to \mathcal{A} and average-power constrained to $\mathcal{E} = \alpha\mathcal{A}$ for some $\alpha \in [\frac{1}{3}, 1]$, or if the channel is only peak-power constrained to \mathcal{A} (which corresponds to $\alpha = 1$), then

$$\mathcal{C}_{\text{Poisson}}(\lambda, \mathcal{A}, \alpha\mathcal{A}) = \frac{1}{2} \log \left(\frac{2\mathcal{A}}{\pi e} \right) + o_{\mathcal{A}}(1). \quad (7.14)$$

If the channel is only average-power constrained to \mathcal{E} , then

$$\mathcal{C}_{\text{Poisson}}(\lambda, \mathcal{E}) = \frac{1}{2} \log \mathcal{E} + o_{\mathcal{E}}(1). \quad (7.15)$$

The high-power asymptotes (7.13), (7.14), and (7.15) do not depend on the dark current λ . Also, for $\alpha \geq \frac{1}{3}$ the asymptote (7.14) does not depend on the average-to-peak power ratio α . Thus, in the high-power asymptote, the average-power constraint is not active if $\mathcal{E} \geq \frac{\mathcal{A}}{3}$.

Notice that asymptotes (7.14) and (7.15) coincide with the high-power capacity-asymptotes of the classical additive white Gaussian noise (AWGN) channel when its noise variance $\sigma^2 = 1$ and when its channel input is either peak-power constrained to \mathcal{A} , i.e., $0 \leq X^2 \leq \mathcal{A}$, or average-power constrained to \mathcal{E} , i.e., $E[X^2] \leq \mathcal{E}$.³

Comparing (7.13) and (7.14) we observe that if *in addition to* a peak-power constraint \mathcal{A} also an average-power constraint $\mathcal{E} = \alpha\mathcal{A}$ is imposed, for $\alpha \in (0, \frac{1}{3})$, then the asymptotic high-power capacity is reduced by $(1 - \alpha)v^* + \log(1 - 2\alpha v^*)$. Similarly, comparing (7.14) and (7.15), we notice that if *instead of* a (stronger) peak-power constraint \mathcal{A} a (weaker) average-power constraint $\mathcal{E} = \mathcal{A}$ is imposed, then the high-power capacity is increased by $\frac{1}{2} \log \left(\frac{\pi e}{2} \right)$.

A different high-power asymptote was investigated by Brady and Verdú in [1]. They focused on the special case with only an average-power constraint and studied the asymptotic behavior of capacity when the dark current grows proportionally with the average-power constraint.

THEOREM 7.2 (Dark current proportional to \mathcal{E} , [1]) *If only the average power is constrained to \mathcal{E} and the dark current λ grows proportionally with \mathcal{E} , i.e., $\lambda = c\mathcal{E}$ for some $c > 0$, then the following two bounds hold: given $\epsilon > 0$ there exists an \mathcal{E}_ϵ such that for all $\mathcal{E} > \mathcal{E}_\epsilon$ the capacity is bounded by*

³ Recall that for the classical AWGN channel the standard input-power constraints act on the square of the input X^2 , and not directly on the input X .

$$\mathcal{C}_{\text{Poisson}}(\lambda = c\mathcal{E}, \mathcal{E}) \geq \frac{1}{2} \log \left(\frac{\mathcal{E}}{2\pi} \right) - \frac{1}{2} \log(1+c) - \epsilon, \quad (7.16)$$

$$\begin{aligned} \mathcal{C}_{\text{Poisson}}(\lambda = c\mathcal{E}, \mathcal{E}) \leq & \frac{1}{2} \log \left(\frac{\mathcal{E}}{2\pi} \right) + \log \left(\frac{1}{\sqrt{c}} \left(1 + \frac{1}{\mathcal{E}_\epsilon} \right) + \sqrt{c} \right) \\ & + 1 + \log \left(\frac{3}{2} \right) + \epsilon. \end{aligned} \quad (7.17)$$

7.2.1.2 Low-power results

The first low-power result treats the case where the dark current λ is fixed and positive.

THEOREM 7.3 (Constant nonzero dark current, [23]) *If a peak-power constraint \mathcal{A} and an average-power constraint \mathcal{E} are imposed, then for fixed dark current $\lambda > 0$:*

$$\lim_{\mathcal{E} \downarrow 0} \frac{\mathcal{C}_{\text{Poisson}}(\lambda, \mathcal{E}, \mathcal{A})}{\mathcal{E}} = \left(1 + \frac{\lambda}{\mathcal{A}} \right) \log \left(1 + \frac{\mathcal{A}}{\lambda} \right) - 1. \quad (7.18)$$

If only an average-power constraint \mathcal{E} is imposed, then for fixed dark current $\lambda > 0$:

$$\lim_{\mathcal{E} \downarrow 0} \frac{\mathcal{C}_{\text{Poisson}}(\lambda, \mathcal{E})}{\mathcal{E} \log \log \left(\frac{1}{\mathcal{E}} \right)} \geq \frac{1}{2}, \quad (7.19)$$

$$\lim_{\mathcal{E} \downarrow 0} \frac{\mathcal{C}_{\text{Poisson}}(\lambda, \mathcal{E})}{\mathcal{E} \log \log \left(\frac{1}{\mathcal{E}} \right)} \leq 2. \quad (7.20)$$

When only an average-power constraint \mathcal{E} is imposed, then our upper and lower bounds on the asymptotic low-power capacity are not tight, but only exhibit similar behavior.

Both (7.18) and (7.19) can be achieved with on–off signalling. In the case of (7.18), the “on” signal is equal to \mathcal{A} ; while in the case of (7.19), the “on” signal tends to infinity as a constant times $\log \left(\frac{1}{\mathcal{E}} \right)$.

For constant (nonzero) dark current λ , the asymptotic low-power capacity depends heavily on the peak-power constraint \mathcal{A} . In particular, it is linear in the average-power constraint \mathcal{E} in the presence of a peak-power constraint, and it is proportional to $\mathcal{E} \log \log \left(\frac{1}{\mathcal{E}} \right)$ in its absence.

The second low-power result treats the case where the dark current is proportional to the average-power constraint \mathcal{E} , and thus tends to 0 with $\mathcal{E} \downarrow 0$.

THEOREM 7.4 (Dark current proportional to \mathcal{E} , [23]) *If a peak-power constraint \mathcal{A} and an average-power constraint \mathcal{E} are imposed, then for a dark-current λ that is proportional⁴ to \mathcal{E} , i.e., $\lambda = c\mathcal{E}$ for some real number $c \geq 0$:*

$$\lim_{\mathcal{E} \downarrow 0} \frac{\mathcal{C}_{\text{Poisson}}(\lambda = c\mathcal{E}, \mathcal{E}, \mathcal{A})}{\mathcal{E} \log \left(\frac{1}{\mathcal{E}} \right)} = 1. \quad (7.21)$$

The result also holds in the absence of a peak-power constraint.

⁴ Note that this includes the case of zero dark current $\lambda = 0$.

Note that (7.21) holds irrespective of the allowed (positive) peak power \mathcal{A} . In fact, (7.21) can be achieved using on–off signalling, where the “on” signal is an arbitrary small but positive constant (i.e., it does not tend to 0 as $\mathcal{E} \downarrow 0$).

Asymptote (7.21) does not depend on the proportionality constant $c \geq 0$. In particular, it does not depend on whether the dark current λ is zero or whether it only tends to 0 with $\mathcal{E} \downarrow 0$. Intuitively this is because – irrespective of the constant c – for sufficiently small \mathcal{E} our constant “on” signal dominates the dark-current floor $\lambda = c\mathcal{E}$.

We conclude this section by discussing the implications of Theorem 7.4 for the continuous-time Poisson channel. We focus on the case where the peak power is not constrained. As mentioned in Section 7.1.1, the discrete-time model studied in Theorem 7.4 can be used to describe a pulse intensity modulated system over a *continuous-time* Poisson channel where the pulse duration T tends to 0 and the average power $\frac{\mathcal{E}}{T}$ (i.e., the average intensity) is constant. Theorem 7.4 thus implies that in such a continuous-time system, as the pulse duration T tends to 0, the data rate $\frac{C_{\text{Poisson}}}{T}$ tends to infinity like a constant times $\log\left(\frac{1}{T}\right)$. This observation is consistent with the fact that in the absence of a peak-power constraint the capacity of a continuous-time Poisson channel is infinite.

7.2.1.3 Nonasymptotic results

The following lower bounds hold for arbitrary dark current and arbitrary peak- and/or average-power constraints.

THEOREM 7.5 (Arbitrary dark current, [21]) *If the channel is peak-power constrained to \mathcal{A} and average-power constrained to $\mathcal{E} = \alpha\mathcal{A}$ for some $\alpha \in (0, \frac{1}{3})$, then*

$$\begin{aligned} C_{\text{Poisson}}(\lambda, \mathcal{A}, \alpha\mathcal{A}) &\geq \frac{1}{2} \log \mathcal{A} - (1 - \alpha)v^* - \log\left(\frac{1}{2} - \alpha v^*\right) \\ &\quad - e^{v^*} \left(\frac{1}{2} - \alpha v^*\right) \left[\log\left(1 + \frac{\lambda + \frac{1}{12}}{\mathcal{A}}\right) \right. \\ &\quad \left. + 2\sqrt{\frac{\lambda + \frac{1}{12}}{\mathcal{A}}} \arctan\left(\sqrt{\frac{\mathcal{A}}{\lambda + \frac{1}{12}}}\right) \right] \\ &\quad + (\mathcal{E} + 1) \log\left(1 + \frac{1}{\mathcal{E}}\right) - 1 - \frac{1}{2} \log(2\pi e), \end{aligned} \quad (7.22)$$

where v^* is the unique solution to (7.11).

If the channel is peak-power constrained to \mathcal{A} and average-power constrained to $\mathcal{E} = \alpha\mathcal{A}$ for some $\alpha \in [\frac{1}{3}, 1]$, or if the channel is only peak-power constrained to \mathcal{A} (which corresponds to $\alpha = 1$), then

$$C_{\text{Poisson}}(\lambda, \mathcal{A}, \alpha\mathcal{A}) \geq \frac{1}{2} \log \mathcal{A} + \left(\frac{\mathcal{A}}{3} + 1\right) \log\left(1 + \frac{3}{\mathcal{A}}\right) - 1 \quad (7.23)$$

$$- \sqrt{\frac{\lambda + \frac{1}{12}}{\mathcal{A}}} \left(\frac{\pi}{4} + \frac{1}{2} \log 2\right) - \frac{1}{2} \log\left(\frac{\pi e}{2}\right). \quad (7.24)$$

If the channel is only average-power constrained to \mathcal{E} , then

$$C_{\text{Poisson}}(\lambda, \mathcal{E}) \geq \frac{1}{2} \log \mathcal{E} - \sqrt{\frac{\pi \left(\lambda + \frac{1}{12} \right)}{2\mathcal{E}}} + (\mathcal{E} + 1) \log \left(1 + \frac{1}{\mathcal{E}} \right) - 1. \quad (7.25)$$

Martinez [24] proved an upper bound and a tighter lower bound on the capacity of the discrete-time Poisson channel for the special case without dark current and under an average-power constraint only.

THEOREM 7.6 (Zero dark current, [24]) *If only the average power is constrained to \mathcal{E} and the dark current $\lambda = 0$, then*

$$C_{\text{Poisson}}(\lambda = 0, \mathcal{E}) \geq \frac{1}{2} \log(1 + \mathcal{E}), \quad (7.26)$$

$$C_{\text{Poisson}}(\lambda = 0, \mathcal{E}) \leq \left(\mathcal{E} + \frac{1}{2} \right) \log \left(\mathcal{E} + \frac{1}{2} \right) - \mathcal{E} \log \mathcal{E} - \frac{1}{2} + \log \left(1 + \frac{\sqrt{2e} - 1}{\sqrt{1 + 2\mathcal{E}}} \right). \quad (7.27)$$

For the case of only an average-power constraint and zero dark current, the bounds of Theorem 7.5 and Theorem 7.6 are shown in Figure 7.1.

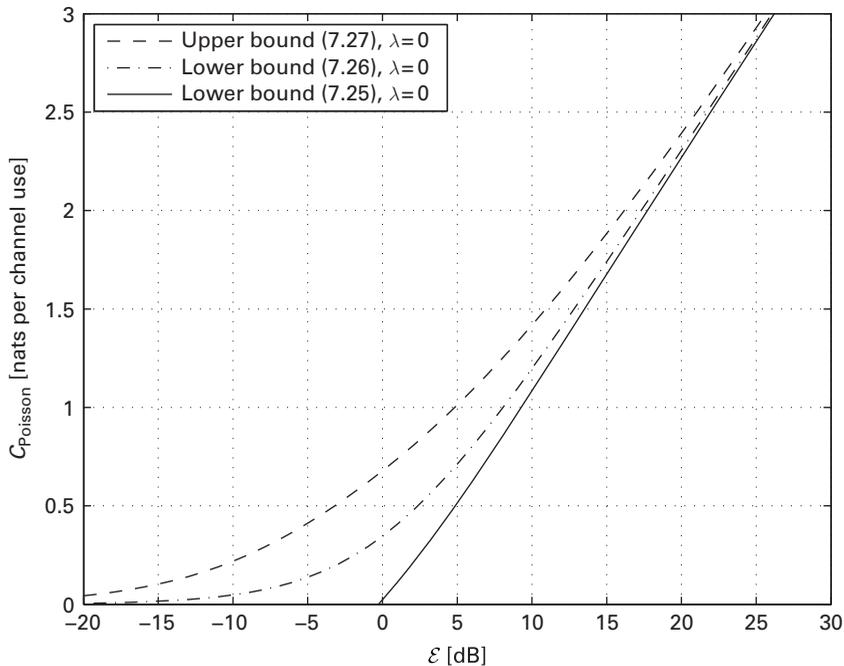


Figure 7.1 This plot depicts some nonasymptotic bounds on the capacity of the discrete-time Poisson channel with only an average-power constraint for the special case of no dark current $\lambda = 0$. The horizontal axis is measured in dB where $\mathcal{E} [\text{dB}] = 10 \log_{10} \mathcal{E}$.

7.2.2 Free-space optical intensity channel

For the free-space optical intensity channel we present results on the asymptotic capacity in the high- and low-power regime, as well as firm (nonsymptotic) bounds on capacity for arbitrary power. In the high- and low-power regime the average-to-peak power ratio $\alpha = \frac{\mathcal{E}}{\mathcal{A}}$ is held fixed and thus both powers grow or vanish proportionally.

Using the symmetry of the channel model, it can be shown that when the average-power constraint exceeds half the peak-power constraint, then the average-power constraint is inactive. Therefore, the capacity $\mathcal{C}_{\text{FreeSpace}}(\mathcal{A}, \mathcal{E})$ with peak-power constraint \mathcal{A} and average-power constraint \mathcal{E} equals the capacity $\mathcal{C}_{\text{FreeSpace}}(\mathcal{A}, \mathcal{A})$ with only a peak-power constraint whenever the average-to-peak power ratio $\alpha \geq \frac{1}{2}$.

Hranilovic and Kschischang [14] studied the capacity of a continuous-time version of the free-space optical intensity channel and derived upper and lower bounds on this capacity under an average-power constraint and under the constraint of using a particular set of pulse shapes for the pulse intensity modulation scheme.

7.2.2.1 High-power results

For every $\alpha \in (0, \frac{1}{2})$ define μ^* as the unique solution to

$$\alpha = \frac{1}{\mu} - \frac{e^{-\mu}}{1 - e^{-\mu}}. \quad (7.28)$$

(It is easily shown that the solution to (7.28) always exists and is unique.)

THEOREM 7.7 ([22]) *If a peak-power constraint \mathcal{A} and an average-power constraint $\mathcal{E} = \alpha\mathcal{A}$ are imposed for some $\alpha \in (0, \frac{1}{2})$, then*

$$\mathcal{C}_{\text{FreeSpace}}(\mathcal{A}, \alpha\mathcal{A}) = \frac{1}{2} \log \left(\frac{\mathcal{A}^2}{2\pi e\sigma^2} \right) - (1 - \alpha)\mu^* - \log(1 - \alpha\mu^*) + o_{\mathcal{A}}(1), \quad (7.29)$$

where μ^* is the unique solution to (7.28).

If a peak-power constraint \mathcal{A} and an average-power constraint $\mathcal{E} = \alpha\mathcal{A}$ is imposed for some $\alpha \in [\frac{1}{2}, 1]$, or if only a peak-power constraint \mathcal{A} is imposed (which corresponds to $\alpha = 1$), then

$$\mathcal{C}_{\text{FreeSpace}}(\mathcal{A}, \alpha\mathcal{A}) = \frac{1}{2} \log \left(\frac{\mathcal{A}^2}{2\pi e\sigma^2} \right) + o_{\mathcal{A}}(1). \quad (7.30)$$

If only an average-power constraint \mathcal{E} is imposed, then

$$\mathcal{C}_{\text{FreeSpace}}(\mathcal{E}) = \frac{1}{2} \log \left(\frac{\mathcal{E}^2}{2\pi e\sigma^2} \right) + 1 + o_{\mathcal{E}}(1). \quad (7.31)$$

The asymptotic capacity (7.30) can be obtained directly from the well-known asymptotic behavior of the capacity of the classical AWGN channel under a peak-power constraint [29].

Comparing (7.29) and (7.30), we observe that the asymptotic high-power capacity is decreased by $(1 - \alpha)\mu^* + \log(1 - \alpha\mu^*)$, if additionally to a peak-power constraint \mathcal{A}

also an average-power constraint $\mathcal{E} = \alpha\mathcal{A} < \frac{\mathcal{A}}{2}$ is imposed. Similarly, comparing (7.30) and (7.31), we observe that the asymptotic high-power capacity is increased by 1 nat, if instead of a peak-power constraint an average-power constraint is imposed.

7.2.2.2 Low-power results

THEOREM 7.8 ([22]) *If both a peak-power constraint \mathcal{A} and an average-power constraint $\mathcal{E} = \alpha\mathcal{A}$ are imposed for some $\alpha \in (0, \frac{1}{2})$, then*

$$\lim_{\mathcal{A} \downarrow 0} \frac{C_{\text{FreeSpace}}(\mathcal{A}, \alpha\mathcal{A})}{\mathcal{A}^2} = \frac{\alpha(1-\alpha)}{2\sigma^2}. \quad (7.32)$$

If both a peak-power constraint \mathcal{A} and an average-power constraint $\mathcal{E} = \alpha\mathcal{A}$ are imposed for some $\alpha \in [\frac{1}{2}, 1]$, or if only a peak-power constraint \mathcal{A} is imposed (which corresponds to $\alpha = 1$), then

$$\lim_{\mathcal{A} \downarrow 0} \frac{C_{\text{FreeSpace}}(\mathcal{A}, \alpha\mathcal{A})}{\mathcal{A}^2} = \frac{1}{8\sigma^2}. \quad (7.33)$$

If only an average-power constraint \mathcal{E} is imposed, then

$$\lim_{\mathcal{E} \downarrow 0} \frac{C_{\text{FreeSpace}}(\mathcal{E})}{\frac{\mathcal{E}}{\sigma} \sqrt{\log\left(\frac{\sigma}{\mathcal{E}}\right)}} \geq \frac{1}{\sqrt{2}}, \quad (7.34)$$

$$\lim_{\mathcal{E} \downarrow 0} \frac{C_{\text{FreeSpace}}(\mathcal{E})}{\frac{\mathcal{E}}{\sigma} \sqrt{\log\left(\frac{\sigma}{\mathcal{E}}\right)}} \leq 2. \quad (7.35)$$

When only an average-power constraint \mathcal{E} is imposed, then the asymptotic upper and lower bound at low power do not coincide in the sense that their ratio is $2\sqrt{2}$ and not 1.

Expression (7.33) is easily obtained from the well-known asymptotic capacity of the classical AWGN channel with peak-power constraint [29].

Asymptotes (7.32), (7.33), and (7.34) can be achieved with binary signalling. In the presence of a peak-power constraint \mathcal{A} the “on”-signal is chosen as \mathcal{A} ; in its absence it is chosen as a constant times $\sqrt{\log\left(\frac{\sigma}{\mathcal{E}}\right)}$, thus tending to infinity as $\mathcal{E} \downarrow 0$.

Results (7.32), (7.34), and (7.35) show that in the presence of an average-power constraint the asymptotic low-power capacity depends heavily on the peak-power constraint. In fact, with a peak-power constraint that tends to 0 proportionally to \mathcal{E} , the low-power capacity grows quadratically in \mathcal{E} . In contrast, without a peak-power constraint the low-power capacity grows sublinearly, i.e., much faster for small values of \mathcal{E} .

Finally, comparing (7.32) and (7.33) we observe that if in addition to a peak-power constraint \mathcal{A} also an average-power constraint $\mathcal{E} = \alpha\mathcal{A} < \frac{\mathcal{A}}{2}$ is imposed, then the asymptotic low-power capacity is reduced by a factor $4\alpha(1-\alpha)$.

7.2.2.3 Nonasymptotic results

An efficient numerical method for computing the channel capacity under a peak-power and/or average-power constraint was proposed in [8], [7]. Here we present the analytic bounds of [22].

THEOREM 7.9 ([22]) *If both a peak-power constraint \mathcal{A} and an average-power constraint $\mathcal{E} = \alpha\mathcal{A}$ are imposed for some $\alpha \in (0, \frac{1}{2})$, then the capacity $\mathcal{C}_{\text{FreeSpace}}(\mathcal{A}, \alpha\mathcal{A})$ is lower-bounded by*

$$\mathcal{C}_{\text{FreeSpace}}(\mathcal{A}, \alpha\mathcal{A}) \geq \frac{1}{2} \log \left(1 + \mathcal{A}^2 \frac{e^{2\alpha\mu^*}}{2\pi e\sigma^2} \left(\frac{1 - e^{-\mu^*}}{\mu^*} \right)^2 \right), \quad (7.36)$$

and upper-bounded by each of the two bounds

$$\begin{aligned} \mathcal{C}_{\text{FreeSpace}}(\mathcal{A}, \alpha\mathcal{A}) &\leq \frac{1}{2} \log \left(1 + \alpha(1 - \alpha) \frac{\mathcal{A}^2}{\sigma^2} \right), \quad (7.37) \\ \mathcal{C}_{\text{FreeSpace}}(\mathcal{A}, \alpha\mathcal{A}) &\leq \left(1 - \mathcal{Q} \left(\frac{\delta + \alpha\mathcal{A}}{\sigma} \right) - \mathcal{Q} \left(\frac{\delta + (1 - \alpha)\mathcal{A}}{\sigma} \right) \right) \\ &\quad \cdot \log \left(\frac{\mathcal{A}}{\sigma} \cdot \frac{e^{\frac{\mu\delta}{\mathcal{A}}} - e^{-\mu(1 + \frac{\delta}{\mathcal{A}})}}{\sqrt{2\pi}\mu(1 - 2\mathcal{Q}(\frac{\delta}{\sigma}))} \right) \\ &\quad + \frac{\delta}{\sqrt{2\pi}\sigma} e^{-\frac{\delta^2}{2\sigma^2}} + \frac{\sigma}{\mathcal{A}} \frac{\mu}{\sqrt{2\pi}} \left(e^{-\frac{\delta^2}{2\sigma^2}} - e^{-\frac{(\mathcal{A} + \delta)^2}{2\sigma^2}} \right) \\ &\quad - \frac{1}{2} + \mathcal{Q} \left(\frac{\delta}{\sigma} \right) + \mu\alpha \left(1 - 2\mathcal{Q} \left(\frac{\delta + \frac{\mathcal{A}}{2}}{\sigma} \right) \right). \quad (7.38) \end{aligned}$$

Here, $\mu > 0$ and $\delta > 0$ are free parameters, μ^* is the unique solution to (7.28), and $\mathcal{Q}(\cdot)$ denotes the Gaussian \mathcal{Q} -function

$$\mathcal{Q}(\xi) \triangleq \int_{\xi}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{1}{2} \left(1 - \operatorname{erf} \left(\frac{\xi}{\sqrt{2}} \right) \right), \quad \forall \xi \in \mathbb{R}. \quad (7.39)$$

A suboptimal but useful choice for the free parameters δ and μ in (7.38) is given in [22, Eqs. (14) and (15)]. This choice suffices to prove that the asymptotic high-power capacity is upper-bounded by the right-hand side of (7.29).

Figure 7.2 depicts the bounds of Theorem 7.9 for $\alpha = 0.1$ where (7.38) is numerically minimized over $\delta, \mu > 0$. The maximum gap between upper and lower bound is numerically identified as ≈ 0.68 nats (for $\frac{\mathcal{A}}{\sigma} \approx 10.5$ dB).

THEOREM 7.10 ([22]) *If a peak-power constraint \mathcal{A} and an average-power constraint $\mathcal{E} = \alpha\mathcal{A}$ are imposed for some $\alpha \in [\frac{1}{2}, 1]$, then $\mathcal{C}_{\text{FreeSpace}}(\mathcal{A}, \alpha\mathcal{A})$ is lower-bounded by*

$$\mathcal{C}_{\text{FreeSpace}}(\mathcal{A}, \alpha\mathcal{A}) \geq \frac{1}{2} \log \left(1 + \frac{\mathcal{A}^2}{2\pi e\sigma^2} \right), \quad (7.40)$$

and upper-bounded by each of the two bounds

$$\mathcal{C}_{\text{FreeSpace}}(\mathcal{A}, \alpha\mathcal{A}) \leq \frac{1}{2} \log \left(1 + \frac{\mathcal{A}^2}{4\sigma^2} \right), \quad (7.41)$$

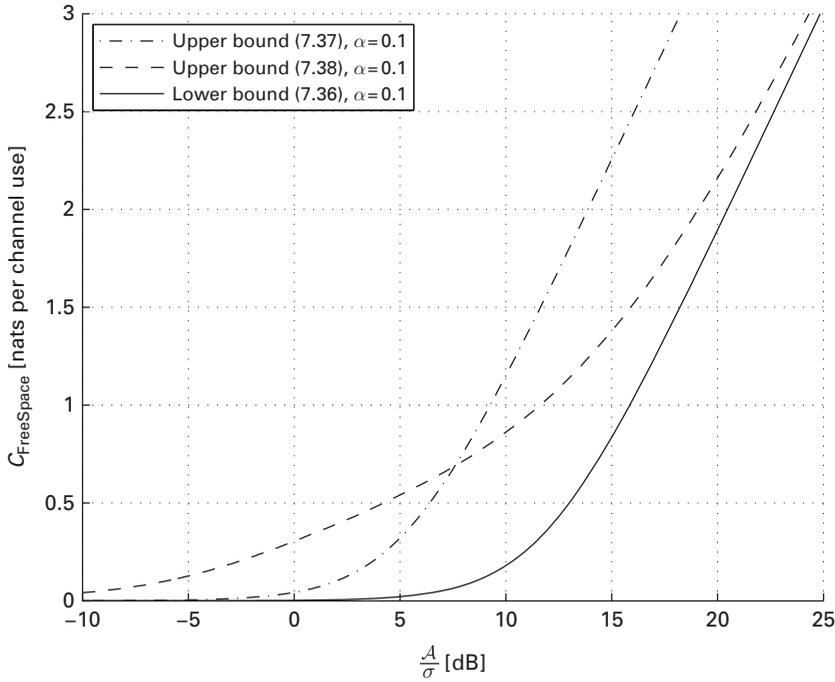


Figure 7.2 Bounds of Theorem 7.9 for $\alpha = 0.1$.

$$\begin{aligned}
 C_{\text{FreeSpace}}(\mathcal{A}, \alpha \mathcal{A}) \leq & \left(1 - 2Q\left(\frac{\delta + \frac{\mathcal{A}}{2}}{\sigma}\right) \right) \log\left(\frac{\mathcal{A} + 2\delta}{\sigma\sqrt{2\pi}(1 - 2Q(\frac{\delta}{\sigma}))}\right) \\
 & - \frac{1}{2} + Q\left(\frac{\delta}{\sigma}\right) + \frac{\delta}{\sqrt{2\pi}\sigma} e^{-\frac{\delta^2}{2\sigma^2}}, \quad (7.42)
 \end{aligned}$$

where $\delta > 0$ is a free parameter.

A suboptimal but useful choice for the free parameter $\delta > 0$ in (7.42) is given in [22, Eq. (21)]. This choice suffices to prove that the asymptotic high-power capacity is upper-bounded by the right-hand side of (7.30).

Figure 7.3 depicts the bounds of Theorem 7.10, where upper bound (7.42) is numerically minimized over $\delta > 0$. The maximum gap between upper and lower bound is numerically identified as ≈ 0.50 nats (for $\frac{\mathcal{A}}{\sigma} \approx 6.4$ dB).

THEOREM 7.11 ([22]) *If only an average-power constraint \mathcal{E} is imposed, then the capacity $C_{\text{FreeSpace}}(\mathcal{E})$ is lower-bounded by*

$$C_{\text{FreeSpace}}(\mathcal{E}) \geq \frac{1}{2} \log\left(1 + \frac{\mathcal{E}^2 e}{2\pi\sigma^2}\right), \quad (7.43)$$

and upper-bounded by each of the bounds

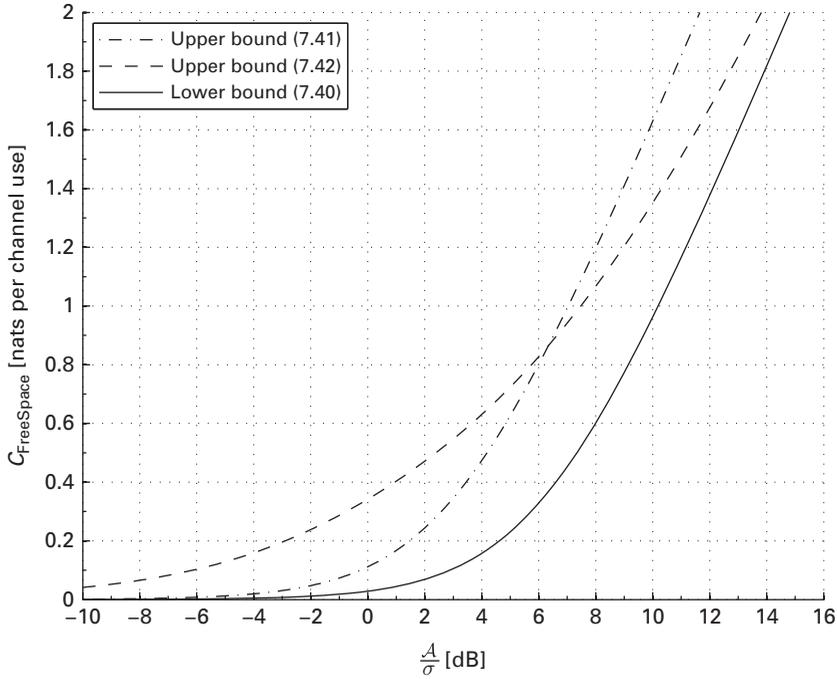


Figure 7.3 Bounds on capacity for $\alpha \in [\frac{1}{2}, 1]$ according to Theorem 7.10.

$$C_{\text{FreeSpace}}(\mathcal{E}) \leq \log \left(\beta e^{-\frac{\delta^2}{2\sigma^2}} + \sqrt{2\pi}\sigma \mathcal{Q} \left(\frac{\delta}{\sigma} \right) \right) - \log \left(\sqrt{2\pi}\sigma \right) - \frac{\delta\mathcal{E}}{2\sigma^2} + \frac{\delta^2}{2\sigma^2} \left(1 - \mathcal{Q} \left(\frac{\delta}{\sigma} \right) - \frac{\mathcal{E}}{\delta} \mathcal{Q} \left(\frac{\delta}{\sigma} \right) \right) + \frac{1}{\beta} \left(\mathcal{E} + \frac{\sigma}{\sqrt{2\pi}} \right), \quad (7.44)$$

$$C_{\text{FreeSpace}}(\mathcal{E}) \leq \log \left(\beta e^{-\frac{\delta^2}{2\sigma^2}} + \sqrt{2\pi}\sigma \mathcal{Q} \left(\frac{\delta}{\sigma} \right) \right) + \frac{1}{2} \mathcal{Q} \left(\frac{\delta}{\sigma} \right) + \frac{\delta}{2\sqrt{2\pi}\sigma} e^{-\frac{\delta^2}{2\sigma^2}} + \frac{\delta^2}{2\sigma^2} \left(1 - \mathcal{Q} \left(\frac{\delta + \mathcal{E}}{\sigma} \right) \right) + \frac{1}{\beta} \left(\delta + \mathcal{E} + \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{\delta^2}{2\sigma^2}} \right) - \frac{1}{2} \log \left(2\pi e\sigma^2 \right), \quad (7.45)$$

where $\beta > 0$ and δ are free parameters. Bound (7.44) only holds for $\delta \leq -\sigma e^{-\frac{1}{2}}$, while bound (7.45) only holds for $\delta \geq 0$.

A suboptimal but useful choice for the free parameters β and δ in (7.44) is given in [22, Eqs. (29) and (30)] and a suboptimal choice for the free parameters β and δ in (7.45) is given in [22, Eqs. (31) and (32)]. The first choice suffices to prove that the asymptotic low-power capacity is upper-bounded as in (7.35), and the second choice suffices to prove that the asymptotic high-power capacity is upper-bounded by the right-hand side of (7.31).

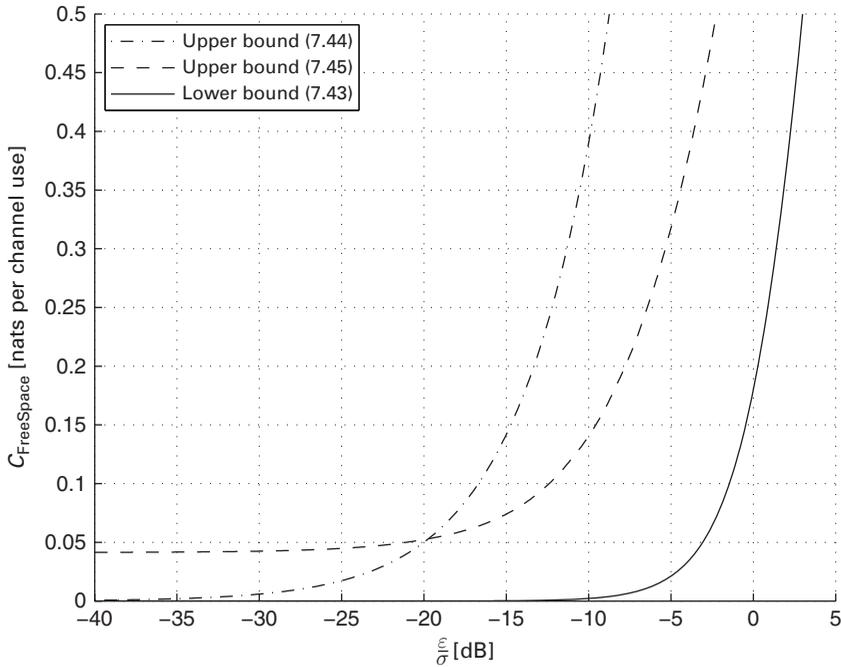


Figure 7.4 Bounds on capacity according to Theorem 7.11.

Figure 7.4 depicts the bounds of Theorem 7.11 when the upper bounds (7.44) and (7.45) are numerically minimized over the allowed values of β and δ . The maximum gap between upper and lower bound is numerically identified as ≈ 0.57 nats (for $\frac{\mathcal{E}}{\sigma^2} \approx 2.8$ dB).

7.2.3 Optical intensity channel with input-dependent Gaussian noise

For the optical intensity channel with input-dependent Gaussian noise we present results on the asymptotic capacities in the high-power and low-power regimes, as well as firm (nonasymptotic) lower bounds on the capacity for arbitrary power. We discuss the three scenarios with peak- and average-power constraints, with peak-power constraint only, and with average-power constraint only.

7.2.3.1 High-power results

THEOREM 7.12 ([25]) *If a peak-power constraint \mathcal{A} and an average-power constraint $\mathcal{E} = \alpha\mathcal{A}$ is imposed for some $\alpha \in (0, \frac{1}{3})$, then*

$$\mathcal{C}_{\text{DepNoise}}(\mathcal{A}, \alpha\mathcal{A}) = \frac{1}{2} \log \left(\frac{2\mathcal{A}}{\pi e} \right) - (1 - \alpha)v^* - \log(1 - 2\alpha v^*) + o_{\mathcal{A}}(1), \quad (7.46)$$

where v^* is the unique solution to (7.11).

If a peak-power constraint \mathcal{A} and an average-power constraint $\mathcal{E} = \alpha\mathcal{A}$ is imposed for some $\alpha \in [\frac{1}{3}, 1]$, or if only a peak-power constraint \mathcal{A} is imposed (which corresponds to $\alpha = 1$), then

$$\mathcal{C}_{\text{DepNoise}}(\mathcal{A}, \alpha\mathcal{A}) = \frac{1}{2} \log \left(\frac{2\mathcal{A}}{\pi e} \right) + o_{\mathcal{A}}(1). \quad (7.47)$$

If only an average-power constraint \mathcal{E} is imposed, then

$$\mathcal{C}_{\text{DepNoise}}(\mathcal{E}) = \frac{1}{2} \log \mathcal{E} + o_{\mathcal{E}}(1). \quad (7.48)$$

We notice that for $\alpha \in [\frac{1}{3}, 1]$ the asymptotic high-power capacity does not depend on α . Thus, in the high-power asymptote, the average-power constraint is inactive when $\alpha \in [\frac{1}{3}, 1]$.

Our asymptotic high-power capacity results (7.46)–(7.48) for the optical intensity channel with input-dependent Gaussian noise coincide with the asymptotic high-power capacity results (7.13)–(7.15) for the discrete-time Poisson channel. Intuitively, this correspondence can be understood by realizing that for large values of x , the cumulative distribution function of a Poisson random variable with mean x approximates the cumulative distribution function of a Gaussian random variable with mean x and variance x (see [25]).

7.2.3.2 Low-power results

THEOREM 7.13 ([25]) *If both a peak-power constraint \mathcal{A} and an average-power constraint $\mathcal{E} = \alpha\mathcal{A}$ are imposed for some $\alpha \in (0, \frac{1}{2})$, then*

$$\lim_{\mathcal{A} \downarrow 0} \frac{\mathcal{C}_{\text{DepNoise}}(\mathcal{A}, \alpha\mathcal{A})}{\mathcal{A}^2} = \alpha(1 - \alpha) \frac{1 + 2\sigma^2}{4\sigma^4}. \quad (7.49)$$

If both a peak-power constraint \mathcal{A} and an average-power constraint $\mathcal{E} = \alpha\mathcal{A}$ are imposed for some $\alpha \in [\frac{1}{2}, 1]$, or if only a peak-power constraint \mathcal{A} is imposed (which corresponds to $\alpha = 1$), then

$$\lim_{\mathcal{A} \downarrow 0} \frac{\mathcal{C}_{\text{DepNoise}}(\mathcal{A}, \alpha\mathcal{A})}{\mathcal{A}^2} = \frac{1 + 2\sigma^2}{16\sigma^4}. \quad (7.50)$$

We notice that the asymptotic low-power capacities (7.49) and (7.50) are very similar to the asymptotic capacities (7.32) and (7.33), especially for large values of σ^2 . Like for the free-space optical intensity channel, the threshold between the case with both a peak- and an average-power constraint and the case where the average-power constraint is inactive is at $\alpha = \frac{1}{2}$, and – contrary to the high-power regime – not at $\alpha = \frac{1}{3}$.

Thus, we conclude that whereas the capacity of the optical intensity channel with input-dependent Gaussian noise at high power behaves like the capacity of the discrete-time Poisson channel, at low power it behaves like the capacity of the free-space optical intensity channel.

7.2.3.3 Nonasymptotic results

THEOREM 7.14 ([25]) *If a peak-power constraint \mathcal{A} and an average-power constraint $\mathcal{E} = \alpha\mathcal{A}$ are imposed for some $\alpha \in (0, \frac{1}{3})$, then*

$$\begin{aligned} C_{\text{DepNoise}}(\mathcal{A}, \mathcal{E}) &\geq \frac{1}{2} \log \mathcal{A} - \frac{1}{2} \log(2\pi e) - (1 - \alpha)v^* - \log\left(\frac{1}{2} - \alpha v^*\right) \\ &\quad - e^{v^*} \left(\frac{1}{2} - \alpha v^*\right) \left(2\sqrt{\frac{\sigma^2}{\mathcal{A}}} \arctan\left(\sqrt{\frac{\mathcal{A}}{\sigma^2}}\right) + \log\left(1 + \frac{\sigma^2}{\mathcal{A}}\right)\right) \\ &\quad + \frac{1}{2} \log\left(1 + \frac{2}{\mathcal{E}}\right) + \sqrt{\mathcal{E}(2 + \mathcal{E})} - \mathcal{E} - 1, \end{aligned} \quad (7.51)$$

where v^* is defined to be the solution to (7.11).

If both a peak-power constraint \mathcal{A} and an average-power constraint $\mathcal{E} = \alpha\mathcal{A}$ are imposed with $\alpha \in [\frac{1}{3}, 1]$, or if only a peak-power constraint is imposed (which corresponds to $\alpha = 1$), then

$$\begin{aligned} C_{\text{DepNoise}}(\mathcal{A}, \mathcal{E}) &\geq \frac{1}{2} \log \mathcal{A} - \frac{1}{2} \log\left(\frac{\pi e}{2}\right) - \frac{1}{2} \log\left(1 + \frac{\sigma^2}{\mathcal{A}}\right) + \frac{1}{2} \log\left(1 + \frac{6}{\mathcal{A}}\right) \\ &\quad - \sqrt{\frac{\sigma^2}{\mathcal{A}}} \arctan\left(\sqrt{\frac{\mathcal{A}}{\sigma^2}}\right) + \sqrt{\frac{\mathcal{A}}{3}\left(2 + \frac{\mathcal{A}}{3}\right)} - \frac{\mathcal{A}}{3} - 1. \end{aligned} \quad (7.52)$$

If only an average-power constraint \mathcal{E} is imposed, then

$$C_{\text{DepNoise}}(\mathcal{E}) \geq \frac{1}{2} \log \mathcal{E} - \sqrt{\frac{\pi\sigma^2}{2\mathcal{E}}} + \frac{1}{2} \log\left(1 + \frac{2}{\mathcal{E}}\right) + \sqrt{\mathcal{E}(2 + \mathcal{E})} - \mathcal{E} - 1. \quad (7.53)$$

7.3 Proof techniques

We do not prove all the results presented in this chapter, but only outline some of the tools used in their derivation.

For scenarios where the capacity at low power grows linearly in the power, the exact asymptotic low-power capacity can be derived using a result from [31]. The result and its application are described in Section 7.3.1. When the inputs are constrained to a peak power \mathcal{A} that tends to 0 (irrespective of whether an average-power constraint is imposed) then the exact asymptotic low-power capacity for the two intensity channels can be derived using a result from [26]; see Section 7.3.2.

In all other cases, only lower and upper bounds on the capacity have been derived. However, as we have seen, in some asymptotic regimes these upper and lower bounds coincide thus yielding the exact asymptotes. In Section 7.3.3 we describe techniques to derive lower bounds. In Section 7.3.4 we describe the technique used to prove most upper bounds in this chapter. In addition to the technique in Section 7.3.4, some of the asymptotic upper bounds for the high-power regime require an additional result on capacity-achieving input distributions presented in Section 7.3.5.

7.3.1 Capacity per unit cost

For scenarios where the capacity at low power grows like a constant times the power, this constant can be found using Verdú's *capacity per unit cost* [31]. This result applies to scenarios without a peak-power constraint or with a peak-power constraint that does not tend to 0.

We first recall the result in [31], and then show how it can be used to compute the asymptotic low-power capacity (7.18) of the discrete-time Poisson channel with constant (nonzero) dark current under a vanishing average-power constraint.

Let a memoryless stationary channel with input alphabet \mathcal{X} , output alphabet \mathcal{Y} , channel law $W(\cdot|\cdot)$, and a nonnegative cost function $g: \mathcal{X} \rightarrow \mathbb{R}_0^+$ be given. For every real number $\Upsilon \geq 0$ the capacity-cost function $\mathcal{C}(\Upsilon)$ is defined as

$$\mathcal{C}(\Upsilon) \triangleq \sup_{Q \in \mathcal{P}(\mathcal{X})} I(Q, W) \quad (7.54)$$

where the supremum is over all input distributions Q that satisfy $E_Q[g(X)] \leq \Upsilon$. The *capacity per unit cost* is defined as $\sup_{\Upsilon \geq 0} \frac{\mathcal{C}(\Upsilon)}{\Upsilon}$.

THEOREM 7.15 (Capacity per Unit Cost, [31]) *Consider a memoryless stationary channel with input alphabet \mathcal{X} , output alphabet \mathcal{Y} , channel law $W(\cdot|\cdot)$, and a nonnegative cost function $g: \mathcal{X} \rightarrow \mathbb{R}_0^+$ such that for some $x_0 \in \mathcal{X}$ the cost is 0, i.e., $g(x_0) = 0$. Then, the capacity per unit cost is given by*

$$\sup_{\Upsilon \geq 0} \frac{\mathcal{C}(\Upsilon)}{\Upsilon} = \lim_{\Upsilon \downarrow 0} \frac{\mathcal{C}(\Upsilon)}{\Upsilon} \stackrel{(a)}{=} \sup_{x \in \mathcal{X} \setminus \{x_0\}} \frac{D(W(\cdot|x) \| W(\cdot|0))}{g(x)}, \quad (7.55)$$

where $D(\cdot|\cdot)$ denotes the relative entropy [4].⁵

For our purpose of computing the asymptotic low-power capacity of a channel, Equation (a) in (7.55) is useful. In fact, for channels with an input alphabet $\mathcal{X} \subseteq \mathbb{R}_0^+$, specializing Equation (7.55a) to the choice $g: x \mapsto x$ and $\Upsilon = \mathcal{E}$, results in

$$\lim_{\mathcal{E} \downarrow 0} \frac{\mathcal{C}(\mathcal{E})}{\mathcal{E}} = \sup_{x \in \mathcal{X} \setminus \{0\}} \frac{D(W(\cdot|x) \| W(\cdot|0))}{x}. \quad (7.56)$$

Equation (7.56) can be used to prove the asymptotic low-power capacity result (7.18) on the discrete-time Poisson channel with constant (nonzero) dark current under a vanishing average-power constraint. We briefly sketch the proof of (7.18). For every positive input x :

$$D(W(\cdot|x) \| W(\cdot|0)) = \sum_y e^{-\lambda+x} \frac{(x+\lambda)^y}{y!} \log \left(e^{-x} \left(1 + \frac{x}{\lambda} \right)^y \right) \quad (7.57)$$

$$= -x + (\lambda + x) \log \left(1 + \frac{x}{\lambda} \right). \quad (7.58)$$

⁵ For finite alphabets the expression of the relative entropy is given in (7.80) ahead.

Thus, by (7.55a) where we choose $g: x \mapsto x$ and $\Upsilon = \mathcal{E}$, we obtain for $\lambda > 0$:

$$\lim_{\mathcal{E} \downarrow 0} \frac{\mathcal{C}_{\text{Poisson}}(\lambda, \mathcal{A}, \mathcal{E})}{\mathcal{E}} = \sup_{x \in (0, \mathcal{A}] } \frac{D(W(\cdot|x) \| W(\cdot|0))}{x} \quad (7.59)$$

$$= \sup_{x \in (0, \mathcal{A}] } \left(1 + \frac{\lambda}{x} \right) \log \left(1 + \frac{x}{\lambda} \right) - 1 \quad (7.60)$$

$$= \left(\frac{\lambda}{\mathcal{A}} \right) \log \left(1 + \frac{\mathcal{A}}{\lambda} \right) - 1. \quad (7.61)$$

While Theorem 7.15 suffices to prove (7.18), it is too weak to derive the asymptotic low-power capacity (7.21) of the discrete-time Poisson channel when the dark current either tends to 0 or equals 0. For example, for zero dark current $\lambda = 0$, (7.58) implies that the relative entropy $D(W(\cdot|x) \| W(\cdot|0)) = \infty$ irrespective of the value of x . Consequently, when $\lambda = 0$ the capacity per unit cost is infinite:

$$\lim_{\mathcal{E} \downarrow 0} \frac{\mathcal{C}_{\text{Poisson}}(\lambda = 0, \mathcal{A}, \mathcal{E})}{\mathcal{E}} = \infty, \quad (7.62)$$

which only proves that at low power the capacity grows faster than linearly in \mathcal{E} , but it does not specify that it grows like a constant times $\mathcal{E} \log(\frac{1}{\mathcal{E}})$ as exhibited by (7.21).

7.3.2 Mutual information of weak input signals

For scenarios where the peak-power constraint tends to 0, a result by Prelov and van der Meulen [26] can be used to obtain the exact asymptotic low-power capacity. The following theorem is included as a special case in [26, Theorem 2].

THEOREM 7.16 ([26]) *Consider a channel that for all sufficiently small inputs x produces an output that is Gaussian distributed with mean m_x and variance σ_x^2 that can depend on x . Then, for sufficiently small \mathcal{A} and $|X| \leq \mathcal{A}$, the mutual information between the channel's input X and output Y satisfies*

$$I(X; Y) = \frac{1}{2} J(0) \text{Var}(X) + o(\mathcal{A}^2), \quad (7.63)$$

where $o(\mathcal{A}^2)$ denotes a term that tends to 0 faster than $\mathcal{A}^2 \downarrow 0$, and where $J(0)$ denotes the Fisher information of the channel at 0:

$$J(x) \triangleq \int_{\mathcal{Y}} \frac{\left(\frac{d}{dx} W(y|x) \right)^2}{W(y|x)} dy. \quad (7.64)$$

It is easily seen that the free-space optical intensity channel and the optical intensity channel with input-dependent Gaussian noise satisfy the assumption in the theorem. Thus, the theorem can be used to derive the asymptotic low-power capacity of the free-space optical intensity channel (7.32) and (7.33) under both a peak-power and an average-power constraint and under a peak-power constraint only [22]. In the same way, also the asymptotic low-power capacity of the optical intensity channel with

input-dependent Gaussian noise under both peak- and average-power constraints (7.49) and under a peak-power constraint only (7.50) can be derived from Theorem 7.16.

We briefly sketch the derivation of (7.32). It is easily proved that for the free-space optical intensity channel:

$$J(0) = \frac{1}{\sigma^2}. \quad (7.65)$$

Moreover, for $\mathcal{E} \leq \frac{\mathcal{A}}{2}$:

$$\max_{\substack{Q \in \mathcal{P}([0, \mathcal{A}]) \\ E_Q[X] \leq \mathcal{E} = \alpha \mathcal{A}}} \text{Var}(X) = \mathcal{E}(\mathcal{A} - \mathcal{E}) = \alpha(1 - \alpha)\mathcal{A}^2. \quad (7.66)$$

Combining (7.65), (7.66) with (7.63) and the definition of channel capacity (7.9), we conclude that for $\alpha \in (0, \frac{1}{2}]$:

$$\mathcal{C}_{\text{FreeSpace}}(\mathcal{A}, \alpha \mathcal{A}) = \frac{\alpha(1 - \alpha)\mathcal{A}^2}{2\sigma^2} + o(\mathcal{A}^2), \quad (7.67)$$

which establishes (7.32).

Expression (7.33) is obtained by specializing (7.67) to $\alpha = \frac{1}{2}$.

7.3.3 Lower bounds on capacity

A lower bound on channel capacity can be obtained by dropping the maximization in (7.9) and considering any input distribution Q satisfying the constraints. For the resulting bound to be tight, this choice of Q should yield a mutual information that is reasonably close to capacity. Finding such a Q can be tricky and might result in $I(Q, W)$ being intractable. In the following, we present two methods for choosing the distribution Q and for lower-bounding $I(Q, W)$.

The first method is based on choosing an entropy-maximizing input distribution (Theorem 7.18) and on relating the entropy of the channel's output to the entropy of its input. For additive noise channels such as the free-space optical intensity channel this can be done using the entropy power inequality (Theorem 7.17); for the optical intensity channel with input-dependent Gaussian noise and the discrete-time Poisson channel this requires a different approach; see [25] and [21].

The second method is based on choosing binary inputs and deriving only an asymptotic lower bound on $I(Q, W)$ for low power.

7.3.3.1 Method 1: Entropy power inequality and entropy-maximizing distributions

In the following we assume that we are given a channel

$$Y = x + Z \quad (7.68)$$

where Y denotes the output of the channel, x the input, and Z the additive noise which has a density. We prove a lower bound on the capacity of such channels based on the entropy power inequality in Theorem 7.17 and the max-entropy result in Theorem 7.18.

THEOREM 7.17 (Entropy power inequality, [4]) *If X and Z are independent random variables with densities, then*

$$e^{2h(X+Z)} \geq e^{2h(X)} + e^{2h(Z)}. \quad (7.69)$$

THEOREM 7.18 (Max-entropy theorem, [4]) *Let an alphabet $\mathcal{X} \subseteq \mathbb{R}$, a nonnegative integer m , a set of m functions $\{r_i: \mathcal{X} \rightarrow \mathbb{R}\}_{i=1}^m$, and m values $\{a_i \in \mathbb{R}\}_{i=1}^m$ be given, and consider the optimization problem*

$$\sup_{Q \in \mathcal{P}(\mathcal{X})} h(X) \quad (7.70)$$

subject to

$$E_Q[r_i(X)] = a_i, \quad i \in \{1, \dots, m\}. \quad (7.71)$$

If there exist real numbers $\kappa_0, \dots, \kappa_m$ such that the distribution \tilde{Q} with density

$$\tilde{Q}'(x) \triangleq e^{\kappa_0 + \sum_{i=1}^m \kappa_i r_i(x)}, \quad x \in \mathcal{X}, \quad (7.72)$$

satisfies constraints (7.71) when Q is replaced by \tilde{Q} , then \tilde{Q} is the unique distribution on \mathcal{X} subject to (7.71) that achieves the supremum in (7.70).

Using the entropy power inequality, a lower bound on the capacity of additive noise channels as in (7.68) is obtained by noting that for any input X that has a density we have

$$\mathcal{C} \geq I(X; Y) \quad (7.73)$$

$$= h(Y) - h(Y|X) \quad (7.74)$$

$$= h(X + Z) - h(Z) \quad (7.75)$$

$$\geq \frac{1}{2} \log \left(e^{2h(X)} + e^{2h(Z)} \right) - h(Z). \quad (7.76)$$

Obviously, the lower bound is made tightest if the density of X is chosen so as to maximize the differential entropy under the given constraints. For many interesting scenarios the desired density can be found using the max-entropy result in Theorem 7.18.

This method was used to derive the lower bounds on the capacity of the free-space optical intensity channel both in the asymptotic high-power regime as well as in the nonasymptotic regime.

We quickly sketch how this method is used to prove lower bound (7.36). By Theorem 7.18 the density that maximizes differential entropy under a nonnegativity constraint, a peak constraint \mathcal{A} , and an average constraint $\mathcal{E} = \alpha \mathcal{A} \leq \frac{1}{2} \mathcal{A}$ is given by⁶

$$Q'(x) = \frac{1}{\mathcal{A}} \cdot \frac{\mu^*}{1 - e^{-\mu^*}} e^{-\frac{\mu^* x}{\mathcal{A}}}, \quad 0 \leq x \leq \mathcal{A}, \quad (7.77)$$

⁶ Using the symmetry of the channel and the uniqueness of the capacity-achieving input distribution [3] it is easily proved that for the free-space optical intensity channel and $\mathcal{E} \leq \frac{1}{2} \mathcal{A}$ the average-power constraint is always active.

where μ^* is the unique solution to (7.28). We also note that since the noise of the free-space optical intensity channel is Gaussian with variance σ^2 :

$$\frac{1}{2} \log \left(e^{2h(X)} + e^{2h(Z)} \right) - h(Z) = \frac{1}{2} \log \left(1 + \frac{e^{2h(X)}}{2\pi e\sigma^2} \right). \quad (7.78)$$

Evaluating the right-hand side of (7.78) for an input X that has density (7.77) and combining the result with (7.76) yields the desired lower bound (7.36).

7.3.3.2 Method 2: Binary inputs and asymptotic lower bound

It is well known that the capacity-achieving input distribution for our channels is discrete and that for sufficiently low power the number of mass points is two [27], [3]. Thus, good lower bounds for the channel capacity at low power can be obtained by evaluating the mutual information $I(X; Y)$ for binary inputs. Unfortunately, for most such binary inputs an analytic expression for the mutual information $I(X; Y)$ seems infeasible even in the asymptotic regime of low power. We therefore propose to lower-bound the mutual information and only derive the low-power asymptote of the lower bound.

This method was used to prove: the lower bound on the asymptotic low-power capacity of the discrete-time Poisson channel with constant dark current (7.19); the achievability of the asymptotic low-power capacity of the discrete-time Poisson channel with vanishing dark current (7.21); and the lower bound on the asymptotic low-power capacity of the free-space optical intensity channel (7.34). The exact derivations are tedious; for details see [23] and [22].

7.3.4 Duality-based upper bounds on capacity

We present an approach from [19], [20] for deriving upper bounds on the channel capacity. The approach is explained here for discrete memoryless channels $W(\cdot|\cdot)$ with finite input and output alphabets \mathcal{X} and \mathcal{Y} . The generalization to continuous alphabets with input constraints is relatively straightforward.⁷

In [5] a dual expression for the channel capacity of discrete memoryless channels in (7.9) was presented:

$$\mathcal{C} = \min_{R \in \mathcal{P}(\mathcal{Y})} \max_{x \in \mathcal{X}} D(W(\cdot|x) \| R(\cdot)), \quad (7.79)$$

where $D(\cdot\|\cdot)$ denotes relative entropy [4]

$$D(W(\cdot|x) \| R(\cdot)) = \sum_{y \in \mathcal{Y}} W(y|x) \log \left(\frac{W(y|x)}{R(y)} \right). \quad (7.80)$$

This dual expression allows us to easily obtain upper bounds on the channel capacity. In fact, every choice of a distribution $R(\cdot)$ on the output \mathcal{Y} leads to such an upper bound:

$$\mathcal{C} \leq \max_{x \in \mathcal{X}} D(W(\cdot|x) \| R(\cdot)). \quad (7.81)$$

⁷ There are some mathematical technicalities that must be addressed, but they do not change the fundamental idea.

However, for every choice of the distribution $R(\cdot)$ a generally tighter upper bound on capacity can be derived. In fact, by considering the identity [30]

$$\sum_{x \in \mathcal{X}} Q(x) D(W(\cdot|x) \| R(\cdot)) = I(Q, W) + D((QW)(\cdot) \| R(\cdot)) \quad (7.82)$$

and by noting that relative entropy is nonnegative [4], we have

$$I(Q, W) \leq \sum_{x \in \mathcal{X}} Q(x) D(W(\cdot|x) \| R(\cdot)), \quad R \in \mathcal{P}(\mathcal{Y}). \quad (7.83)$$

This leads to the following upper bound on capacity [19], [20]. For every choice of a distribution $R(\cdot)$ on the channel output alphabet

$$\mathcal{C} \leq E_{Q^*} [D(W(\cdot|X) \| R(\cdot))], \quad (7.84)$$

where Q^* denotes a capacity-achieving input distribution, i.e., a distribution that achieves the maximum in (7.9). It is important to note that $R(\cdot)$ in (7.84) can be any distribution on the output alphabet: it need not be an output distribution that is induced by some input distribution.

Two problems make the evaluation of the right-hand side of (7.84) difficult. Firstly, evaluating $D(W(\cdot|x) \| R(\cdot))$ analytically is generally difficult; and secondly, the right-hand side of (7.84) involves an expectation over the capacity-achieving input distribution Q^* which in general is not known.⁸ So in order for the bound to be useful, a good but relatively simple distribution $R(\cdot)$ has to be found for which the expectation in (7.84) over Q^* can be computed or reasonably upper-bounded (without knowing Q^*). Often the expectation with respect to Q^* can be computed/bounded by using some of the known properties of Q^* like, e.g., that it satisfies the input constraints or that it escapes to infinity (see Section 7.3.5), or by using Jensen's inequality [4] which states that for every concave function $f(\cdot)$:

$$E[f(X)] \leq f(E(X)). \quad (7.85)$$

Most of the upper bounds presented in this chapter have been proved using this technique. Typically, in these proofs the output distribution $R(\cdot)$ was chosen depending on some parameters that can be optimized at a later stage. Thus, a *family* of output distributions was chosen and a particular choice was singled out only after the evaluation of (7.84). This way one can even make the choice of $R(\cdot)$ depend on the channel parameters such as the allowed peak and average powers.⁹

7.3.5 Input distributions that escape to infinity

As we have seen, the capacity with peak power \mathcal{A} and average power $\mathcal{E} = \alpha\mathcal{A}$ of many channels of interest can be expressed at high input powers as

$$\mathcal{C}(\mathcal{A}, \alpha\mathcal{A}) = \Pi \log \mathcal{A} + \chi + o_{\mathcal{A}}(1), \quad (7.86)$$

⁸ If it were known, then capacity could be (at least numerically) computed directly, and no bounding would be necessary.

⁹ Note, however, that by definition $R(\cdot)$ must not depend on the channel input x .

where the correction term $o_{\mathcal{A}}(1)$ tends to zero as \mathcal{A} tends to infinity. The constant Π is sometimes called the *pre-log* and is given by

$$\Pi = \lim_{\mathcal{A} \uparrow \infty} \frac{\mathcal{C}(\mathcal{A}, \alpha \mathcal{A})}{\log \mathcal{A}}. \quad (7.87)$$

For example, the pre-log in (7.29) is¹⁰ 1 and the pre-logs in (7.13) and (7.46) are $\frac{1}{2}$. To compute Π and χ the notion of *input distributions that escape to infinity* can be useful.

For many channels of interest the asymptotic capacity or even the pre-log can only be achieved if the input distributions *escape to infinity*. Loosely speaking this means that the probabilities that they assign to any fixed ball tend to zero as the allowed power tends to infinity. We next define this concept more formally and explain how it can be used to compute Π and χ . (For an example, see [18].)

The definition and the statements in this section are valid for general channels; they are not restricted to optical channels. We will only assume that the input and output alphabets \mathcal{X} and \mathcal{Y} of some channel are separable metric spaces and that for any set $\mathcal{B} \subset \mathcal{Y}$ the mapping $x \mapsto W(\mathcal{B}|x)$ from \mathcal{X} to $[0, 1]$ is Borel measurable. We then consider a general cost function $g: \mathcal{X} \rightarrow [0, \infty)$ which is assumed measurable.¹¹ In this setting, the capacity-cost function with an average- and a peak-cost constraint is defined as

DEFINITION 7.1 Given a channel $W(\cdot|\cdot)$ over the input alphabet \mathcal{X} and the output alphabet \mathcal{Y} and given some nonnegative cost function $g: \mathcal{X} \rightarrow \mathbb{R}_0^+$, we define the capacity-cost function $\mathcal{C}: ([\inf_{x \in \mathcal{X}} g(x), \infty))^2 \rightarrow \mathbb{R}_0^+$ by

$$\mathcal{C}(\mathcal{A}, \mathcal{E}) \triangleq \sup_{Q(\cdot)} I(Q, W), \quad \mathcal{A}, \mathcal{E} \geq \inf_{x \in \mathcal{X}} g(x) \quad (7.88)$$

where the supremum is over all input distributions $Q(\cdot)$ that satisfy

$$Q(\{x \in \mathcal{X}: g(x) > \mathcal{A}\}) = 0 \quad (7.89)$$

and

$$E_Q[g(X)] \leq \mathcal{E}. \quad (7.90)$$

The following results also hold in the absence of a peak-power constraint. For brevity we omit the explicit statements for this case.

DEFINITION 7.2 Fixing the ratio of available average-to-peak cost $\alpha \triangleq \frac{\mathcal{E}}{\mathcal{A}}$, we say that a family of input distributions $\{Q_{\mathcal{A}, \mathcal{E}}(\cdot)\}_{\mathcal{A} \geq \inf_x \frac{g(x)}{\alpha}, \mathcal{E} = \alpha \mathcal{A}}$ on \mathcal{X} parameterized by \mathcal{A} and \mathcal{E} *escapes to infinity* if for any fixed $\mathcal{A}_0 > 0$

$$\lim_{\mathcal{A} \uparrow \infty} Q_{\mathcal{A}, \alpha \mathcal{A}}(\{x \in \mathcal{X}: g(x) \leq \mathcal{A}_0\}) = 0. \quad (7.91)$$

¹⁰ Note the \mathcal{A}^2 in the logarithm that will multiply the factor $\frac{1}{2}$ in front of the logarithm by 2.

¹¹ For an intuitive understanding of the following definitions and some of their consequences, it is best to focus on the example of the channel models shown in Section 7.1 where the channel inputs are nonnegative real numbers and where the cost function $g(\cdot)$ is $g(x) = x, \forall x \geq 0$.

Based on this definition, a general theorem was presented in [19] and [20] demonstrating that if the ratio of mutual information to channel capacity is to approach 1, then the input distribution must escape to infinity:

THEOREM 7.19 (Escaping to infinity, [20]) *Let the capacity-cost function $\mathcal{C}(\cdot, \cdot)$ of a channel $W(\cdot|\cdot)$ be finite but unbounded. Let $\mathcal{C}_{\text{asy}}(\cdot)$ be a function that captures the asymptotic behavior of the capacity-cost function $\mathcal{C}(\mathcal{A}, \alpha\mathcal{A})$ in the sense that*

$$\lim_{\mathcal{A} \uparrow \infty} \frac{\mathcal{C}(\mathcal{A}, \alpha\mathcal{A})}{\mathcal{C}_{\text{asy}}(\mathcal{A})} = 1. \quad (7.92)$$

Assume that $\mathcal{C}_{\text{asy}}(\cdot)$ satisfies the growth condition

$$\lim_{\mathcal{A} \uparrow \infty} \left\{ \sup_{\mu \in (0, \mu_0]} \frac{\mu \mathcal{C}_{\text{asy}}\left(\frac{\mathcal{A}}{\mu}\right)}{\mathcal{C}_{\text{asy}}(\mathcal{A})} \right\} < 1, \quad \forall 0 < \mu_0 < 1. \quad (7.93)$$

Let $\{Q_{\mathcal{A}, \alpha\mathcal{A}}(\cdot)\}_{\mathcal{A} \geq \inf_x \frac{g(x)}{\alpha}}$ be a family of input distributions satisfying the cost constraints (7.89) and (7.90) such that

$$\lim_{\mathcal{A} \uparrow \infty} \frac{I(Q_{\mathcal{A}, \alpha\mathcal{A}}, W)}{\mathcal{C}_{\text{asy}}(\mathcal{A})} = 1. \quad (7.94)$$

Then $\{Q_{\mathcal{A}, \alpha\mathcal{A}}(\cdot)\}_{\mathcal{A} \geq \inf_x \frac{g(x)}{\alpha}}$ escapes to infinity.

Proof A first version of this theorem was proved in [19]. In [20, Sec. VII.C.3] the theorem was further generalized to the form presented here and the proof was simplified. \square

Examples of functions $\mathcal{C}_{\text{asy}}(\mathcal{A})$ that satisfy (7.93) include

$$\log(1 + \log(1 + \mathcal{A})), \quad \log(1 + \mathcal{A}), \quad (\log(1 + \mathcal{A}))^\beta \text{ for } \beta > 0 \quad (7.95)$$

and any positive multiple thereof. Hence, most channels of interest will satisfy Theorem 7.19 in the sense that for these channels there exists a function $\mathcal{C}_{\text{asy}}(\mathcal{A})$ that satisfies (7.92) and (7.93).

Returning to the problem of computing Π and χ of (7.86), we note that if a sequence of input distributions achieves Π (in the sense that the limiting ratio of the mutual information they induce to $\log \mathcal{A}$ tends to Π), then they must escape to infinity (see Theorem 7.19 with $\mathcal{C}_{\text{asy}}(\mathcal{A})$ taken as $\Pi \log \mathcal{A}$). Once we have computed Π – or even only established the existence of the limit in (7.87) – we can now proceed to compute χ (and Π if needed) by limiting ourselves to input distributions that escape to infinity. Limiting ourselves to such inputs may allow us to tighten the bounds obtained using the duality approach. For example, a straightforward consequence of Definition 7.2 is the following result about expectations with respect to input distributions that escape to infinity.

PROPOSITION 7.1 *Let $f: \mathcal{X} \rightarrow \mathbb{R}$ be a bounded function that maps input symbols of large cost to small values in the sense that for every $\epsilon > 0$ there exists some \mathcal{A}_ϵ such*

that $|f(x)| < \epsilon$ whenever $g(x) \geq A_\epsilon$. If the input distributions $\{Q_{A,\epsilon}(\cdot)\}_{A \geq \inf_x \frac{g(x)}{\alpha}, \epsilon = \alpha A}$ on \mathcal{X} escape to infinity, then

$$\lim_{A \uparrow \infty} E_{Q_{A,\alpha A}}[f(X)] = 0. \quad (7.96)$$

References

- [1] D. Brady and S. Verdú, "The asymptotic capacity of the direct detection photon channel with a bandwidth constraint," in *Proc. 28th Allerton Conf. Commun., Contr. and Comput.*, Monticello, IL, USA, Oct. 3–5, 1990, pp. 691–700.
- [2] K. Chakraborty and P. Narayan, "The Poisson fading channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2349–2364, Jul. 2007.
- [3] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073–2088, Jun. 2005.
- [4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd edn. John Wiley & Sons, 2006.
- [5] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.
- [6] M. H. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 710–715, Nov. 1980.
- [7] A. A. Farid and S. Hranilovic, "Capacity of optical intensity channels with peak and average power constraints," in *Proc. IEEE International Conference on Communications*, Dresden, Germany, Jun. 2009.
- [8] A. A. Farid and S. Hranilovic, "Channel capacity and non-uniform signalling for free-space optical intensity channels," *IEEE J. Select. Areas Commun.*, vol. 27, no. 9, pp. 1553–1563, Dec. 2009.
- [9] M. R. Frey, "Information capacity of the Poisson channel," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 244–256, Mar. 1991.
- [10] M. R. Frey, "Capacity of the L_p norm-constrained Poisson channel," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 445–450, Mar. 1992.
- [11] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.
- [12] S. M. Haas and J. Shapiro, "Capacity of wireless optical communications," *IEEE Trans. Inf. Theory*, vol. 21, no. 8, pp. 1346–1357, Oct. 2003.
- [13] K.-P. Ho, *Phase-Modulated Optical Communication Systems*. Springer Verlag, 2005.
- [14] S. Hranilovic and F. R. Kschischang, "Capacity bounds for power- and band-limited optical intensity channels corrupted by Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 784–795, May 2004.
- [15] Y. Kabanov, "The capacity of a channel of the Poisson type," *Theory of Probability and Its Appl.*, vol. 23, pp. 143–147, 1978.
- [16] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proc. IEEE*, vol. 85, no. 2, pp. 265–298, Feb. 1997.
- [17] M. Katz and S. Shamai, "On the capacity-achieving distribution of the discrete-time non-coherent and partially coherent AWGN channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2257–2270, Oct. 2004.

-
- [18] A. Lapidoth, "On phase noise channels at high SNR," in *Proc. IEEE Inf. Theory Workshop*, Bangalore, India, Oct. 20–25, 2002, pp. 1–4.
- [19] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
- [20] A. Lapidoth and S. M. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 437–453, Feb. 2006.
- [21] A. Lapidoth and S. M. Moser, "On the capacity of the discrete-time Poisson channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 303–322, Jan. 2009.
- [22] A. Lapidoth, S. M. Moser, and M. A. Wigger, "On the capacity of free-space optical intensity channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4449–4461, Oct. 2009.
- [23] A. Lapidoth, J. H. Shapiro, V. Venkatesan, and L. Wang, "The Poisson channel at low input powers," in *Proc. 25th IEEE Conv. El. & Electr. Eng. in Israel*, Eilat, Israel, Dec. 3–5, 2008.
- [24] A. Martinez, "Spectral efficiency of optical direct detection," *J. Opt. Soc. Am. B*, vol. 24, no. 4, pp. 739–749, Apr. 2007.
- [25] S. M. Moser, "Capacity results of an optical intensity channel with input-dependent Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, Jan. 2012.
- [26] V. V. Prelov and E. C. van der Meulen, "An asymptotic expression for the information and capacity of a multidimensional channel with weak input signals," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1728–1735, Sept. 1993.
- [27] S. Shamai (Shitz), "Capacity of a pulse amplitude modulated direct detection photon channel," in *Proc. IEE*, vol. 137, pt. I (Communications, Speech and Vision), no. 6, Dec. 1990, pp. 424–430.
- [28] S. Shamai (Shitz) and A. Lapidoth, "Bounds on the capacity of a spectrally constrained Poisson channel," *IEEE Trans. Inf. Theory*, vol. 39, no. 1, pp. 19–29, Jan. 1993.
- [29] C. E. Shannon, "A mathematical theory of communication," *Bell System Techn. J.*, vol. 27, pp. 379–423 and 623–656, Jul. and Oct. 1948.
- [30] F. Topsøe, "An information theoretical identity and a problem involving capacity," *Studia Sci. Math. Hungar.*, vol. 2, pp. 291–292, 1967.
- [31] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sept. 1990.
- [32] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel – part I and II," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1462–1471, Nov. 1988.

Part III

Unique channels

8 Modeling and characterization of ultraviolet scattering communication channels

Haipeng Ding, Brian M. Sadler, Gang Chen, and Zhengyuan Xu

In this chapter, we examine basic modeling and characterization issues that arise from short-range communications over non-line-of-sight (NLOS) ultraviolet (UV) atmospheric scattering channels. We start by presenting the unique channel properties and history of NLOS UV communications, and introduce outdoor NLOS UV scattering link geometries. Both single and multiple scattering effects are considered, including channel impulse response and link path loss. Analysis and Monte Carlo simulation are employed to investigate the UV channel properties. We also report on experimental outdoor channel measurements, and compare with theoretical predictions.

8.1 Introduction

8.1.1 Why NLOS UV communications?

Within the ultraviolet (UV) spectrum 4~400 nm, the UV-C band (sometimes called deep UV, or mid-UV) with wavelength range 200~280 nm has unique atmospheric propagation characteristics that enable non-line-of-sight (NLOS) communications, as described in [33]. This band is essentially solar blind (e.g., see [17]), so that a ground-based photodetector can approach quantum-limited photon counting detection performance, and low-power sources can be used. At these wavelengths, scattering due to photon interaction with atmospheric molecules and aerosols is very pronounced, and the relatively angle-independent scattering yields spatially diverse NLOS communications paths, e.g., see [21]. This rich scattering diversity enables unique NLOS optical communications, although the great majority of photons may be scattered away from the receiver and so the corresponding path loss is typically much higher than that occurring with a conventional optical LOS communication channel. The low solar background helps significantly in this regard, so that short-range links can be maintained with relatively low transmit power. In addition, significant atmospheric absorption acts to limit communication range, while also enhancing the potential for short-range networking by reducing multi-user interference.

For eye and skin safety, the International Commission on Non-Ionizing Radiation Protection [1] and the International Electrotechnical Commission [2] govern UV exposure power limits. In the deep UV region, the minimum allowable continuous exposure occurs at 270 nm at a level as low as 3 mJ per cm² per second, while increasing to 100 mJ at 200 nm and 3.4 mJ at 280 nm. So, while solar background radiation is very low, human sensitivity is correspondingly high, and this is an important consideration in system design.

Thus, for conditions in which low-power consumption, NLOS operability, and security are essential, while range and bandwidth requirements are modest, UV technology provides an interesting alternative to conventional approaches such as radio, and so for example may find broad-ranging applications in data communication on the move, surveillance sensor networks, homeland security, unattended ground sensor (UGS) networks, and small unit communications in urban terrain environments described in [29].

8.1.2 History of NLOS UV communications

Early studies of NLOS UV communications date to the 1960s. An early experimental long-range link employed a high-power xenon flashtube UV source with shortest wavelength of 280 nm, and a photomultiplier tube (PMT) receiver [30]. A transmit–receive separation of 26 km was demonstrated, with an equivalent NLOS path length of 40 km. A NLOS UV system based on an isotropic radiating mercury arc lamp with a modulation rate of 40 kHz was also demonstrated by Fishburne *et al.* [12], and later Puschell and Bayse [20] reported an improved UV local area network test-bed spanning a kilometer range based on a collimated mercury-xenon lamp increasing the modulation rate up to 400 kHz at an effective wavelength of 265 nm. A UV laser communication system at 266 nm that radiated short pulses at high peak power was also reported by Charles *et al.* [4], although the modulation rate was only 600 Hz. During this early period of NLOS experimentation, Reilly and Warde [22] developed an elegant theoretical channel response model to describe the temporal characteristics of scattered radiation in 200~300 nm. Later Luetzgen *et al.* [18] extended the model to examine scattering-induced angular signal distribution and path losses.

The early UV NLOS systems were limited to employing rather bulky, power-hungry, bandwidth-limited flashtube/lamp/laser UV sources. Semiconductor UV optical sources offer the potential of low cost, small size, low power, high reliability, and high bandwidth. State-of-the-art device technology has enabled miniaturized UV light-emitting diodes (LEDs), for example as reported in Bai *et al.* [3] and Shen *et al.* [28]. These LEDs are no bigger than one square millimeter, offer a series of wavelengths in the deep UV region, typically consume electrical power of 150 mW, and radiate optical power of 0.5 mW. Consequently, the arrival of this new generation of UV sources has inspired new studies of LED-based UV systems, with applications including sensing and short-range communications.

Shaw *et al.* [21–27] report on experiments using a 24-element LED array at 274 nm, producing 40 mW optical power. They used a relatively large transmit divergence angle of about 50°. A NLOS link was configured by setting transmitter (Tx) and receiver (Rx) pointing angles equal to 90° (vertical pointing). At the receiver, a Perkin Elmer PMT

(MP-1921) was combined with a solar blind filter to achieve an overall rejection ratio exceeding 10^{10} in the visible through infrared bands. This system was shown to reliably deliver a data rate of 2.4kbps using 4-PPM with transmitter and receiver separation of 11 m (or 200 bps at 100 m). Kedar and Arnon [16] introduced a NLOS optical wireless sensor network based on UV multiple backscattering. A mathematical model for the optical received signal power was developed using the bistatic lidar equation approach. The model was then applied to evaluate the viability and limitations of the inter-node link, and predict the communication performance. Their analysis is valuable for design and study of miniature operational sensor networks.

8.1.3 The communication medium: atmosphere

The atmosphere can be viewed as a complex composition of molecules and aerosols, such as gases, water vapor, pollutants, and other chemical particles. These atmospheric particles interact with the radiation fields generated by an emitter, and cause the signal power losses and wavefront distortion. These interactions are generally classified as scattering and absorption. From a photon propagation point of view, scattering implies a redirection of the angle of propagation, whereas absorption means the photon is annihilated. The magnitude of the atmospheric interaction varies considerably with meteorological conditions, generally becoming stronger as the optical wavelength approaches the cross-sectional dimension of the atmospheric particles (which explains the strong scattering observed in the UV spectrum). Scattering is generally modeled through Rayleigh and Mie theory, as a function of the wavelength and particle size, e.g., [14]. Both of these are included in the models described in the following. Scattering enables NLOS communications, although the great majority of photons are either absorbed or scattered away from the intended detector yielding large path loss.

Atmospheric turbulence due to random non-homogeneity in the atmosphere will cause local changes in the refractive index, that induces distortion in the optical wavefront. This effect is well known in coherent line-of-sight laser communications, where a distorted wavefront arriving at the detector can limit communications rates or received signal quality. Here, we consider NLOS communications due to scattering over relatively short range, and ignore turbulence effects. We consider coherent scattering, where energy lost from the primary field is redistributed to a scatter field which has the same wavelength as the primary. Absorption considers all other interactions (true absorption, incoherent scattering, etc.) in which “absorbed” energy is lost to the total radiation field (primary and coherently scattered radiation).

8.1.4 The NLOS UV scattering channel

The NLOS UV channel involves both molecular scattering and absorption, and aerosol scattering and absorption. We ignore atmospheric turbulence and focus on relatively short-range propagation where turbulence typically plays a very small role. For NLOS UV communications, scattering serves as the vehicle for information exchange between the transmitter and receiver. The scattered light that reaches the receiver depends on the link geometry and the optical properties of the atmosphere, as described next.

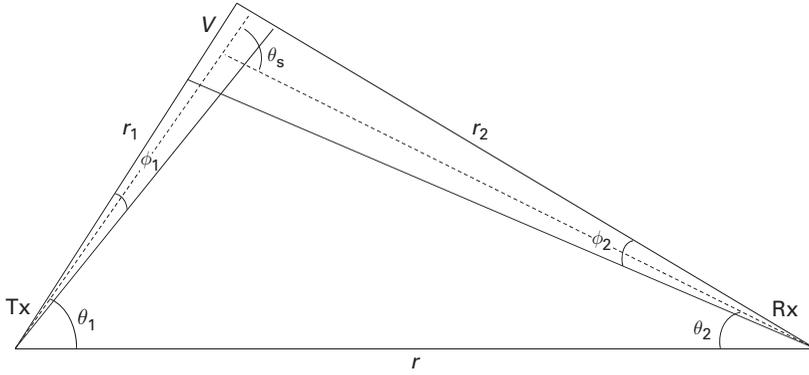


Figure 8.1 NLOS UV communication link geometry, depicting elevated transmitter beam and receiver field-of-view.

Consider a typical NLOS communication geometry illustrated by Reilly and Warde [22] and Luetgen *et al.* [18], as shown in Figure 8.1. Denote the Tx beam full-width divergence angle by ϕ_1 , the Rx field-of-view (FOV) angle by ϕ_2 , the Tx elevation angle by θ_1 , Rx elevation angle by θ_2 , the Tx and Rx baseline separation by r , and the distances of the intersected (overlap) volume V to the Tx and Rx by r_1 and r_2 , respectively.

From the communications viewpoint, scattering and absorption are the two dominant types of photon interactions with the atmosphere over a short communication range. We assume a homogeneous atmosphere and use the following coefficients: the Rayleigh (molecular) scattering coefficient k_s^{Ray} , Mie (aerosol) scattering coefficient k_s^{Mie} , absorption coefficient k_a , and extinction coefficient k_e . The total scattering coefficient is defined as the sum of the two scattering coefficients $k_s = k_s^{Ray} + k_s^{Mie}$, and the extinction coefficient is given by the sum of the scattering and absorption coefficients as $k_e = k_s + k_a$.

The scattering phase function is modeled as a combination of Rayleigh and Mie scattering phase functions based on the corresponding scattering coefficients, as suggested by [34]

$$P(\mu) = \frac{k_s^{Ray}}{k_s} p^{Ray}(\mu) + \frac{k_s^{Mie}}{k_s} p^{Mie}(\mu), \quad (8.1)$$

where $\mu = \cos \theta_s$ is defined from the scattering angle θ_s . The two phase functions follow a generalized Rayleigh model and a generalized Henyey–Greenstein function, respectively, given by

$$p^{Ray}(\mu) = \frac{3[1 + 3\gamma + (1 - \gamma)\mu^2]}{16\pi(1 + 2\gamma)}, \quad (8.2)$$

$$p^{Mie}(\mu) = \frac{1 - g^2}{4\pi} \left[\frac{1}{(1 + g^2 - 2g\mu)^{3/2}} + f \frac{0.5(3\mu^2 - 1)}{(1 + g^2)^{3/2}} \right], \quad (8.3)$$

where γ , g , and f are model parameters.

8.2 Single scattering models

In this section, we investigate the UV scattering channel for NLOS free-space communication links based on a single scattering assumption. An analytical model is reviewed, and an approximate closed-form expression is derived.

8.2.1 Analytical single scattering model

Single scattering models assume that photons traversing the medium between the Tx and the Rx and impinging on the photodetector undergo scattering at most once. Therefore with this model in NLOS geometry all photons arriving at the detector will have undergone scattering once such that their resulting trajectory lands them in the Rx detector. Reilly and Warde [22] proposed an analytical single scattering model for a short-range NLOS UV scattering channel, and Luetzgen *et al.* [18] generalized the model to examine angular spectra and path losses. The model assumes an impulse is transmitted, and predicts the atmospheric scattering and extinction, yielding the channel impulse response. The path loss is then obtained by integrating over the impulse response.

For a NLOS system, the transmitter and receiver are arranged as shown in Figure 8.1. The model is based on a prolate spheroidal coordinate system, illustrated in Figure 8.2. Each point in space is defined by a radial coordinate ξ , an angular coordinate η , and an azimuthal coordinate ϕ . An advantage of this coordinate system is that the sum of the distances between the two foci and any point on a given surface ξ is a constant. This property indicates that a prolate spheroidal surface can be considered as an equitemporal surface, which means that the total single scattering radiation arriving at the receiver at a given time simply requires integration over the surface defined by a given ξ .

It is assumed that an impulse of energy E_t is emitted at time $t = 0$ uniformly over the transmitter solid cone angle into a homogeneous scattering and absorbing medium. The system geometry determines the common volume regarding the transmitter and receiver, from which single-scattered photons could reach the receiver. For simplicity, it is assumed that the axes of the transmitter and receiver lie in a common plane (coplanar geometry), and thus the integration volume is symmetric over coordinate ϕ . The received single scattered intensity at time t can be expressed as in [22] and [18]

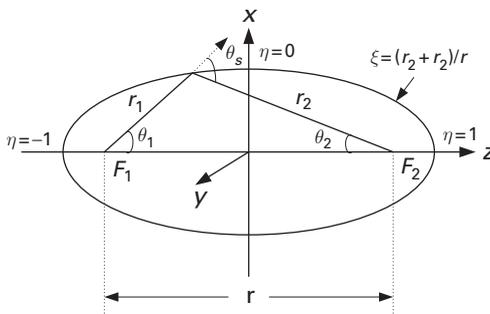


Figure 8.2 Prolate spheroidal coordinate system.

$$h(t) = \frac{E_t A_r c k_s e^{-k_e c t}}{2\pi \Omega_t r^2} \times \int_{\eta_1(ct/r)}^{\eta_2(ct/r)} \frac{2G[\phi(ct/r, \eta)]p(\theta_s)}{(ct/r)^2 - \eta^2} d\eta, \quad (8.4)$$

where the function $G[\cdot]$ is defined in the cited works.

With reference to Figure 8.1, the parameters and functions are defined as follows. Let Ω_t be the solid angle of the Tx cone, r the Tx–Rx baseline separation, and r_1 and r_2 the distances of the common volume to the Tx and Rx, respectively. Define ϕ_1 and ϕ_2 as the Tx beam angle and Rx FOV, and θ_1 and θ_2 as the Tx and Rx focal angles (interchangeable with elevation angles in this context) between each axis and the horizontal axis. The elevation angles are also referred to as pointing or apex angles; perhaps apex is the better term in NLOS scenarios. k_e , k_s , k_a are the atmosphere coefficients, θ_s is the scattering angle, A_r is the area of the receiving aperture, and $p(\theta_s)$ the scattering phase function.

8.2.2 Approximate single scattering path loss model

The impulse response model in (8.4) does not lead to a tractable closed form and so requires numerical analysis. We can obtain a closed-form approximation for $h(t)$ that is amenable to incorporation into communications performance analysis, as follows. The derivation assumes that the geometry is such that the intersecting cones are small. Following the geometry and parameters defined above, the energy at the Rx scattered from the differential volume dV in V can be modeled as [33]

$$\delta E_r = \frac{E_t k_s P(\mu) A_r \delta V \cos(\zeta) e^{-k_e(r_1+r_2)}}{\Omega_t r_1^2 r_2^2}, \quad (8.5)$$

where ζ is the angle between the Rx axis and a vector from the Rx to the common volume, and phase function $P(\mu)$ is given by (8.1). The total energy at the receiver can be found by integrating over the common volume using the prolate spheroidal coordinate system, but this requires numerical evaluation. To obtain a tractable analytical expression, we assume that the common volume is small so that $\zeta \approx 0$, and use this to simplify $h(t)$ and the resulting expression for the total received energy. Note that $\Omega_t = 2\pi(1 - \cos \frac{\phi_1}{2})$, $r_1 = r \sin \theta_2 / \sin \theta_s$, $r_2 = r \sin \theta_1 / \sin \theta_s$, and $\theta_s = \theta_1 + \theta_2$. Now, we can approximate the total received energy as

$$E_r \approx \frac{E_t k_s P(\mu) A_r V \sin^4 \theta_s \exp[-\frac{k_e r}{\sin \theta_s}(\sin \theta_1 + \sin \theta_2)]}{2\pi r^4 \sin^2 \theta_1 \sin^2 \theta_2 (1 - \cos \frac{\phi_1}{2})}. \quad (8.6)$$

Define the path loss as the energy ratio $L = \frac{E_t}{E_r}$. Here L depends on the common volume V that in turn depends on the shape of the intersection. For a small overlap volume, V can be well approximated by a frustum¹ of the right cone of volume $V = \frac{1}{3}\pi(D_1^2 h_1 - D_2^2 h_2)$, where $h_1 = r_1 + r_2 \frac{\phi_2}{2}$ and $D_1 = h_1 \frac{\phi_1}{2}$ are the height and the radius of the bottom surface of the large cone, and $h_2 = r_1 - r_2 \frac{\phi_2}{2}$ and $D_2 = h_2 \frac{\phi_1}{2}$ for the smaller cone. Both cones share a common apex at the Tx location. Substituting r_1 , r_2 ,

¹ The frustum is the portion lying between two parallel planes.

and θ_s into V and the corresponding result into Eq. (8.6), we obtain the following path loss expression

$$L \approx \frac{96r \sin \theta_1 \sin^2 \theta_2 (1 - \cos \frac{\phi_1}{2}) \exp[\frac{k_e r (\sin \theta_1 + \sin \theta_2)}{\sin \theta_s}]}{k_s P(\mu) A_r \phi_1^2 \phi_2 \sin \theta_s (12 \sin^2 \theta_2 + \phi_2^2 \sin^2 \theta_1)}. \quad (8.7)$$

This provides L as an explicit function of the system and geometric parameters, and is easily applied to link analysis for a particular choice of the atmospheric parameters. However, this model only provides high fidelity path loss approximations when the common volume is small, which limits its applicability for geometries with longer baseline range or large common volume overlap such as can occur with wide optical field of view.

8.3 Multiple scattering models

Next we remove the single scattering assumption, and describe models that allow multiple scattering. Generally, higher atmospheric particle density and longer propagation range lead to a higher probability of multiple scattering, and atmospheric particles can be quite abundant and reactive to UV photons. Thus, in some cases, a multiple scattering model can lead to more accurate channel prediction, and the model enables separation and study with respect to the scattering order effect.

As we will show, multiple scattering interactions of photons with the atmospheric constituents result in additional pulse broadening beyond that due to single scattering, i.e., there is increased channel delay spread. The additional pulse spreading occurs because, with multiple scattering, both longer and shorter path lengths are now possible. The pulse spreading limits the channel bandwidth available for communications due to the introduction of inter-symbol interference (ISI). Accounting for multiple scattering also enables more accurate prediction of path loss. With multiple scattering there are more opportunities for photons to successfully transit from the Tx to the Rx, so that the predicted path loss under the multiple scattering model is generally less than that from the corresponding single scattering path loss model. These channel characteristics are a function of transmitter–receiver geometry, including transmitter beam width, elevation angles, range, and receiver FOV. While these are crucial for UV NLOS communications system performance analysis, there is no simple closed-form multi-scattering NLOS UV channel model. In this section, first we review a Monte Carlo numerical technique to obtain both the impulse response and the corresponding path loss, and then we consider an analytical solution for the path loss.

8.3.1 Numerical Monte Carlo multiple scattering channel model

The method described here computes the impulse response due to multiple scattering via Monte Carlo (MC) simulation of each photon's arrival probability along with the associated propagation delay to the Rx. Path loss can then be easily obtained from the area under the impulse response curve as illustrated by [8].

Following [19], the main idea of the MC method is to simulate a complex process as a succession of elementary events whose probability laws are known. Light is decomposed into a set of photons and each individual photon's migration path is probabilistically modeled. This process is repeated for a large number of photons, yielding a statistical picture of the propagation channel. After emission, each photon will follow a migration path along which it may be scattered and/or absorbed. The length of each migration between scattering events is governed probabilistically, as is the departure angle after scattering. Each photon has an initial survival probability that is renewed after each migration. The photon is repeatedly migrated until it either reaches the receiver or its survival probability is smaller than a threshold value whereupon it is considered lost. The process is repeated for many photons, and the aggregated arrival probabilities as a function of time represent the expected received signal intensities, corresponding to the channel impulse response. Figure 8.3 shows a block diagram of the Monte Carlo simulation method, whose details are described next. A recent paper provides pseudo-code for an instantiation of a Monte-Carlo multiple scattering model [11].

The Monte Carlo method begins by launching a photon from the light source located at r_0 into the channel. A collimated or divergent source can be modeled. For the configuration shown in Figure 8.1, each photon's initial direction must be confined

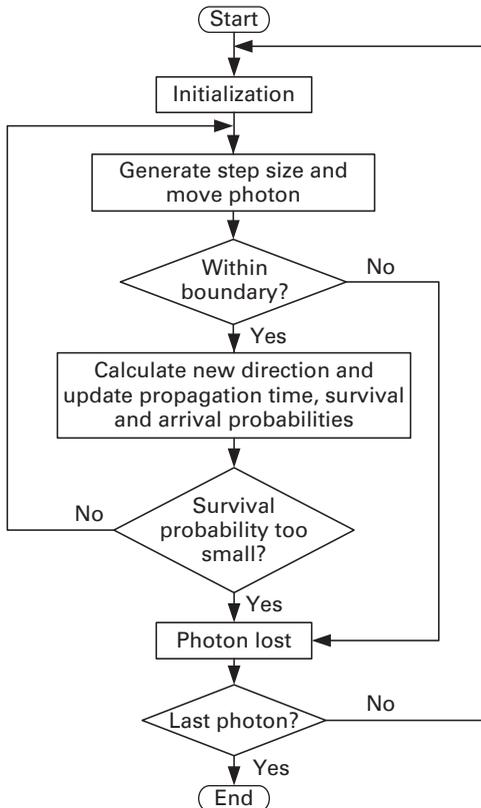


Figure 8.3 Monte Carlo algorithm for multiple scattering impulse response.

within the full beam angle ϕ_1 with solid angle $2\pi(1 - \cos \frac{\phi_1}{2})$. We place the light source in the coordinate system such that the beam axis has a zenith angle $(\pi/2 - \theta_1)$ and azimuth angle 0° . Let the initial direction of an individual photon with respect to the zenith angle be θ_{ini} , and with respect to the azimuth angle be ψ_{ini} . Then, the initial direction from the beam axis is

$$U = \theta_{ini} - (\pi/2 - \theta_1).$$

For a uniformly distributed light source, the initial angles are generated to satisfy

$$\begin{aligned} \cos U &= 1 - \xi^{(\theta)}(1 - \cos \frac{\phi_1}{2}), \\ \psi_{ini} &= 2\pi\xi^{(\psi)}, \end{aligned} \quad (8.8)$$

where $\xi^{(\theta)}$ and $\xi^{(\psi)}$ are independent standard uniform random variables (uniform between zero and one).

The photon trajectory is then simulated by successive migration paths among different scattering centers. After each interaction between a photon and a scattering center, the photon's propagation is assumed to follow the law of single scattering until it reaches the next scattering center or arrives at the receiver. Following this theory, the distance between scattering interactions is given by the random variable

$$\Delta s = -\frac{\ln \xi^{(s)}}{k_s}, \quad (8.9)$$

where $\xi^{(s)}$ is a standard uniform random variable.

An individual photon's spatial migration path can be uniquely described by five variables: three spatial coordinates for the position and two direction angles θ and ψ . The photon's spatial position can be represented with three Cartesian coordinates and the migration direction can be described with three direction cosines from the two direction angles. The direction cosines are specified by taking the cosines of the angles that the photon's direction makes with each of the x , y , and z axes, respectively. As an example, the initial direction cosines are specified by $\mu_x^{ini} = \cos \psi_{ini} \sin \theta_{ini}$, $\mu_y^{ini} = \sin \psi_{ini} \sin \theta_{ini}$, and $\mu_z^{ini} = \cos \theta_{ini}$. For a photon located at (x_i, y_i, z_i) travelling a distance Δs in the direction $(\mu_x^i, \mu_y^i, \mu_z^i)$, its coordinates are updated in the 3-dimensional geometry by

$$\begin{aligned} x_{i+1} &= x_i + \mu_x^i \Delta s, \\ y_{i+1} &= y_i + \mu_y^i \Delta s, \\ z_{i+1} &= z_i + \mu_z^i \Delta s. \end{aligned} \quad (8.10)$$

The new migration direction is governed by the normalized phase function that describes the angular intensity of scattered light. When applied to a single photon, this describes the scattered direction (the scattering angle). The deviation from the current direction is determined from a standard uniform random variable $\xi^{(\mu)}$ and the phase function by [31]

$$\xi^{(\mu)} = 2\pi \int_{-1}^{\mu_1} P(\mu) d\mu. \quad (8.11)$$

Here, $\mu_1 = \cos \theta$, where θ is the angle of the new photon path with respect to the current direction. A numerical solution for θ can be obtained from Eq. (8.11). Note that because there is no azimuth dependence in the phase function, the azimuth angle ψ is then set as a uniformly distributed random variable between 0 and 2π , and so this step provides both angles θ and ψ .

The next step is to compute the direction cosines in the absolute Cartesian coordinate system previously described. If a photon is scattered at angles (θ, ψ) offset from the incoming direction with direction cosines $(\mu_x^i, \mu_y^i, \mu_z^i)$, then the new direction cosines are easily calculated via

$$\begin{aligned}\mu_x^{i+1} &= \frac{\sin \theta}{\sqrt{1 - (\mu_z^i)^2}} (\mu_x^i \mu_z^i \cos \psi - \mu_y^i \sin \psi) + \mu_x^i \cos \theta, \\ \mu_y^{i+1} &= \frac{\sin \theta}{\sqrt{1 - (\mu_z^i)^2}} (\mu_y^i \mu_z^i \cos \psi + \mu_x^i \sin \psi) + \mu_y^i \cos \theta, \\ \mu_z^{i+1} &= -\sin \theta \cos \psi \sqrt{1 - (\mu_z^i)^2} + \mu_z^i \cos \theta.\end{aligned}\quad (8.12)$$

Each photon is assigned a survival probability, initialized to one when first emitted by the source. The probability that the photon arrives at the receiver within its FOV after n scattering steps is given by

$$p_{1n} = \int_{\Omega_n} P(\cos \theta) d\Omega, \quad (8.13)$$

where Ω_n is the solid angle along the scattered direction that can be seen by the receiver aperture of area A_r , and $P(\cdot)$ is the phase function defined earlier. The probability that the photon cannot be seen by the receiver (i.e., it moves out of the receiver FOV) is thus $(1 - p_{1n})$.

Between scattering interactions, the photon is modeled to move with a random step size, and along this path we next consider energy loss between two consecutive scattering centers. The photon location changes from the $(n - 1)$ th scattering center location r_{n-1} to the n th scattering center location r_n according to the direction cosine and coordinate updates described above. Between these scattering centers, the propagation distance $|r_n - r_{n-1}|$ is given by the random variable Δs in Eq. (8.9). Along this path between scatterers, the photon undergoes an energy loss of $e^{-k_a |r_n - r_{n-1}|}$, which is a function of k_a . Impinging upon the n th scattering center, the photon's survival probability is reduced due to this energy loss and so is updated according to

$$w_n = (1 - p_{1n}) e^{-k_a |r_n - r_{n-1}|} w_{n-1}. \quad (8.14)$$

After the n th scattering interaction, the probability that the photon successfully arrives at the receiver is given by

$$\mathcal{P}_n = w_n p_{1n} p_{2n}, \quad (8.15)$$

where p_{2n} represents the propagation loss of the photon from the n th scattering center to the receiver

$$p_{2n} = e^{-k_e|r_n-r'|}, \quad (8.16)$$

r' is the location vector of the receiver, and k_e is employed to model the path loss.

To calculate the channel impulse response, it is necessary to find the propagation time (i.e., relative time delay) of each photon, in addition to its probability of arrival at the detector. This is easily found from the photon migration path, and is given by d_n/c , where d_n is the accumulated propagation distance $d_n = \sum_{j=1}^n |r_j - r_{j-1}|$ and c is the speed of light. For example, suppose a photon undergoes N scattering interactions. Then, associated with this photon are a set of probabilities ($\mathcal{P}_1, \dots, \mathcal{P}_N$), and a corresponding set of propagation times between each pair of scatterers. The Monte Carlo procedure is carried out for a large number of photons (e.g., $M = 10^6$ or more). To display the expected impulse response, the probabilities are superimposed. If we denote each photon energy by E_p , then the probabilities versus time represent the channel response due to M photons. The curve is normalized by $E_p M$ to obtain the impulse response corresponding to a transmitted pulse with energy equal to one joule.

8.3.2 Analytical multiple scattering path loss model

Following the same physical scattering law as the Monte Carlo technique, Ding *et al.* [9] proposed a stochastic analytical method to theoretically derive the n th order scattered signal energy collected by the detector. The model considers all photons to be stochastically scattered and/or absorbed by the atmospheric particles, and involves probabilistic modeling of the photon direction, distance, energy loss, and receiver capture for a specified scattering order. In order to obtain the n th order scattered signal at the receiver, we trace the migration route of a single photon through the medium, similar to [13]. The migration distance and scattering angles follow probability distribution functions (PDFs) that are determined by Beer's law and the phase function, respectively. The PDF of the migration distance r is given by

$$f_r(r) = k_s e^{-k_s r}. \quad (8.17)$$

The zenith scattering angle under combined Rayleigh and Mie scattering has the following PDF

$$f_\theta(\theta) = \frac{k_s^{Ray}}{k_s} f_\theta^{Ray}(\theta) + \frac{k_s^{Mie}}{k_s} f_\theta^{Mie}(\theta), \quad (8.18)$$

where $f_\theta^{Ray}(\theta)$ and $f_\theta^{Mie}(\theta)$ are the phase functions, and θ takes values in $[0, \pi]$. We consider a uniform distribution between 0 and 2π for azimuth scattering angle ϕ because there is no azimuth dependence in the phase function (symmetry is assumed). Thus, the azimuth scattering angle PDF is simply

$$f_\phi(\phi) = \frac{1}{2\pi}, \quad \phi \in [0, 2\pi]. \quad (8.19)$$

Assume a photon from a UV emitter is uniformly emitted at initial direction angles θ_0 and ϕ_0 within the beam divergence, and migrates a distance r_0 before the first scattering occurs. The solid angle within the beam can thus be modeled as a uniform

random variable with a constant PDF of $1/\Omega_t$, where Ω_t is the source beam solid angle given by $\Omega_t = 2\pi(1 - \cos \frac{\phi_t}{2})$. In the current context, we use $(\phi_r, \phi_r, \theta_r, \theta_r)$ as Tx beam width, Rx FOV, Tx and Rx pointing angles respectively, not to be confused with notations in Figure 8.1. The probability that it is scattered in the infinitesimal solid angle $d\Omega_0 = \sin(\theta_0)d\theta_0d\phi_0$ becomes $1/\Omega_t \sin(\theta_0)d\theta_0d\phi_0$, and the probability that it moves an incremental distance dr_0 with attenuation is $e^{-k_a r_0} f_{r_0}(r_0) dr_0$. Therefore the probability that this photon lies away from the scattering center by r_0 and further moves dr_0 along the infinitesimal solid angle is the product of these two probabilities

$$dQ_0 = \frac{e^{-k_a r_0}}{\Omega_t} f_{r_0}(r_0) \sin(\theta_0) d\theta_0 d\phi_0 dr_0. \quad (8.20)$$

After arriving at the first scattering center, the photon is scattered and then continues to move with attenuation. The scattering center is regarded as a secondary point source emitting photons spatially following the scattering angle PDFs described by Eqs. (8.18) and (8.19). From the i th scattering center to the $(i+1)$ th scattering center ($i = 1, 2, \dots$), a similar small probability dQ_i conditioned on all previous events can be written as

$$dQ_i = e^{-k_a r_i} f_{r_i}(r_i) f_{\theta_i}(\theta_i) f_{\phi_i}(\phi_i) \sin(\theta_i) d\theta_i d\phi_i dr_i. \quad (8.21)$$

The same procedure can be successively applied for each scattering process. We assume each scattering is self-governed, and the distances and angles for different scattering events are conditioned on previous quantities. Therefore, the arrival probability for a photon that is scattered n times before arriving at the Rx can be derived based on these transitions. Figure 8.4 shows the photon trajectory corresponding to n possible scatterings.

After the n -th scattering, we focus on an infinitesimal solid angle within the receiver FOV for the receiver to obtain this photon. Let the direction angle ψ_n be the angle between the line connecting the receiver and the n th scattering center, and the transmitter-receiver separation line. Since the FOV has angular range of $[\theta_r - \phi_r/2, \theta_r + \phi_r/2]$, we can confine the photon direction using an indicator function I_n , which equals one when the condition $(\theta_r - \phi_r/2) < \psi_n < (\theta_r + \phi_r/2)$ is satisfied and zero otherwise. Therefore, the probability of the photon leaving the n th scattering center and reaching the detector is

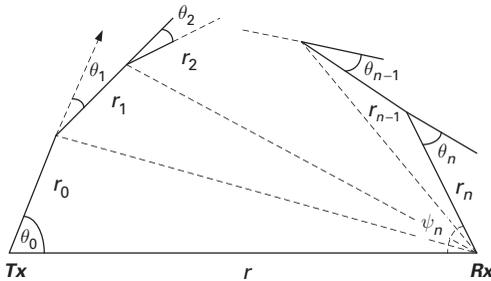


Figure 8.4 Photon migration path for n th order scattering.

$$dQ_n = I_n e^{-k_e r_n} f_{\theta_n}(\theta_n) f_{\phi_n}(\phi_n) \sin(\theta_n) d\theta_n d\phi_n. \quad (8.22)$$

The photon arrival probability after n scatterings is denoted P_n , given by

$$P_n = \int \int \cdots \int dQ_0 \times dQ_1 \cdots \times dQ_n. \quad (8.23)$$

If the UV source emits a pulse with energy E_t , then the received total energy up to n scatterings is

$$E_{r,n} = \sum_{i=1}^n E_r(i) = E_t \sum_{i=1}^n P_i, \quad (8.24)$$

and the corresponding path loss can be expressed as

$$PL_n = \frac{E_t}{E_{r,n}} = \frac{1}{\sum_{i=1}^n P_i}. \quad (8.25)$$

We typically express the path loss in decibels, given by $10 \log_{10}(PL_n)$. As n goes to ∞ , the energy represents the total possible over all scattering orders. Later we will consider some specific cases. Sometimes single scattering dominates, while for some cases scattering up to about 4th-order can make a meaningful contribution.

8.4 NLOS UV channel measurement systems

To validate the channel models, we have experimentally measured both impulse response and path loss, employing a UV pulsed-laser for impulse response and UV LEDs for path loss, respectively. In this section we briefly describe the measurement systems. In the next section we compare channel models and experimental data.

8.4.1 LED path loss measurements

Chen *et al.* [5, 6] describe a solar blind UV path loss experimental test-bed at 260 nm wavelength. As shown in Figure 8.5, the transmitter employed a signal generator to feed a current driver circuit that powered an array of seven ball-lens UV LEDs. The driving current for each LED was 30 mA, yielding an average radiated optical power of 0.3 mW. Using a beam profiler, the beam divergence angle was measured to be 17° . At the receiver, a solar blind filter was mounted on top of the circular sensing window of a Perkin-Elmer PMT module MP1922. The PMT's output current was fed to a high bandwidth preamplifier, whose output was further sent to a photon counter for photon detection. The LEDs, PMT, and filter were attached to Tx and Rx angular control modules, respectively. Each module utilized two perpendicular rotation stages to achieve a precise angular control up to 360° in the azimuth and zenith directions. Experiments reported below have focused on scenarios where the Tx beam axis and Rx FOV axis were coplanar, and only Tx and Rx apex angles were adjusted. The Rx filter had full-width half maximum (FWHM) bandwidth of 15 nm. The filter in-band transmission was 8% and visible-band transmission was below 10^{-10} . The PMT sensing window had

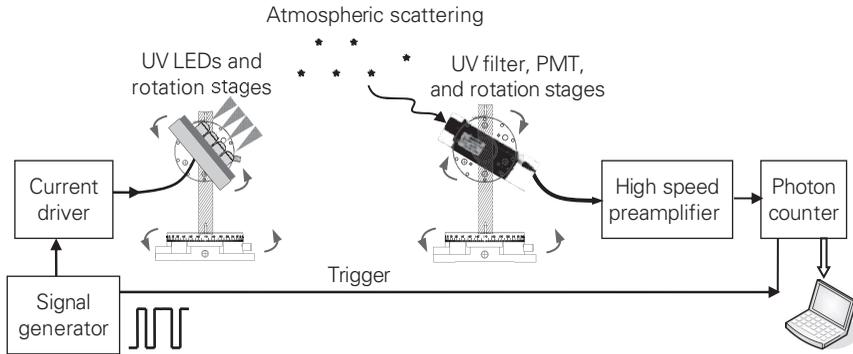


Figure 8.5 NLOS path loss measurement system diagram.

a diameter of 1.5 cm, resulting in an active detection area of $\pi(1.5/2)^2 = 1.77 \text{ cm}^2$. The PMT had an average of 10 dark counts per second (10 Hz), and an in-band UV detection efficiency of 13%. Combining the solar blind filter and PMT, the detector's effective FOV was estimated to be 30° based on the shape and dimension of the filter and PMT.

To obtain path loss L the ratio of transmitted and received power, $10 \log_{10}(P_t/P_r)$ in decibels, was experimentally obtained as follows. In most of the testing configurations, the received power was typically too weak to be measured via a power measurement unit. Consequently, the high sensitivity PMT can be employed for photon counting, and the number of received photons per pulse can be used for calculating L . On the other hand, at the Tx, direct measurement of the total number of emitted photons from the LEDs by a photon counter is impossible because the reading easily overflows the counting limit. Thus, the transmitted optical power was measured using a high-performance power meter and the corresponding photon count was calculated from the measured power and wavelength. Each photon carries energy hc/λ where c is the speed of light, h is Planck's constant, and λ is the wavelength. Denoting the transmitted pulse duration as T_p , then the average number of transmitted photons per pulse is $N_t = P_t T_p \lambda / (hc)$. The number of photons detected per pulse is N_d , which is a percentage of the number of photons N_r impinging on the receiver (the solar blind filter in series with the PMT). This is given by $N_d = \eta_f \eta_r N_r$ where η_f is the filter transmission and η_r is the PMT detection efficiency. Finally, the path loss is given by $10 \log_{10}(N_t/N_r)$ dB.

8.4.2 Laser impulse response measurements

Figure 8.6 depicts a solar blind UV communication impulse response test-bed reported by Chen *et al.* [7]. The critical elements include a high-power short-time pulsed UV source, a high-sensitivity PMT detector, and a solar-blind filter with good out-of-band rejection.

The transmitter was a compact Q-switched fourth harmonic Nd:YAG laser at 266 nm triggered by a rectangular pulse at 10 Hz from a signal generator, producing a corresponding laser pulse train at 10 Hz. Each pulse had width of (3–5) ns and energy

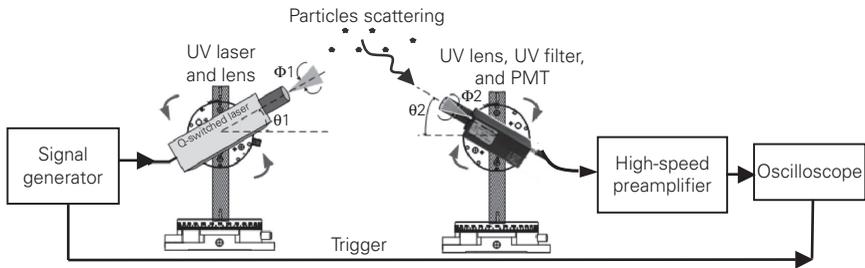


Figure 8.6 NLOS impulse response measurement system diagram.

up to (3–5) mJ, and can be attenuated manually by a factor of 10^3 with an integrated attenuator. The laser head was mounted on a rotation stage with precise motorized angular control. A synchronization signal from the signal generator was output to the oscilloscope at the receiver through a cable. On top of the laser output window, an optical beam expander and a UV focus lens were mounted to adjust the beam angle. The experimental set of beam angles were obtained via offline measurements from a set of lenses of different focal lengths.

A pulsed laser UV source was used to provide sufficient power and bandwidth to make direct impulse response measurements. In a typical measurement scenario, at 100 m baseline range for example, the path loss can exceed 100 dB. No commercial UV laser diodes were available in the solar blind region. Available commercial UV LEDs delivered about 1 mW optical power, with LED arrays extending that up to 50 mW, so these devices did not provide sufficient power for direct impulse response measurement. In addition, the UV LED modulation bandwidths were limited to about 50 MHz, yielding a time domain resolution worse than 20 nsec. Consequently, a Q-switched solid state UV laser was employed. A drawback is that the laser could not be modulated at continuously varying frequencies. Therefore, the frequency domain channel sounding technique successfully applied for indoor infrared channel testing is not applicable, and time domain waveform recording techniques had to be used.

At the receiver, the detector was followed by a customized high gain (34 dB) preamplifier module at 1.5 GHz, feeding either an APD or PMT detector. A 3 GHz oscilloscope was used to view and record the detector output waveform. Figure 8.7 shows the receiver response in a LOS configuration with negligible range, for both APD and PMT detectors. The figure shows the APD has a fast response time of about 1 nsec, with an FWHM pulse width less than 5 ns. The ripple in the tail was mainly due to electric discharge in the detector circuit. While the APD has a fast response time, its relatively small active area limits gain to about 10^2 – 10^4 . Consequently, the APD is not suitable for many high path loss NLOS measurement scenarios of interest due to inadequate weak signal response.

Figure 8.7 shows the PMT detector output measured with LOS at negligible range. While the PMT gain is very high (10^5 – 10^7), its multiple electrode structure tends to spread the response time. The fastest commercial UV PMT available at the time of

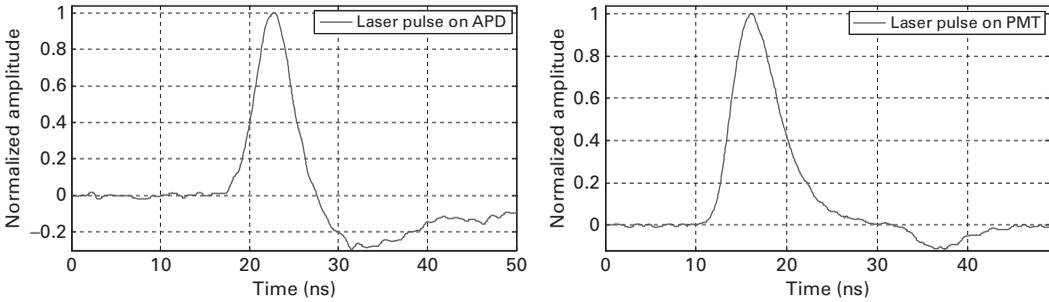


Figure 8.7 APD and PMT photodetector outputs due to an input laser pulse.

the experiment was a Hamamatsu PMT whose response time was about 6 ns with an active diameter of 8 mm. The FWHM pulse width was 6.2 ns, and this represented the minimum time resolution in the channel response measurements due to non-ideal devices in the system, including the laser, detector, and amplifier circuitry. However, as we will show, the channel response typically spans tens to thousands of nano-seconds. Consequently, in the impulse response experimental results reported in the next section, we ignore the small smearing introduced by the PMT detection process.

All subsequent results reported are based on a Hamamatsu PMT module H10304 with an integrated high voltage circuit and active diameter of 8 mm, gain 2.3×10^6 , and a response time of 6 ns. A focusing lens and a 1-inch solar blind UV filter were mounted before the PMT. The Rx FOV was easily changed by using a set of lenses of different focal lengths, as in the Tx beam angle control. A mechanical module incorporated a rotation stage to achieve high-resolution zenith angular control.

8.5 Numerical and experimental results

In this section we evaluate the channel models for some cases of interest, and compare them with experimental measurements taken as described in the previous section.

8.5.1 Numerical results

Table 8.1 compares properties for the four models: Reilly’s single scattering model, our closed-form approximation to Reilly’s model (labeled “Approximation” in the table), our Monte Carlo simulation (labeled “Monte Carlo”), and our analytical path loss model (labeled “Analytical”). The Monte Carlo technique is the most generally applicable, incorporating multiple scattering and yielding both path loss and impulse response estimates. The multi-scattering analytical model yields an expression for the distribution of the probability of photon detection, but we do not have a closed form for the expectation over this distribution, hence the need for the Monte Carlo approach.

Ding *et al.* [9] evaluated the models described above for the impulse response and path loss, varying the geometry including range, and Tx and Rx angles. The parameters

Table 8.1 Scattering model comparison

| Model property | Reilly | Approximation | Monte Carlo | Analytical |
|---------------------|--------|---------------|-------------|------------|
| Closed form | No | Yes | No | No |
| Multiple scattering | No | No | Yes | Yes |
| Impulse response | Yes | No | Yes | No |
| Path loss | Yes | Yes | Yes | Yes |
| Arbitrary geometry | Yes | No | Yes | Yes |

Table 8.2 Atmosphere model parameters

| atmosphere | $k_s^{Ray}(\text{km}^{-1})$ | $k_s^{Mie}(\text{km}^{-1})$ | $k_a(\text{km}^{-1})$ |
|-------------|-----------------------------|-----------------------------|-----------------------|
| tenuous | 0.266 | 0.284 | 0.972 |
| thick | 0.292 | 1.431 | 1.531 |
| extra thick | 1.912 | 7.648 | 1.684 |

were chosen to be realistic with respect to our experimental measurement systems. We assume gas concentrations and optical features of the atmosphere as described by Table II of [15] for middle UV at wavelength 260 nm. As we might expect, variation in the absorption and Mie scattering coefficients indicate that weather conditions may significantly affect the UV signal propagation. However, an explicit correspondence between the parameter settings and weather conditions is not available in the literature. Therefore we consider atmosphere coefficients for three cases: typical tenuous, thick, and extra thick atmosphere conditions (corresponding to clear, overcast, and foggy), given in Table 8.2. The tenuous condition will be adopted for all the following results unless stated otherwise.

Figure 8.8 shows a Monte Carlo simulated impulse response (solid curve). The geometric and model simulation parameters were set as follows: $(\phi_1, \phi_2, \theta_1, \theta_2) = (17^\circ, 30^\circ, 90^\circ, 90^\circ)$, range r at 100 meters, $\gamma = 0.017$, $g = 0.72$, $f = 0.5$, and the detector area is 1.77 cm^2 . Also shown for comparison is a single scattering impulse response model, given by Eq. (3.13a) in [18]. Note that the multiple scattering model predicts both higher intensity and longer duration than the model based on a single scattering assumption, which generally underestimates the effect of scattering.

The multiple scattering path loss for an extra thick (e.g., foggy) atmosphere is plotted in Figure 8.9 for four different pointing angle pairs, with the baseline range varying from 10 to 1000 meters. The higher-order scattering in this example is enhanced due to the thick atmosphere assumption, and highlights the need to account for higher-order scattering in some scenarios. The path loss generally increases as the pointing angles become more vertical, proceeding from panels (a) to (d). The curves are parameterized by scattering order n , and indicate the predicted decrease in path loss as higher-order scattering is incorporated. In these examples the path loss in decibels does not significantly decrease for scattering orders above $n = 4$. Note in Figure 8.9(d) that with vertical pointing angles multiple scattering ($n > 1$) predicts dramatically less path loss than the single scattering ($n = 1$) model.

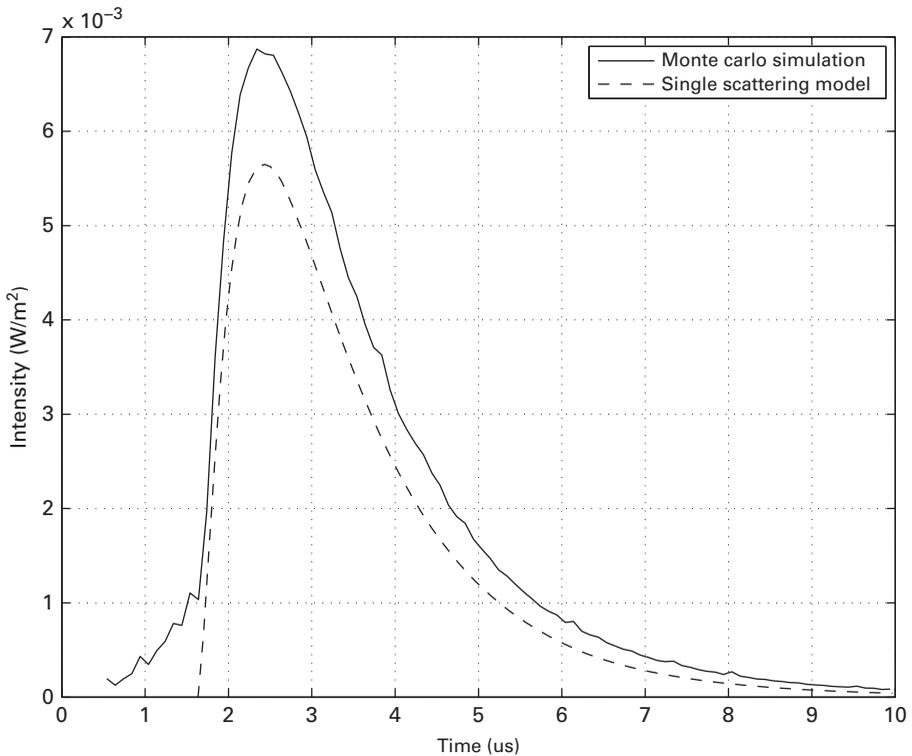


Figure 8.8 Simulated impulse response under single and multiple scattering assumptions. The single scattering assumption yields lower delay spread and intensity.

8.5.2 Experimental results and comparison with simulation

The calibration of our NLOS measurement system was performed in a line-of-sight (LOS) configuration as reported in [6]. Many measurements of path loss and impulse response have been taken outdoors using the systems described in the previous section. We report on some of these, and compare with simulated channel response.

Using a UV LED source array, a variety of path loss measurements were carried out with different Tx/Rx geometries and separation distances. The path loss was calculated as the ratio between the transmitted photons radiated from the UV LEDs and the signal photons impinging upon the receiver. The former was calculated based on the measured source radiated power, and the latter was calculated from measured received photons divided by the total percentage loss from the filter and PMT with receiving area of 1.77 cm^2 . If the path loss per unit area is of interest, then the results can be normalized by this area.

Figure 8.10 from [5] presents path loss versus distance on a logarithmic scale, with each curve parameterized by choice of Tx and Rx elevation angle pairs. We observe that the path loss increases by about 18 dB for each order of magnitude increase in distance r , i.e., path loss is proportional to $r^{1.8}$ under this geometry. However, a path loss exponent

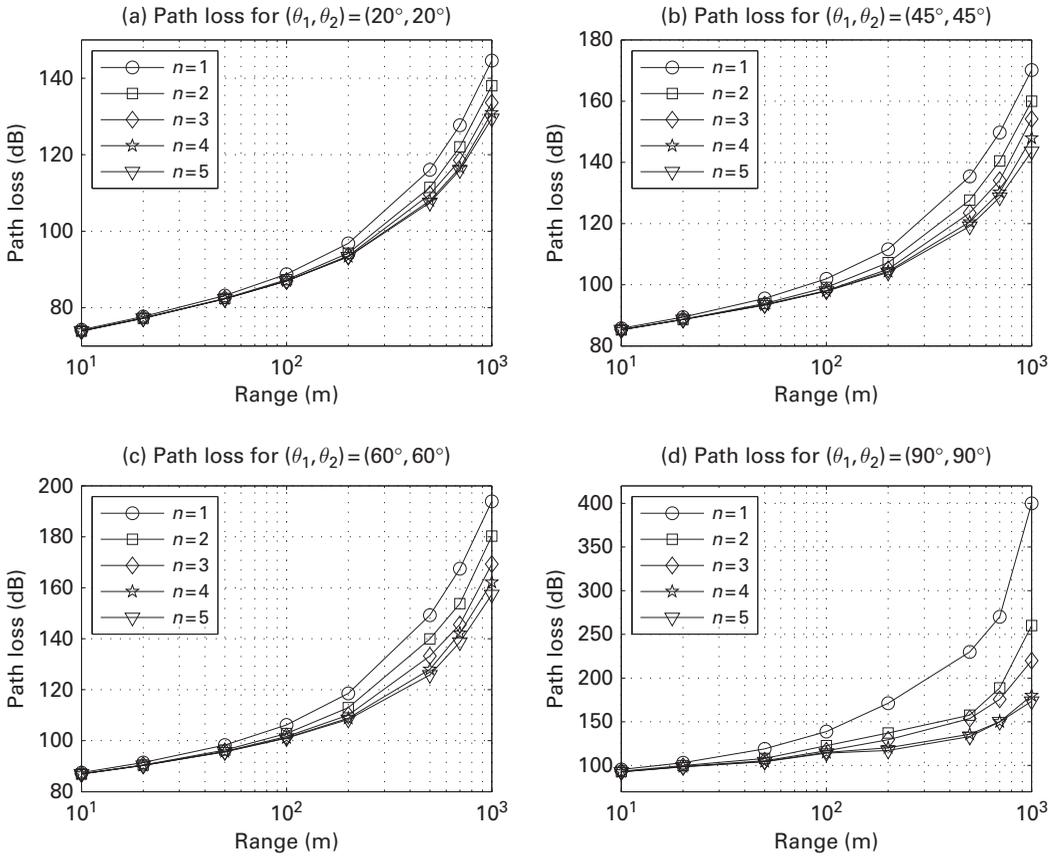


Figure 8.9 Predicted path loss for an extra thick atmosphere.

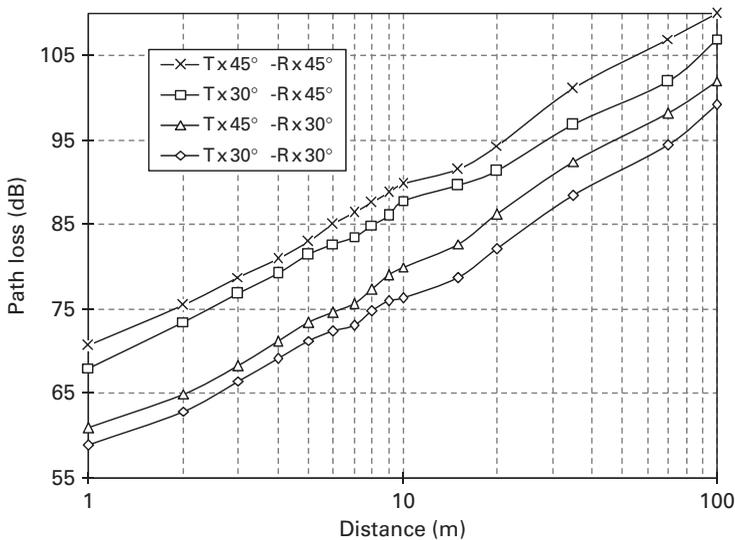


Figure 8.10 Measured path loss for different Tx and Rx elevation angles.

of 1.8 is not universal and may change with geometry. For example, for range up to 10 m and Tx/Rx angle of 90° , the path loss exponent was found to be close to 1.

Recently, a paper by Drost *et al.* [10] used Monte-Carlo simulation to show channel linearity in the absence of a LOS link. They also observed that the nonunity path loss exponent obtained from experiments could be due to a nonideal LED beam pattern and the detector. Weak emission outside the main beam width can create a line of sight path that may impact the measured channel response. This highlights a difficulty with measuring scattering channels.

Figure 8.11 from [32] compares path loss measurement and simulation assuming single scattering. The model parameters matched the experimental measurement system as closely as possible, and assumed a tenuous atmosphere, with the measurements taken outdoors on a clear day. The simulation shows a reasonable agreement, within a few dB. In general, as expected, we observe that the loss increases as either the Tx or Rx angle increases. This is due to the longer propagation path as well as the inherent scattering loss.

A series of impulse response measurements were also obtained, while varying the Tx and Rx elevation angles from 0° to 90° , and changing the Tx beam angle, Rx FOV, and

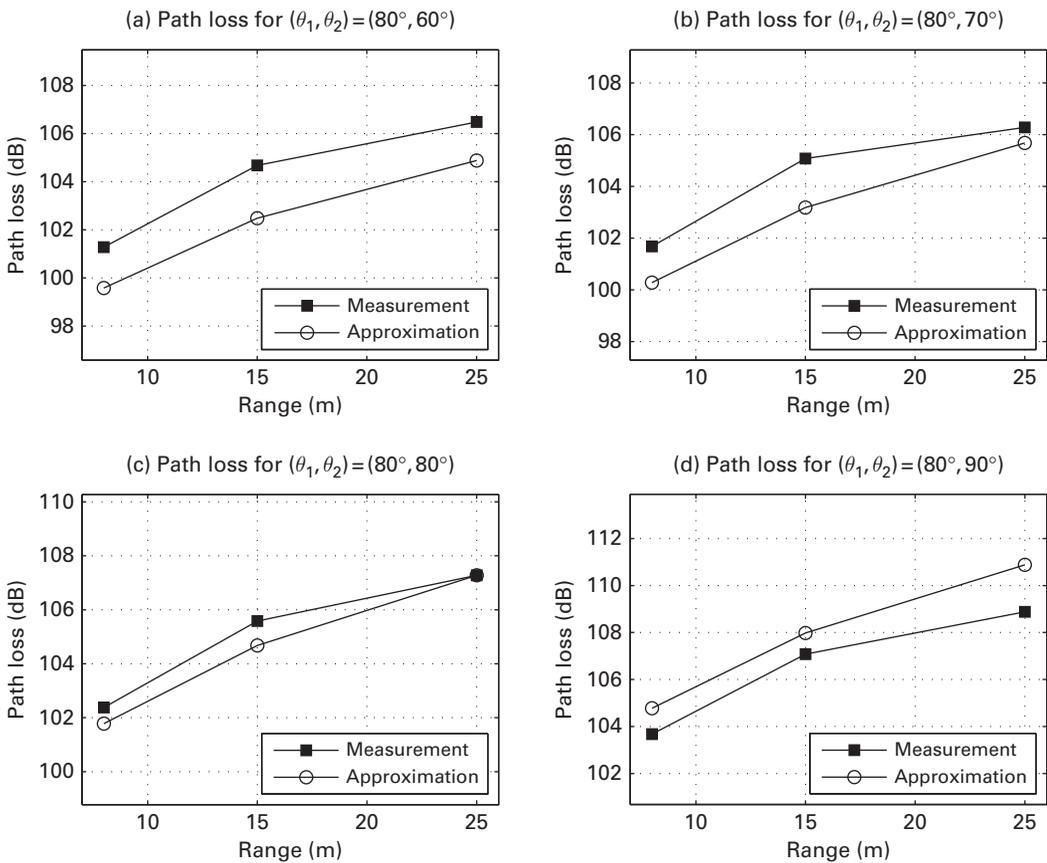


Figure 8.11 Comparison of measured and simulated path loss with four different pointing angle pairs.

baseline distance up to 100 m. The laser pulse rate was set at 10 Hz, and the output energy after the optical beam shaping system was 3 mJ. Beam divergence was less than 3 mrad, and line width was 1 cm^{-1} . The 270 nm solar blind filter had FWHM bandwidth of 15 nm. The peak transmission was 10.4% at 271 nm, and the out-of-band transmission was less than 10^{-8} at 290 nm, 10^{-10} at 390 nm, and 10^{-11} at 305 ~ 750 nm. The spectral mismatch between the laser and the filter was found to be less than 30%. The PMT had a circular sensing window with a diameter of 8 mm, and its dark current was 4 nA.

Figure 8.12 from [7] shows typical average impulse responses under different Tx and Rx pointing angles at a range of 100 meters, progressing from a small angle of 10° to a large angle of 90° . Each figure corresponds to averaging 50 measured impulse responses. The resulting FWHM pulse width varies from 20 ns to 1500 ns. Note the faster rising edge, followed by slower decay. The decay increases as the elevation angle increases, due to the larger spread in propagation times for scattered photons. With a smaller elevation angle (lower to the horizon), the decay becomes significantly faster and the pulse width is decreased. Note also that, at higher elevation angles (approaching vertical pointing), the responses have more variation from measurement to measurement, and the averaged response continues to exhibit significant fluctuation, even after averaging 50 measurements.

Figure 8.13 compares simulated impulse response with measured response, for four different pointing angle pairs. The pulse spreading is roughly (200, 300, 400, 500) ns for the four pointing geometries indicated in the figure captions. There is a reasonable agreement between the simulation and experimental results. However, other cases have shown

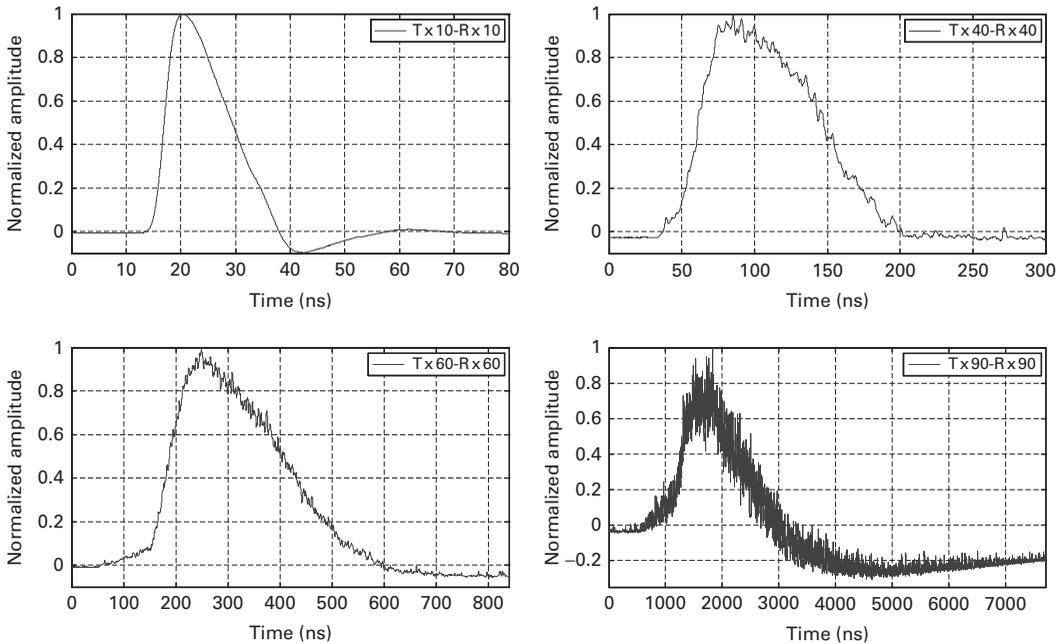


Figure 8.12 Impulse response measurements with four different Tx and Rx pointing angle pairs.

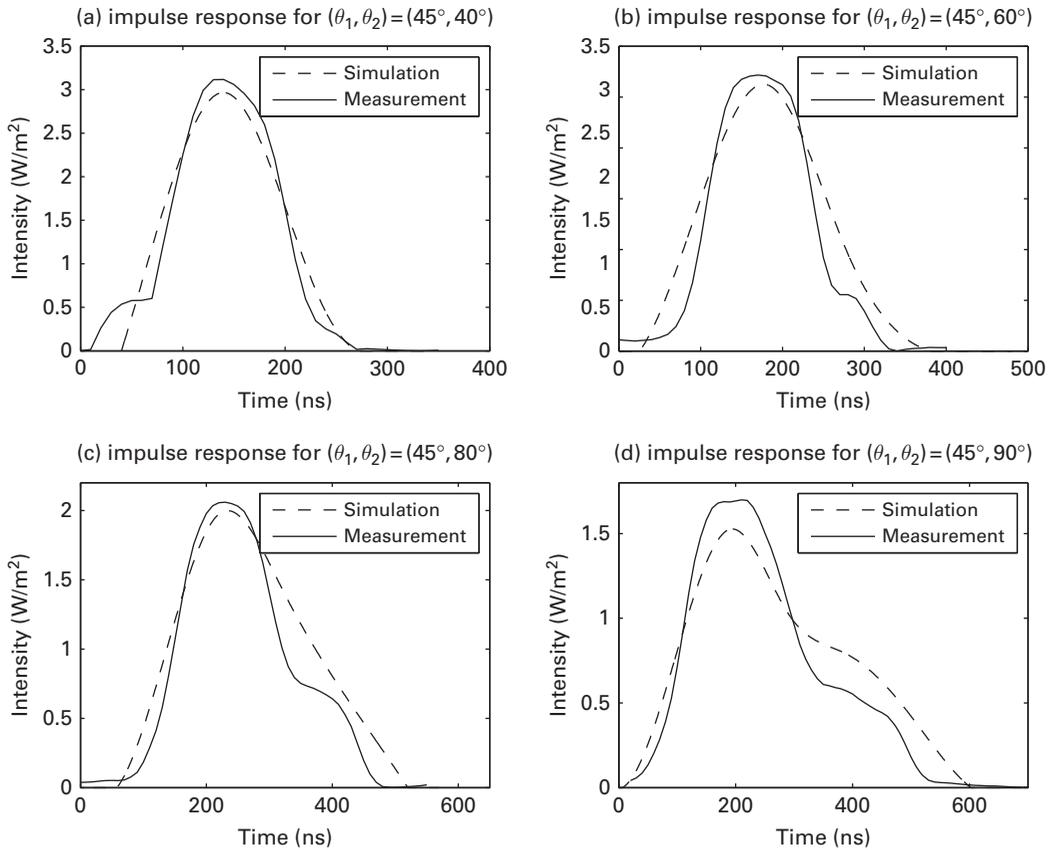


Figure 8.13 Comparison of impulse response simulation with experimental measurement.

larger error between the measurement and simulation, and continued investigation is necessary. We conjecture that some of the mismatch between experiment and prediction can be attributed to system calibration errors, inaccurate atmospheric parameters for the system operating conditions, and other measurement errors.

8.6 Summary

This chapter presented analytical and experimental methods for investigating the NLOS UV scattering channel. Both impulse response and path loss were modeled and analyzed, and simulation and experimental measurements were shown and compared. Incorporating multi-scattering theory into NLOS UV channel modeling yields more accurate path loss and channel delay spread estimates than single scattering theory, especially for larger pointing angles (approaching vertical). The models provide reasonably accurate path loss prediction for the cases studied, at baseline ranges up to 100 meters, as validated by extensive outdoor measurements. These results are useful for

system design and communications link performance analysis. Topics of further interest include developing an analytical impulse response model, developing a stochastic path loss model, analyzing non-coplanar propagation geometry, study of multi-user channels and multi-sensor cases, and modeling turbulence effects at longer ranges.

References

- [1] International Commission on Non-Ionizing Radiation Protection (ICNIRP): Guidelines on limits of exposure to ultraviolet radiation of wavelengths between 180nm and 400nm (incoherent optical radiation), 2004, initially published in *Health Physics* 49:331–340, 1985; amended in *Health Physics* 56:971–972, 1989; reconfirmed by ICNIRP in *Health Physics* 71:978, 1996; republished in *Health Physics* 87(2):171–186.
- [2] International Electrotechnical Commission (IEC), IEC 60825-12: Safety of laser products – part 12: Safety of free space optical communication systems used for transmission of information, 2005.
- [3] Bai, X., McIntosh, D., Liu, H., and Campbell, J. C. Ultraviolet single photon detection with Geiger-mode 4H-SiC avalanche photodiodes, *IEEE Photon. Technol. Lett.*, **19(22)**, 1822–1824, 2007.
- [4] Charles, B., Hughes, B., Erickson, A., Wilkins, J., and Teppo, E. An ultraviolet laser based communication system for short range tactical applications, *Proc. of SPIE*, **2115**, 79–86, 1994.
- [5] Chen, G., Abou-Galala, F., Xu, Z., and Sadler, B. M. Experimental evaluation of LED-based solar blind NLOS communication links, *Optics Express*, **16(19)**, 15059–15068, 2008.
- [6] Chen, G., Xu, Z., Ding, H., and Sadler, B. M. Path loss modeling and performance trade-off study for short-range non-line-of-sight ultraviolet communications, *Opt. Express*, **17(5)**, 3929–3940, 2009.
- [7] Chen, G., Xu, Z., and Sadler, B. M. Experimental demonstration of ultraviolet pulse broadening in short-range non-line-of-sight communication channels, *Opt. Express*, **18(10)**, 10500–10509, 2010.
- [8] Ding, H., Chen, G., Majumdar, A. K., Sadler, B. M., and Xu, Z. Modeling of non-line-of-sight ultraviolet scattering channels for communication, *IEEE J. Sel. Topics Comm.*, **27(9)**, 1535–1544, 2009.
- [9] Ding, H., Xu, Z., and Sadler, B. M. A path loss model for non-line-of-sight ultraviolet multiple scattering channels, *EURASIP Journal on Wireless Communications and Networking: Special Issue on Optical Wireless Communications and Networking*, **2010(598572)**, 1–12, 2010.
- [10] Drost, R. J., Moore, T. J., and Sadler, B. M. Monte-Carlo-based multiple-scattering channel modeling for non-line-of-sight ultraviolet communications, *Proc. SPIE*, 8038, 803802-1–803802-9, April 2011.
- [11] Drost, R. J., Moore, T. J., and Sadler, B. M. UV communications channel modeling incorporating multiple scattering interactions, *JOSA A*, **28(4)**, 686–695, 2011.
- [12] Fishburne, E. S., Neer, M. E., and Sandri, G. Voice communication via scattered ultraviolet radiation. Final report of Aeronautical Research Associates of Princeton, Inc., NJ, 1976.
- [13] Gillespie, D. T. Stochastic-analytic approach to the calculation of multiply scattered lidar returns, *J. Opt. Soc. Am. A.*, **2(8)**, 1307–1324, 1985.
- [14] Ishimaru, A. *Wave Propagation and Scattering in Random Media*. Academic Press, 1978.

- [15] Junge, D. M. Non-line-of-sight electro-optic laser communications in the middle ultraviolet, M.S Thesis. Naval Postgraduate School, Monterey, MA, 1977.
- [16] Kedar, D. and Arnon, S. Non-line-of-sight optical wireless sensor network operating in multiscattering channel, *App. Optics*, **45(33)**, 8454–8461, 2006.
- [17] Koller, L. R. *Ultraviolet Radiation*. 2nd edn., John Wiley & Sons, 1965.
- [18] Luetngen, M. R., Shapiro, J. H., and Reilly, D. M. Non-line-of-sight single-scatter propagation model, *J. Opt. Soc. Am.*, **8(12)**, 1964–1972, 1991.
- [19] Mobley, C. D. *Light and Water: Radiative Transfer in Natural Waters*. Academic Press, 1994.
- [20] Puschell, J. J. and Bayse, R. High data rate ultraviolet communication systems for the tactical battlefield, *Proc. Tactical Communications Conf.*, **1**, 253–267, 1990.
- [21] Reilly, D. M., Moriarty, D. T., and Maynard, J. A. Unique properties of solar blind ultraviolet communication systems for unattended ground sensor networks, *Proc. of SPIE*, **5611**, 244–254, 2004.
- [22] Reilly, D. M. and Warde, C. Temporal characteristics of single-scatter radiation, *J. Opt. Soc. Am.*, **69(3)**, 464–470, 1979.
- [23] Shaw, G. A., Fitzgerald, J., Nischan, M. L., and Boettcher, P. W. Collaborative sensing test bed and experiments, *Proc. of SPIE*, **5101**, 27–38, 2000.
- [24] Shaw, G. A., and Nischan, M. L. Short-range NLOS ultraviolet communication testbed and measurements, *Proc. of SPIE*, **4396**, 31–40, 2001.
- [25] Shaw, G. A., Nischan, M. L., Iyengar, M. A., Kaushik, S., and Griffin, M. K. NLOS UV communication for distributed sensor systems, *Proc. of SPIE*, **4126**, 83–96, 2000.
- [26] Shaw, G. A., Siegel, A. M., Model, J., and Greisokh, D. Recent progress in short-range ultraviolet communication, *Proc. of SPIE*, **5796**, 214–225, 2005.
- [27] Shaw, G. A., Siegel, A. M., and Model, J. Extending the range and performance of non-line-of-sight ultraviolet communication links, *Proc. of SPIE.*, **6231**, 62310C.1–62310C.12, 2006.
- [28] Shen, S. C, Zhang, Y., Yoo, D. *et al.* Performance of deep ultraviolet GaN avalanche photodiodes grown by MOCVD, *IEEE Photon. Technol. Lett.*, **19(21)**, 1744–1746, 2007.
- [29] Stotts, L. B. and Bryski, D. D. Unattended ground sensor related technologies: an Army perspective, *Proc. of SPIE*, **4743**, 1–9, 2002.
- [30] Sunstein, D. E. A scatter communications link at ultraviolet frequencies, B.S Thesis. MIT, 1968.
- [31] Witt, A. N. Multiple scattering in reflection nebulae I: A Monte Carlo approach, *Astrophys. J. Suppl. Ser.*, **35**, 1–6, 1977.
- [32] Xu, Z., Ding, H., Sadler, B. M., and Chen, G. Analytical performance study of solar blind non-line-of-sight ultraviolet short-range communication links, *Opt. Lett.*, **33(16)**, 1860–1862, 2008.
- [33] Xu, Z. and Sadler, B. M. Ultraviolet communications: potential and state-of-the-art, *IEEE Commun. Mag.*, **46(5)**, 67–73, 2008.
- [34] Zachor, A. S. Aureole radiance field about a source in a scattering—absorbing medium, *Appl. Opt.*, **17(12)**, 1911–1922, 1978.

9 Free-space optical communications underwater

Brandon Cochenour and Linda Mullen

Wireless optical communications underwater is enjoying a renewed interest from researchers due to the wide advances in laser sources and receivers, digital communications, and signal processing. Underwater free space optics (uFSO) fulfills several niche applications for wireless communications in ocean waters. While RF communications have become ubiquitous in our everyday lives above water, the RF portion of the electromagnetic spectrum exhibits high attenuation in seawater. Acoustics on the other hand have long enjoyed success for detection and communication underwater, given their ability to propagate long distances underwater ($>$ km). However, for high-speed data transfer ($>$ Mbps), acoustics are at a disadvantage, as it is well known that acoustic energy exhibits increasing attenuation with increasing frequency. Supported by enormous growth in the telecommunications industry over the past few decades, optical techniques are garnering serious consideration for underwater communications due to the high data rates they may provide. Additionally, as we will learn, the blue/green portion of the visible spectrum exhibits minimal absorption in seawater. Still, scattering of light by organic and inorganic particulates in ocean water can cause significant spatial and temporal dispersion, which may have a measurable impact on link range and available bandwidth.

This chapter serves as both an introduction to the field of light propagation underwater, as well as a survey of current literature pertaining to uFSO. We begin with a simple examination of a link budget equation. Next, we present an introduction of ocean optics in order to gain an appreciation for the challenges involved with implementing free-space optical links underwater. We then discuss state-of-the-art theoretical and experimental methods for predicting beam propagation in seawater. Finally, we present some common uFSO link types, and discuss the system-level design issues associated with each.

9.1 Introduction: towards a link equation

The goal of any communications receiver, optical or otherwise, is to maximize the signal-to-noise (SNR) before decisions on bits are made. In optical communications, the

Advanced Optical Wireless Communication Systems, ed. Shlomi Arnon, John R. Barry, George K. Karagiannidis, Robert Schober, and Murat Uysal. Published by Cambridge University Press.
© Cambridge University Press 2012.

SNR is directly proportional to the optical power that is detected by the photoreceiver. While the exact expression will vary depending on the link geometry and type, let us assume that the transmitter and receiver are positioned in a line-of-sight configuration, where the receiver is accurately aligned with the beam axis. The received power in this case may be modeled by,

$$P_R(t, r, z) = P_T(t) D_T L_W(t, r, z) D_R \quad (9.1)$$

where $P_R(t, r, z)$ is the received power as a function of time (t), lateral displacement from the beam axis (r), and range (z); $P_T(t)$ is the transmitted power, D_T represents the aperture and divergence of the optical source, and D_R represents the photoreceiver aperture and field of view. The most important term, $L_W(t, r, z)$, is the channel loss term. *Wrapped up in this single term are decades worth of experimental and theoretical research that has aimed to characterize the spatial and temporal characteristics of light propagation in seawater.*

We might ask ourselves, “what exactly do we mean by channel *loss*?”. It may be convenient to describe “loss” in terms of dispersion, in both space and time. With *spatial dispersion*, we mean that the initial laser source distribution (collimated) has spread out in space due to either natural divergence, or most likely, scattering by particles in the water. Certainly, this means that fewer photons are incident on the photoreceiver and hence, a lower SNR. With *temporal dispersion*, we refer to a loss of modulation depth, or contrast, of the information signal placed upon the optical carrier. Temporal dispersion underwater is result of spatial dispersion. In other words, a photon that has been multiply scattered accumulates a path difference relative to a photon that has not been altered by scattering. As such, spatial dispersion can limit the maximum signalling bandwidth available in the uFSO channel.

While the end result of uFSO channel loss may be similar to that of other wireless optical channels (i.e. – reduced range, bandwidth, reliability, etc.), the processes by which they occur can be markedly different. In order to gain appreciation for this fact, in the next section we introduce some basic fundamentals in ocean optics. This will serve as a foundation for understanding how different ocean environments influence the channel loss term.

9.2 Introduction to ocean optics

In this section, we aim to gain further insight into the complex and dynamic nature of the free space underwater optical channel. The biological and chemical processes that take place in seawater produce a metaphorical “soup” of particles, both organic and inorganic.

Typically, we are concerned with “optically significant” particles, i.e. ones that contribute to a measurable change to the optical properties of ocean water. These properties can be segregated into two broad areas: inherent or apparent. Inherent optical properties (IOPs) are those that are dependent only upon the medium. Examples include the absorption and scattering coefficients, as well as the index of refraction. Apparent

optical properties (AOPs) depend on *both* the medium *as well as* the distribution of the incident light source used to probe the medium *and the characteristics of the system used to measure the optical signal*. Examples include the system attenuation coefficient, or the diffuse attenuation coefficient which describes the attenuation of sunlight irradiance versus depth. Depending on the circumstance, we may express link loss in terms of either IOPs or AOPs.

We will focus mainly on IOPs, namely absorption and scattering. Consider the following experiment where a collimated laser source and photoreceiver are placed opposite each other, a distance z meters apart. Obviously, with no medium between the two, there is no change in optical power along the path z , or $dP/dz = 0$. Now consider the transmitter and receiver placed in an absorbing and scattering medium (see Figure 9.1). What is the change in power over the path z in the presence of absorption and scattering? Intuitively, the change in power should be proportional to the initial power, and the path over which scattering and absorption may occur. In this case we have,

$$dP = cPdz \quad (9.2)$$

where a proportionality constant c (in units of m^{-1}) has been introduced to account for losses due to absorption and scattering. In the limit of thin slabs, Eq. (9.2) is a first-order differential equation. Integrating both sides yields,

$$P(z) = P_0 \exp(-cz). \quad (9.3)$$

This result is commonly referred to as Beer's Law, and describes the power received $P(z)$, given initial power P_0 , after traveling through a medium of thickness z . We refer to the variable c (m^{-1}) as the *beam attenuation coefficient*. Note that the total attenuation is equal to the sum of absorption and scattering, or $c = a + b$. In a completely absorbing or completely scattering medium, the total attenuation coefficient in Eq. (9.3) can be replaced with the absorption coefficient ($a \text{ m}^{-1}$) or scattering coefficient ($b \text{ m}^{-1}$), respectively. As we will later see, the absorption, scattering, and total attenuation coefficients are functions of wavelength. It is also common to refer to the product cz as the *attenuation length*. For example, at one attenuation length, the received power would be decreased by a factor of $\exp(-1)$, or $\sim 63\%$.

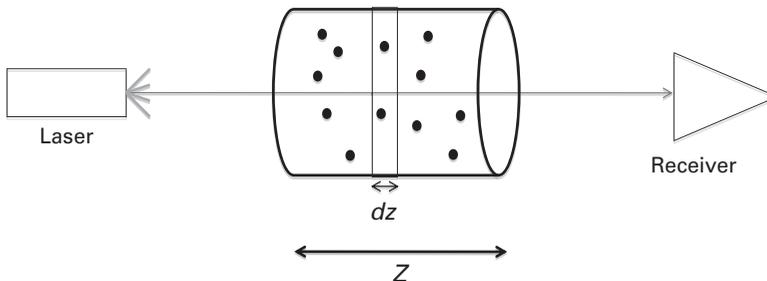


Figure 9.1 The differential loss in intensity of an optical beam propagating in an absorbing and scattering medium.

Beer's Law represents our first step towards characterizing the channel loss term, $L_W(t, r, z)$, in Eq. (9.1). Obviously, Beer's Law is only a partial estimation of channel loss as it ignores the temporal and spatial dependency inherent in $L_W(t, r, z)$ and only includes range dependency. In fact as we will later see, the applicability of Beer's Law is limited only to a certain amount of attenuation lengths even in this simple LOS configuration. In these cases, attenuation is still exponential, but with a coefficient other than $c = a + b$. Regardless, the point we wish to make here is that overcoming an exponential loss term when designing the uFSO link is no trivial task. Nature has not been kind.

It may now appear obvious that in order to further characterize the overall attenuation, we must know more about the absorption and scattering properties of the organic and inorganic particles found in seawater. The following sections present a brief introduction to the processes that govern absorption and scattering in seawater.

9.2.1 Absorption

9.2.1.1 Absorption by pure water

We first consider the absorption properties of pure water, without the presence of any optically significant particles. Here, photons are absorbed through a transfer of energy to water molecules by way of ionization/electronic excitation, or vibrational or rotational excitation of water molecules. The absorption spectra of pure water has been exhaustively measured over the decades with increasingly accurate apparatuses (for example, [47], [43]). A summary of these measurements is shown in Figure 9.2 from 300nm–700nm where we see that the absorption coefficient exhibits a clear minimum in the “blue/green” portion of the spectrum between 400–500nm. The variability between measurements is due to the extreme sensitivity and wide dynamic range needed by instruments to measure absorption coefficients that are already quite negligible to begin with.

It had been posited that perhaps the variation in absorption coefficient observed in the blue/green region may have been due to variations in temperature and/or salinity between studies. However, studies have shown that the dependence on temperature and salinity of the absorption spectra of both pure and salt water is rather negligible for visible wavelengths [39].

To further illustrate the potential impact of pure water absorption on the uFSO link, the link loss as a function of range using results from [43], is shown in Figure 9.3 at various wavelengths. Note that red wavelengths exhibit > 25 dB more attenuation than blue/green wavelengths at 10 m in pure water. The loss is even more devastating at IR wavelengths used in atmospheric FSO or fiber optic links. In other words, from a system design point of view, the uFSO link designer is largely unable to take advantage of the high degree of technical maturity of many of the IR technologies developed for other optical communication links, as IR wavelengths exhibit significantly more absorption in water than visible wavelengths do. We will discuss more about these system trade-offs later in the chapter.

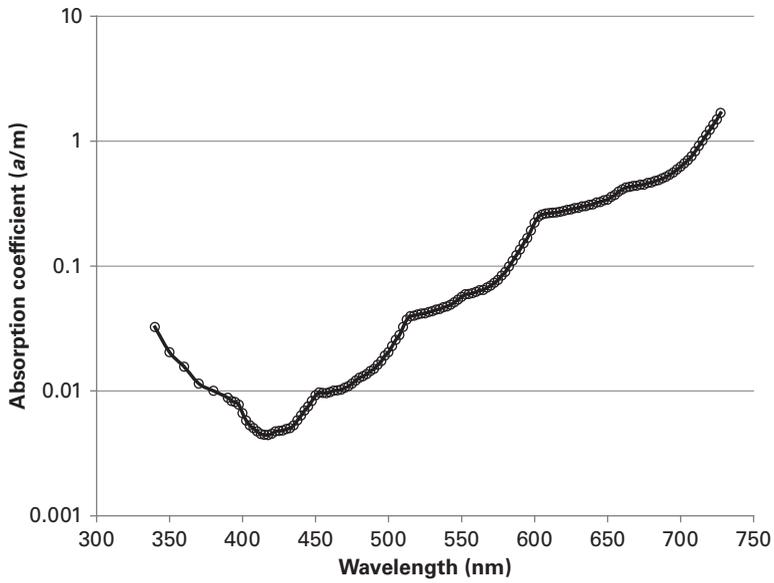


Figure 9.2 Absorption properties of pure water between 340–700nm. Data reproduced from [47] (340–390nm), and [43] (400–700nm)

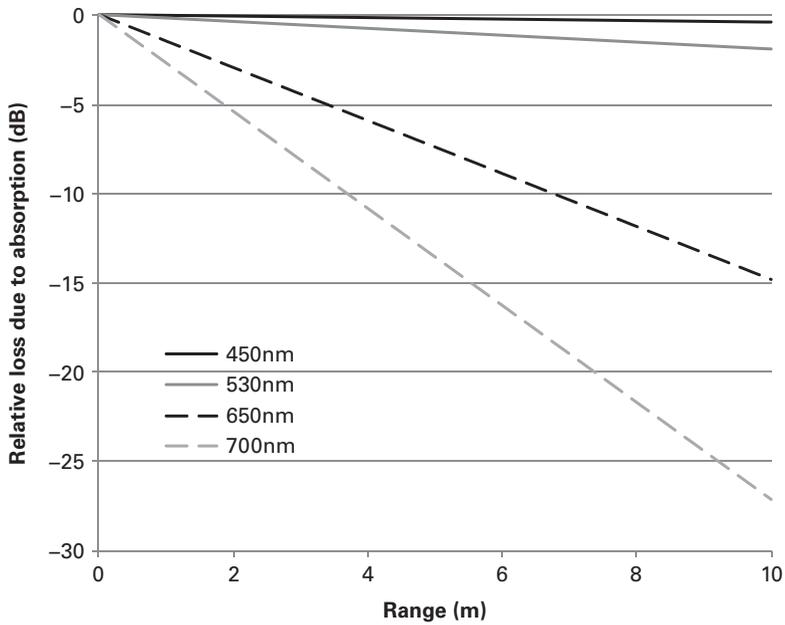


Figure 9.3 Comparison of the attenuation (due to absorption only) versus distance for pure water only for several wavelengths. (Using data of [43].)

9.2.1.2 Absorption by colored dissolved organic matter

Colored dissolved organic matter (CDOM) (also known as yellow substance or gelbstoff) is the result of dead or decaying organisms and cells. Absorption due to CDOM is high in the UV and blue, and typically falls off exponentially with wavelength (see for example, [26]). As we will later see, this tends to shift the wavelength of minimum absorption from the blue into the green.

9.2.1.3 Absorption by phytoplankton

The word “phytoplankton” is derived from Greek words that translate literally to “wandering plant.” As its name suggests, phytoplankton produce energy via photosynthesis, and therefore it is often found near the water surface where sunlight is abundant. As such, phytoplankton plays a significant role in the absorption of light in the sea. Chlorophyll-*a* is a main “ingredient” in the phytoplankton photosynthesis process, and tends to exhibit strong absorption to UV, blue, and red wavelengths ([33], [44]). (For the non-specialist, this is precisely the reason that the leaves of trees, which also contain chlorophyll-*a*, appear green).

While beyond the scope of this discussion, the impact of chlorophyll-*a* on the absorption coefficient in seawater depends on the amount of pigment contained within a phytoplankton species as well as the concentration of phytoplankton in the water. Additionally, concentration and species can be depth dependent. Suffice to say, the window of minimum absorption at blue wavelengths for pure water observed in Figure 9.2, tends to shift towards green wavelengths in waters containing phytoplankton.

9.2.2 Scattering

Scattering is the process whereby a photon’s path is altered by interaction with a molecule or particle. Mobley [33] separates scattering into three broad regimes:

1. *Molecular scattering* ($\ll \lambda$): Scattering caused by local molecule density fluctuations. For example, local concentrations of sea salts in pure seawater.
2. *Scattering by “large” particles* ($> \lambda$): Particles typically 10 times λ are considered “large” particles. These are the organic and inorganic particles that typically dominate the scattering properties of seawater.
3. *Turbulent scattering* ($\gg \lambda$): Temperature, pressure, or salinity changes that cause significant changes in the index of refraction.

9.2.2.1 Molecular scattering

Scattering by pure water arises from the random nature of molecular motion. Scattering by pure seawater is similar, though with the included contribution of various salt ions found in seawater. The scattering coefficient for pure water and pure seawater is found in Figure 9.4 [34]. The slight increase in pure seawater is due to the extra contribution from salt ions. However, in relation to the absorption coefficient, Figure 9.4 shows that scattering of pure water contributes minimally to the overall attenuation. As such, we will focus our discussion on the scattering caused by large particles.

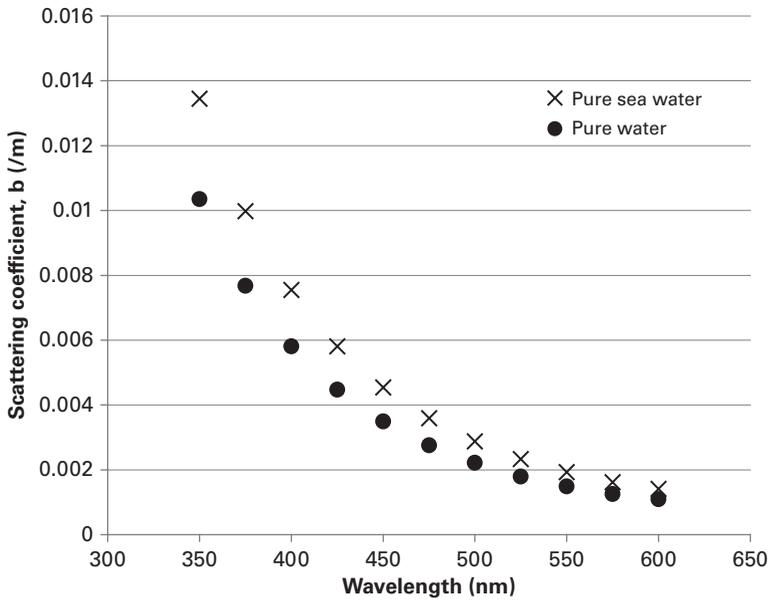


Figure 9.4 The scattering coefficient for pure water and pure seawater as a function of wavelength. Losses due to scattering are less than those due to absorption for pure water at blue/green wavelengths. Data from [34].

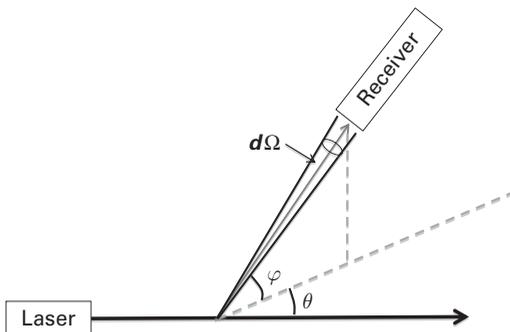


Figure 9.5 Geometry for defining the VSF.

9.2.2.2 Scattering by large particles

Perhaps the most obvious question to ask when beginning a discussion on scattering by large particles is that given a scattering particle(s), what is the angular distribution of the scattered light? To answer this, it is instructive to consider the following mock experiment in Figure 9.5.

A collimated beam of power P_0 is incident upon an infinitesimally thin slab of particles of length dz . Light scatters into the *polar angle* θ and *azimuth angle* ϕ . At distance r , a detector with area A subtending a solid angle $d\Omega$ measures the scattered intensity. The received power at the detector can then be expressed as,

$$P_S(\theta, \phi) = P_0 \beta(\theta, \phi) dz d\Omega \quad (9.3)$$

where we have introduced a term $\beta(\theta, \phi)$ that describes the angular distribution of light. Rearranging terms,

$$\beta(\theta, \phi) = \frac{P_S(\theta, \phi)}{P_0 dz d\Omega} \quad (9.4)$$

Note that the received scattered differential *intensity* is $dI(\theta, \phi) = P_S(\theta, \phi) / d\Omega$ (W sr^{-1}). Also, if the incident power illuminates an area on the slab dA , then the incident *irradiance* is $E = P_0 / dA$ (W m^{-2}). If we consider the incident beam illuminates an infinitesimal volume $dV = dz dA$, then Eq. (9.4) can be re-written as,

$$\beta(\theta, \phi) = \frac{dI(\theta, \phi)}{E dV} \quad (9.5)$$

Thus, the physical meaning of $\beta(\theta, \phi)$ is the scattered intensity (in direction (θ, ϕ)) per unit incident irradiance per unit volume. Therefore, $\beta(\theta, \phi)$ is often referred to as the *volume scattering function* (VSF) and has the units $\text{m}^{-1} \text{sr}^{-1}$. Integrating the VSF over all 4π steradians is analogous to determining the “strength” of scattering over all angles.

$$b = \int_{\phi} \int_{\theta} \beta(\theta, \phi) \sin \theta d\theta d\phi = 2\pi \int_0^{\pi} \beta(\theta) \sin \theta d\theta \quad (9.6)$$

We refer to b as the scattering coefficient (in units of m^{-1}). Note that the last step in Eq. (9.6) arises due to the fact that we have assumed a spatially homogenous medium, and that the VSF is symmetric in the azimuthal direction. If we take the VSF (angular distribution of scattered light) and normalize it by the scattering coefficient, b , we have

$$p(\theta) = \frac{\beta(\theta)}{b} \quad (9.7)$$

which is commonly referred to as the *phase function*. A warning to the reader, despite its name, the phase function has nothing to do with any phase. Rather, it borrows its name from astronomy, where it is used to describe how solar fluxes are reflected versus angle. Other popular names for the phase function include *scattering diagram* or *scattering indicatrix*.

One may ask what physical meaning the phase function would have, given that we’ve already described the distribution of scattered intensity with the VSF. While the exercise is left to the reader, integrating Eq. (9.7) from $\theta : 0 \rightarrow \pi$ results in unity, which is the requirement for a probability distribution. As we will see later, probability distributions are useful in analytical and numerical modeling techniques for simulating optical propagation underwater.

9.2.2.3 Phase functions of natural waters

In the 1960s and 1970s, a great deal of work was performed at Scripps Institute of Oceanography in San Diego, CA aimed at characterizing the visibility of

ocean water. While these studies were initially aimed at characterizing underwater photography, the timing of the experiments also coincided with the development and commercial availability of laser sources. The result of these studies was a wide body of data that provided great insight into how collimated laser light propagates in seawater.

One particularly important series of studies were measurements of various VSFs published by Theodore Petzold in 1972 [40]. While the exact shape of the VSF is dependent on geographical location and depth, Petzold's VSFs are still often cited as being largely representative of typical ocean waters.

Petzold's measurements are shown in Figure 9.6, along with the VSF of pure water. Note that the general shape of each VSF is extraordinarily similar in that they are highly peaked in the very near forward direction as compared to pure water. This suggests that for each scattering event, a photon is most likely to continue in a similar direction from which it came. It should be noted that the unique shape of ocean VSFs are due to the shape of the particle, as well as its size relative to the incident wavelength. For the interested reader, Bohren and Huffman [7] provide an excellent analytical treatment on the scattering properties of small particles.

Also shown in Figure 9.6 is the VSF of Maalox antacid, whose basic ingredients (aluminum hydroxide and magnesium hydroxide) were discovered to make an excellent laboratory replacement for real ocean particles due to its similar peaked shape at small angles [20]. Later, we will discuss the differences between different scattering agents

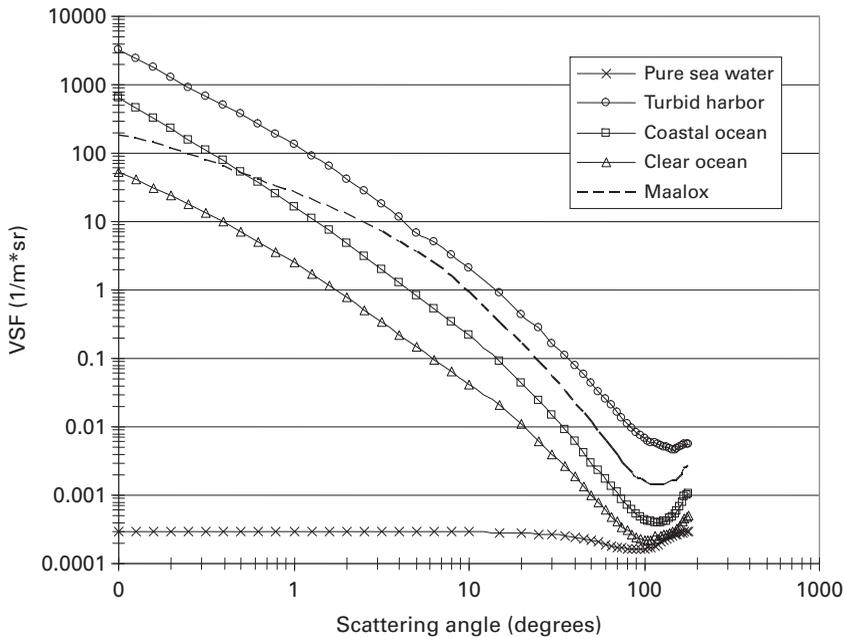


Figure 9.6 The VSF of various ocean water types as measured by [40].

Table 9.1 Representative absorption, scattering, and total attenuation coefficient values as measured by [40].

| Water Type | $a(/m)$ | $b(/m)$ | $c(/m)$ | w_0 |
|---------------|---------|---------|---------|-------|
| Clear ocean | 0.114 | 0.037 | 0.151 | 0.245 |
| Coastal ocean | 0.179 | 0.219 | 0.398 | 0.55 |
| Turbid harbor | 0.266 | 1.824 | 2.19 | 0.83 |

commonly used in laboratory experiments, and how these differences may influence the extrapolation of results to real world environments.

9.2.3 Scattering albedo

Scattering albedo is another useful metric derived from the scattering and absorption IOPs. Quite simply, it is the ratio of scattering to total attenuation, or,

$$\omega_0 = \frac{b}{c}. \quad (9.8)$$

In high albedo waters, a photon is more likely to be scattered. Naturally, in low albedo waters, absorption plays a larger role in overall attenuation.

Table 9.1 shows representative values of the scattering and total attenuation coefficient, as well as scattering albedo for a variety of ocean water types at 530 nm (as measured by [40]). The table illustrates the differences in absorption and scattering that could be expected in different water types. Naturally in clear open ocean waters, the overall attenuation coefficient is low. Also note that the scattering albedo is low in clear ocean waters, meaning that the majority of attenuation is due to absorption. In contrast, for turbid waters (for example, a harbor), the attenuation coefficient can be quite high. Additionally, turbid waters often have higher scattering albedos, meaning that the overall attenuation is dominated by scattering. We will see how the balance of absorption and scattering can affect the uFSO link in later sections.

9.2.4 Turbulence

Readers with experience in atmospheric optical links are more than familiar with the concept of turbulence. Turbulence is often caused by changes in air temperature, which result in changes of the refractive index, n , of the surrounding air. These refractive index changes are seen by the optical beam as it propagates through the air, causing it to bend and turn. The impact on an optical communications link can be significant, as turbulence causes the beam to wander at the receiver aperture, causing fluctuations and dropouts in the link. For schemes that encode information using the coherence of the optical signal, the effect of turbulence can be devastating.

Similar situations exist underwater, as the refractive index of seawater can change as functions of temperature, salinity, pressure, and wavelength. Austin and Halikas [4] (also from the Scripps Institute) compiled several studies in this area and developed

parameterized algorithms to determine the index of refraction over a wide range of conditions. In general, they found that n decreases with increasing temperature or wavelength, and n increases with increasing pressure or salinity. In general though, the value does not fluctuate more than a few percent, and a value of $n = 1.33$ is widely used, particularly for blue/green wavelengths. Furthermore, when in the presence of large particles (as is often the case in ocean waters), large particle scattering tends to dominate over fluctuations due to refractive index in the sea. This is especially true in short-range links (10s of meters) where significant temperature, pressure, or salinity changes are not to be expected.

In general, we will consider turbulence a negligible, secondary effect. However, turbulence may be an issue in open ocean waters where large particle scattering is at a minimum and link ranges could be expected to operate at 100 m or so. While there has been work in examining very small angle scattering and coherence properties of laser beams in ocean water, there has been little that examines the exact effect on the underwater communications channel specifically. Some researchers are taking note, as Simpson, Hughes, and Muth [46] looked at MIMO techniques for measuring fading due to turbulence and bubbles for uFSO links.

9.2.5 Apparent optical properties

The absorption and scattering coefficients, as well as the volume scattering function, are considered *inherent optical properties* (IOPs). This is to say, they are properties that depend only on the water, or the particles contained within them. *Apparent* optical properties (AOPs) are similar to IOPs, but also depend on the distribution of the light used to measure them. For example, AOPs are useful in determining the amount of ambient light undersea. In this case, the initial distribution of the sunlight on the surface, angle of arrival, atmospheric conditions, etc., are all dependent variables. One of the more common AOPs is the diffuse attenuation coefficient, K_D , which can be used to describe the exponential attenuation of the downwelling plane irradiance of the sun. The spectral dependency of K_D is largely dependent upon the IOPs of the water, which obviously varies between geographic locations. In an effort to bring order to the spectral dependency measurements of K_D in various locations of the world, Jerlov [26] proposed a classification system to group like bodies of water together. In Figure 9.7, the attenuation (in dB/m) for different Jerlov water types are shown as a function of wavelength. Note that in more turbid waters closer to land, the wavelength of minimal attenuation shifts from blue wavelengths to green. This is due to the increased concentration of particulate matter (organic and inorganic) in these regions. Because K_D is a measurement of diffuse sunlight, as opposed to a collimated laser source, the data presented in Figure 9.7 may be more indicative of a NLOS link rather than a point-to-point link. Regardless, the reader can gain an intuitive feel of the relative differences in loss between real water types.

Another example of an AOP is the *system* attenuation coefficient, defined here as k_{sys} , which depends not only on the water optical properties and source distribution, but also on the characteristics of the system used to make the measurement. To see the value

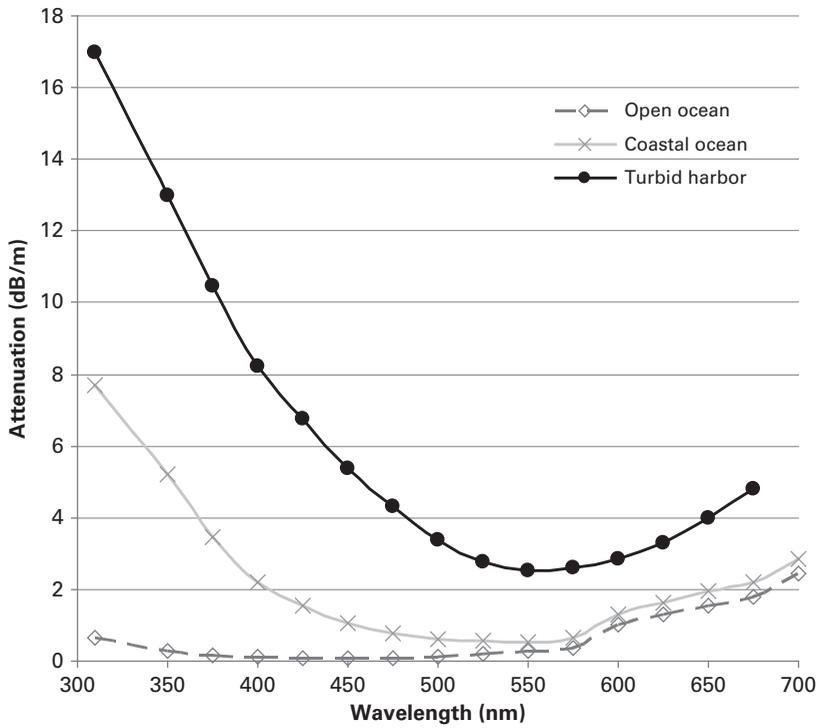


Figure 9.7 Approximate loss in dB/meter for various ocean waters as defined by Jerlov (data adapted from [26]).

of k_{sys} , let us consider the line-of-sight configuration. We have seen that using the total attenuation coefficient, $c = a + b$, in Beer's Law can accurately describe the line-of-sight scenario (i.e. Eq. (9.2)). This model for channel loss is predicated upon any photon that is absorbed or scattered does not contribute to the received optical power at the photoreceiver. However, now consider a scenario where there is significantly more scattering events. Here, multiple scattering causes photons that were once "lost" to scatter back into the receiver aperture and field of view. In this case, the photoreceiver collects more light than predicted by the attenuation coefficient. Additionally, we might consider the effect of increasing the receiver aperture and field of view. This would certainly increase the collection of scattered light.

The point is that modeling optical power loss by the attenuation coefficient, c , is only valid if we assume that *only* non-scattered light is collected. If scattered light is collected (due to either significantly turbid waters, a change in system parameters, or both), then another term must be used to describe attenuation. This "other" term is k_{sys} , and it provides us with the ability to modify Beer's Law as given in Eq. (9.2) to account for both *properties of the water* (multiple scattering, ratio of scattering to absorption, etc.) as well as the *properties of the system* (receiver aperture, field of view, source divergence, etc.).

9.2.6 Relating AOPs to IOPs

It should be clear now that it is important for the link designer to take into account *both* the environment as well as the system design when determining how to predict uFSO channel loss. Said another way, while the IOPs may be used to characterize the water and its constituents, it is an AOP that is often measured by the system. Understanding this relationship is crucial.

We now formalize the above discussion, following the notation of [10]. We begin by modifying Eq. (9.2) as,

$$P(z) = P_0 \exp(-k_{\text{sys}}z) \quad \text{where} \quad k_{\text{sys}} = a + (1 - \eta)b \quad (9.9)$$

where we have introduced a *scattering factor* $0 \leq \eta \leq 1$ that is used to account for the collection of multiply scattered light. Observe that a high scattering factor effectively scales down the scattering coefficient. Intuitively, this makes sense since the collection of scattered photons would have the net effect of reducing the attenuation due to scattering. A scattering factor of unity physically means that we have collected *all* light that has been scattered. In this case, the received power will be attenuated by the absorption coefficient. A scattering factor of 0 physically means that we have collected *none* of the scattered light. In this case, Eq. (9.9) reduces to Eq. (9.2), since $k_{\text{sys}} = c$. As we will see in later sections, neither Eq. (9.2) nor Eq. (9.9) can fully describe the attenuation for all ranges or water types. More often, we find that attenuation is exponential by c in low or minimal scattering regimes, and by k_{sys} in multiple scattering regimes.

9.3 Channel characterization: theory

In the previous section, we learned about some of the physical properties of ocean water, and gained insight into their potential impact to the uFSO link. We saw that while IOPs can be used to characterize the water, the exponential attenuation given by Beer's Law using the beam attenuation coefficient may only provide a first-order understanding of the ocean channel. In reality, we must consider physical subtleties like multiple scattering, or the balance between the degree of absorption and scattering, as well as system subtleties such as aperture, field of view, pointing and tracking, etc.

It should be clear at this point then, that in order to more accurately model uFSO links, we should simply find a way to express the spatial and temporal characteristics of a propagated laser beam at all points in space. Once this is known, one can accurately predict both the expected received intensity, as well as the maximum bandwidth available, at all points in the underwater channel.

On the surface, this seems like a simple task. After all, based on our previous discussions in the last section, we have the ability to accurately measure absorption, scattering, and phase functions of seawater. In reality however, arriving at a complete spatial and temporal description of the light field at all points in space is easier said than done. For example, while we may be able to describe the probability distribution of scattering versus angle of a single photon interacting with a single particle, how do we extrapolate this

to millions of photons interacting with millions of particles, altering their paths many times over the length of propagation?

The starting point for many theoretical descriptions is the Radiative Transfer Equation (RTE),

$$\left[\frac{1}{v} \frac{\partial}{\partial t} + \vec{n} \cdot \nabla \right] I(t, \vec{r}, \vec{n}) = -cI(t, \vec{r}, \vec{n}) + \int_{4\pi} \beta(\vec{r}, \vec{n}, \vec{n}') I(t, \vec{r}, \vec{n}') d\vec{n}' + E(t, \vec{r}, \vec{n}) \quad (9.10)$$

where v is the speed of light, c is the beam attenuation coefficient, t is time, \vec{n} is the direction vector, \vec{r} is the position vector, and ∇ is the divergence operator with respect to position \vec{r} . I is the radiance, β is the VSF, and E is the source radiance.

Simply stated, the RTE is a formulation in energy conservation: the change in intensity over a differential length (l.h.s.) is equal to the extinction losses (absorption and scattering) over that path, plus the contribution of any scattered intensity from other places in space (r.h.s.). The first term on the right, the one that describes extinction losses over a differential length, is simply Beer's Law. The second term on the right-hand side however, represents all the possible contributions of scattered intensity from all other points in space. This term is far less simple. In any case, we see the RTE is a complex integro-differential equation of several variables, and finding tractable solutions is no trivial task. In this section, we introduce some of the common theoretical methods used to describe the intensity distribution in space, as well as the temporal dispersion.

9.3.1 Numerical methods

Numerical methods, such as Monte Carlo simulations, are perhaps the most intuitive. Numerical methods essentially "mimic" the real world by individually tracking photons through the scattering medium. Armed with information such as the absorption and scattering coefficients, as well as the phase function, a Monte Carlo computer program "marches" photons through a medium, and "counts" those that arrive at a particular location. Naturally, these techniques can also account for source aperture and divergence, receiver aperture and field of view, and can also keep track of the polarization of individual photons. Numerical methods are most intuitive to understand, since we physically "keep track" of the pathways and histories of individual photons. Naturally, knowing the precise location of every photon allows us to determine the spatial distribution of light. Additionally, because we know the path each photon travels, we can calculate the time it takes each photon to arrive at a point in space. Comparing these time differences that arise due to different scattering paths allows us to calculate the temporal dispersion (i.e. pulse spread, or channel bandwidth).

Of course, numerical techniques are also the most computationally time consuming, as millions (or billions) of photons may be needed in order to accurately simulate a real-world application. Of course, this downside continues to become less of an issue as faster computers with more processing power become available.

While many Monte Carlo models describing photon transport in seawater have been reported (e.g. [29]), we wish to highlight a few new studies that were performed particularly with the uFSO channel in mind. Hanson and Radic [24] presented a Monte

Carlo model for predicting the spreading of an optical pulse in various water types with different FOVs. They found that at low frequency, increasing the FOV can improve the SNR of the received information signal, because scattered light does not adversely effect the low-frequency modulation. Conversely, at high frequencies, more temporal dispersion is observed. In this case, only non-scattered light maintains the temporally encoded information. Therefore widening the receiver FOV does not provide any benefit, other than perhaps increasing the DC component which contains no information. It is worth noting that this result applies to the LOS link. NLOS and retro links can be different. In general, the results show that 1 GHz rates in the LOS configuration are possible over moderate distances (<50m) in open ocean waters. Even in worst-case harbor waters, the 3 dB bandwidth was observed to be ~ 100 MHz, which may be quite sufficient for high-speed data transfer.

Dalgleish *et. al.* [17] also performed Monte Carlo modeling to investigate pulse spreading underwater. Monte Carlo results were compared to experimental tank measurements using a 500 ps (FWHM) laser source over a 12.5 m range. Results show that the FWHM increases from ~ 700 ps (~ 1.4 GHz) with $c = 0.4/\text{m}$ ($cz = 5$) to ~ 2 ns (500 MHz) at $c = 2.0/\text{m}$ ($cz = 25$). Also presented were model results showing the average number of scattering events versus attenuation length at various angular receiver positions away from the main beam (see Figure 9.8). When the receiver is on-axis, we see that on average, the received photons have undergone < 5 scattering events for $cz < 15$. After which, a transition occurs, and the number of scattering events shows a sharp rise with increasing attenuation length. We will later see similar transitions between “minimal scattering” and “multiple scattering” regimes in later sections.

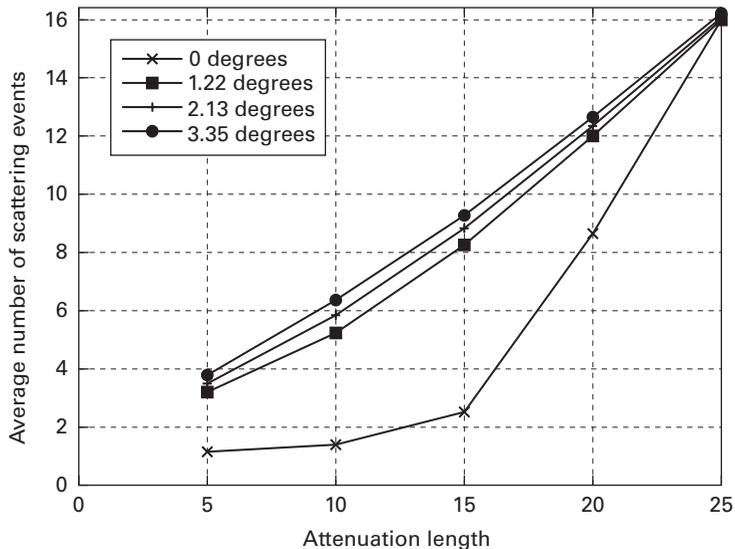


Figure 9.8 Monte Carlo results of the average number of scattering events versus attenuation length at various lateral positions from the main beam axis. Figure courtesy of [17].

For off-axis receiver positions, the average number of scattering events appears to grow somewhat linearly with attenuation length.

9.3.2 Analytical methods

Analytical methods are sometimes preferred over numerical techniques because they are more flexible, and are often less computationally intensive. The drawback to analytical methods are that they are often complex, less intuitive, and in some cases, do not yield tractable solutions without some simplifying assumptions. In this section, we discuss some of these methods.

9.3.2.1 Small-angle approximation

The small-angle approximation (SAA) is a popular simplification of the RTE. While the SAA has its origins in neutron scattering theory, its application to describing photon transport in ocean waters was made popular by a series of papers by H. G. Wells [51–52]. As its name suggests, the SAA assumes that scattering events are confined to near forward angles to the beam axis. In this case, we can assume that $\cos(\theta) = 1$ and $\sin(\theta) = \tan(\theta) = \theta$, where θ is the angle off axis. This assumption is only valid when the VSF is peaked at forward angles. Fortunately, we have seen that this is precisely the case for ocean waters. Also note that this effectively assumes that scattered photons travel the same distance as non-scattered photons. This inherently means that the SAA ignores time-dispersion.

Wells derives the beam-spread function (BSF) as $f(\theta, z)$, where θ is the angle off axis and z is the range. Wells showed the convenience of representing $f(\theta, z)$ as its 2D Fourier transform $F(\psi, z)$, where

$$f(\theta, z) = 2\pi \int_0^{\infty} F(\psi, z) J_0(2\pi\theta\psi) \psi d\psi \quad (9.11)$$

$$F(\psi, z) = 2\pi \int_0^{\theta_{\max}} f(\theta, z) J_0(2\pi\theta\psi) \theta d\theta$$

where J_0 is a zeroth-order Bessel function and θ_{\max} is the maximum angle of scattering as defined by the SAA. Equation (9.11) represent a Hankel transform, which simplifies the 2D Fourier transform by taking advantage of radial symmetry. Wells derived the BSF in the spatial frequency domain as,

$$F(\psi, z) = \exp \left[-cz + \int_0^z S \left(\psi \frac{r}{z} \right) dr \right] \quad (9.12)$$

where $S(\psi)$ is the Hankel transform of the VSF, $\beta(\theta)$. Knowledge of the VSF allows for transform to $S(\psi)$. Integration of $S(\psi)$ in Eq. (9.12) provides $F(\psi, z)$, and transform of $F(\psi, z)$ in Eq. (9.11) provides $f(\theta, z)$, the desired BSF.

Wells suggested a cone of $\theta_{\max} = 10^\circ$ for the small-angle limit, though the exact θ_{\max} can vary depending on the shape of the VSF and geometry to which the SAA is applied. Zege *et al.* [54] showed that the SAA provides good results in regions where $x/z < 1$, where x is the lateral distance from the beam axis and z is range. This limit on the region of applicability is to be expected given our assumption that scattering events are confined to near forward angles.

9.3.2.2 Vector radiative transfer modeling

Jaruwatanadilok [25] approached the problem of predicting temporal dispersion by using vector radiative transfer theory. This theory takes into account both the spatial and temporal dispersion, as well as modeling how the polarization changes during propagation. The impulse response for both polarization components was modeled in water with $c = 0.305/\text{m}$ at 30 m and 50 m. Results show that at both ranges, the co-polarized component is significantly more peaked than the cross-polarized component. This should make intuitive sense, as the co-polarized component is that which should not have undergone multiple scattering, and hence retains a high temporal resolution. On the other hand, multiple scattering would stand to alter the orientation of the polarization vector. As such, the cross-polarized component should contain mostly light that has been multiply scattered, and this multiple scattering will result in a spread of the impulse response. The effect of FOV on BER was also studied. Results supported the findings of Hanson and Radic [24], as they showed a diminishing improvement in BER when increasing FOV at high data rates. It was observed however that at short range, one does stand to benefit from increasing the FOV, even at high frequency. This is because at short range, there is less multiple scattering. However, as range increases, so does the degree of multiple scattering. In this regime, the benefit of a wider FOV decreases.

9.3.2.3 Other analytical techniques

We have previously seen that a common simplification in analytical models is to apply the small-angle approximation, and assume that all scattering events occur within very near forward angles. This effectively ignores any time dispersion, since we essentially presume that all photons travel approximately the same path length through the medium. Of course, time dispersion is ignored only due to the mathematical convenience of the first-order assumption, $\cos(\theta) = 1$ and $\sin(\theta) = \tan(\theta) = \theta$. However, by considering higher-order effects, temporal dispersion can be included. To do this, one attempts to describe the spatial dispersion and variation in photon transit time statistically.

Approaches may vary. Stotts [48] presents a simple statistical approach that applies the scattering statistics of electron scattering theory to that of photons undergoing scattering by large particles. In this way, he was able to calculate the average path length difference a scattered photon has relative to a non-scattered photon. Using the speed of light, the average pulse spreading can be given by,

$$\Delta t = \frac{R}{v} \left\{ \frac{0.30}{bR\theta^2} \left[\left(1 + \frac{9}{4}bR\theta^2 \right)^{1.5} - 1 \right] - 1 \right\}. \quad (9.13)$$

Here, R is the link range, v is the speed of light, b is the scattering coefficient, and θ is the mean scattering angle given by the scattering phase function. McLean, Crawford, and Hindman [31], compared Eq. (9.13) to a Monte Carlo model, and found fairly good agreement out to 16 scattering lengths (i.e. $bz = 16$). The Monte Carlo calculations integrated a 10-m disk centered on the beam axis, and integrated over the entire plane of interest.

More accurate statistical investigations are found in [30]. Other authors have also arrived at statistical expressions for mean multipath time, displacement, and direction. The interested reader is referred to [32] which presents a sufficient review of some of the work in this area. For now, we simply note that the applicability and accuracy of a certain method is largely dependent upon the order of scattering considered, as well as the distributions chosen to represent the scattering process.

9.4 Experimental research in wireless optical communications underwater

Now that an appreciation for the complexities of modeling and characterizing the underwater channel has been gained, we shift focus to present a growing body of work aimed at experimentally characterizing the channel in both the spatial and temporal domain, as well as efforts in validating the theoretical models presented in the previous section.

9.4.1 Spatial dispersion

In Section 9.3, we were introduced to some of the theoretical techniques used to predict spatial dispersion. In particular, we examined the small-angle approximation that is often used to aid the calculation of the beam spread function (BSF). Cochenour, Mullen, and Laux [10] experimentally validated a formulation of the SAA given by Zege [53]. The BSF can be measured experimentally in a large test tank, by scanning a photodetector laterally to the beam axis after the beam has propagated some distance. In this way, the intensity versus lateral distance is obtained, and can be compared to analytical results from the model. Results from this particular experiment are shown in Figure 9.9 for various water types. In particular, we note that in clear waters, the intensity falls off rapidly with lateral distance from the beam. This makes sense, since under little or no scattering, the laser source maintains much of its directionality. From a practical point of view, this tells the link designer that in clear waters, pointing and tracking between the transmitter and receiver are essential. It is interesting to consider that in this case, while link range and bandwidth may be easy to achieve in the absence of scattering (i.e. no temporal dispersion), the overall system design may be harder due to the pointing and tracking requirements.

As turbidity increases, light is scattered into near forward angles, and the intensity distribution spreads out away from the beam axis. As such, the power penalty associated with off-axis viewing decreases as turbidity increases. In this sense, we might even say scattering “helps” the link by making it less prone to pointing errors. The reader should bear in mind however that the overall signal level will be lower due to

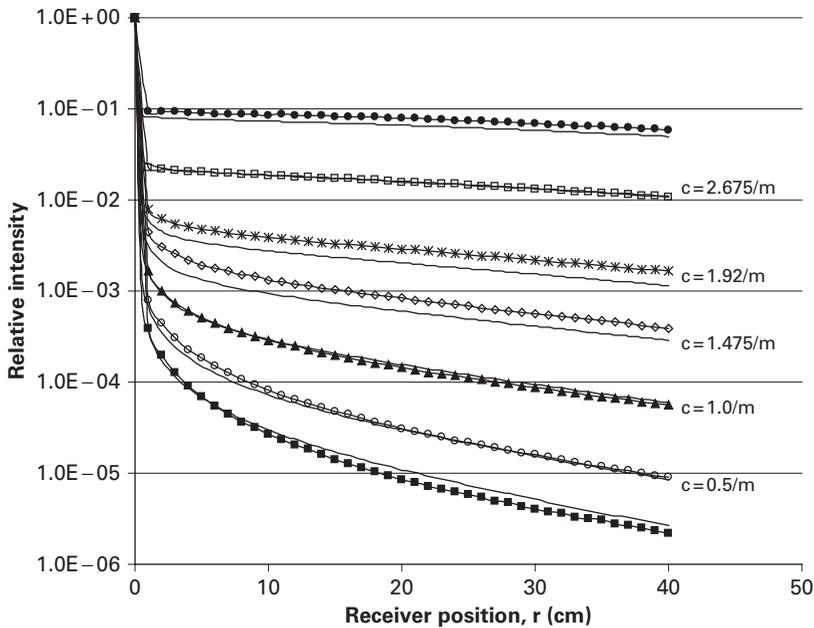


Figure 9.9 Comparison of the beam spread function (BSF) between experiment and small-angle scattering theory for various water types. From [10].

increased losses due to scattering and absorption at the higher turbidities. This is not reflected in Figure 9.9 due to the on-axis normalization for each BSF. Furthermore, we have not mentioned anything about how the temporal signal is affected by off-axis viewing.

Recall our discussion from Section 9.2, where we discussed the differences between IOPs and AOPs. We learned that while IOPs (e.g. the beam attenuation coefficient) are useful in defining the properties of the water, AOPs (e.g. the system attenuation coefficient) are useful in incorporating the effects of the measurement system, as well as distribution of the light field itself (like when a collimated laser beam undergoes multiple scattering).

With this in mind we refer to Figure 9.10, also from [10], who used a SAA model to investigate the rate of attenuation versus range for different lateral displacements of the receiver from the main beam axis. (A summary of the modeled “link” is shown in Table 9.2. The reader is referred to the reference for more detail.) Figure 9.10 shows the received intensity as a function of range for various lateral positions from the beam for $c = 0.275/\text{m}$. When the receiver is accurately aligned with the transmit beam ($r = 0$), the received power attenuates exponentially according to the beam attenuation coefficient, c , out to $z = 75 \text{ m}$ ($cz = 20$), at which point a transition occurs, and the received signal attenuates according to $k_{\text{sys}} = a + 0.16b$ ($\eta = 0.84$). As predicted, under enough scattering events, multiply scattered photons make their way back into the receiver aperture and field of view, and contribute positively to the overall received

Table 9.2 Modeled link parameters of the data shown in Figure 9.11, from [10].

| Parameter | Value |
|--|-------------------|
| Laser wavelength | 532nm |
| Initial power | 3W |
| Receiver aperture | 8mm |
| Receiver gain | 5×10^5 |
| Anode responsivity | 3.9×10^4 |
| Q.E. | 18% |
| Dark current (minimum detectable signal level) | 20nA |

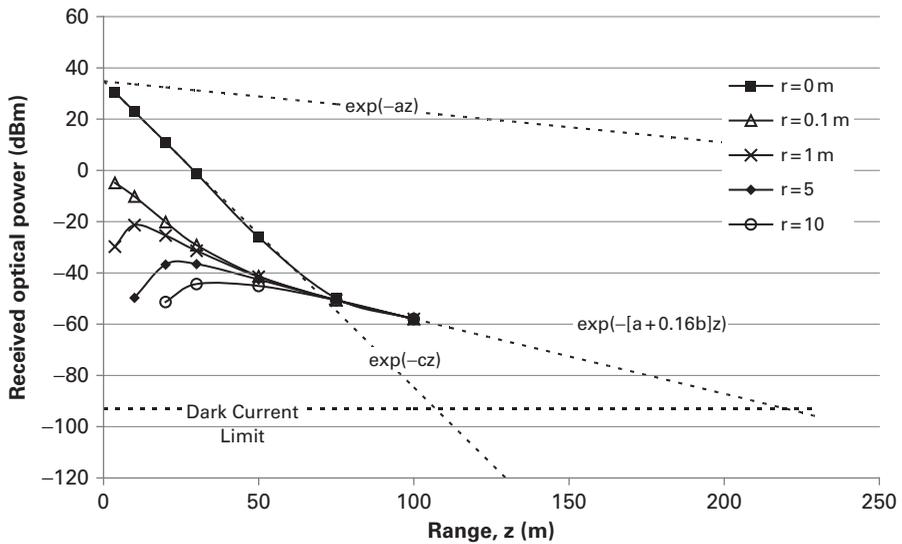


Figure 9.10 The received intensity as a function of range for $c = 0.275/\text{m}$. Data is shown for different lateral receiver positions, r m from the beam axis. From [10].

signal. In some sense, the collection of scattered light can effectively extend the maximum range of operation. Had we only considered non-scattered light, we would in this case underestimate the maximum range of operation by a factor of 2 (or about 100 m).

Another interesting conclusion comes when considering lateral receiver positions away from the main beam axis ($r > 0$). At short range, the penalty for pointing mis-match is significant. This is intuitively plausible since the water is not extremely turbid ($c = 0.275/\text{m}$) and the beam has not propagated far enough to have undergone enough multiple scattering events that would significantly spread the beam out in space. However, as the link range increases, the power penalty for relaxed pointing decreases relative to on-axis viewing. In fact, after the transition point from $\exp(-cz)$ to $\exp(-k_{yy}z)$, there is virtually no penalty. This is because at longer ranges, the beam has undergone so much spatial dispersion and divergence, that it approaches a uniform distribution in space, and changes in receiver location become irrelevant.

9.4.1.1 Characterizing k_{sys} and the c - k_{sys} transition

The reader may be wondering if it is possible that, given some properties of the water, one may predict at what range the attenuation will transition between c and k_{sys} , as well as what value k_{sys} will take on. In essence, we would ideally like to parameterize k_{sys} based upon all of the variables that influence the rate of attenuation. These would include *both* the water optical properties (absorption and scattering coefficients, VSF, etc.) as well as system parameters (receiver aperture, FOV, beam radius, divergence, etc.). This parameterization however is no trivial task due to the number of variables involved [13]. However, some experimental work has gone into examining the impact of a subset of variables. Cochenour, Mullen, and Muth [12] experimentally showed the dependence of albedo and on overall attenuation versus attenuation length, as well as polarization. In this light, it may be more convenient to represent Eq. (9.9) in terms of the scattering albedo. Rearranging terms results in,

$$P(cz) = P_0 \exp(-[1 - \eta\omega_0]cz). \quad (9.14)$$

Here we see that the natural logarithm of the received power versus attenuation length decays with a slope $m = 1 - \eta\omega_0$. When the scattering factor $\eta = 0$ (no scattered light collected), Eq. (9.14) reduces to Eq. (9.2). As $\eta \rightarrow 1$, the received signal is dominated by multiply scattered light, and reduces to,

$$P(cz) = P_0 \exp(-[1 - \omega_0]cz). \quad (9.15)$$

Because $\omega_0 = b/c \leq 1$, the received power in Eq. (9.15) decays slower versus attenuation length than the non-scattered regime in Eq. (9.14). Intuitively, this makes sense since the collection of multiply scattered light in effect increases the optical power seen at the receiver. Finally, note that in the single scattering regime ($\eta = 0$) P_0 is equal to the initial transmitted power. However in the multiple scattering regime ($\eta \rightarrow 1$), P_0 is approximately equal to the received power immediately before transition.

Equations (9.14) and (9.15) describe the rate of attenuation in the single scattering and multiple scattering regimes. We now ask whether or not we can predict at what attenuation length this transition might occur. To do this, we borrow analysis from Lerner and Summers [29], who via Monte Carlo model, found that light remains relatively confined to the original axis of propagation for distances less than a diffusion length, where a diffusion length is defined as,

$$L_D = \frac{L_{\text{sca}}}{1 - \langle \cos \theta \rangle}. \quad (9.16)$$

The scattering length is given as $L_{\text{sca}} = b^{-1}$, and the term $\langle \cos \theta \rangle$ is the average cosine of the scattering angle,

$$\langle \cos \theta \rangle = 0.5 \int_0^\pi \rho(\theta) \cos \theta \sin \theta d\theta, \quad (9.17)$$

where $\rho(\theta)$ is the scattering phase function. When the VSF is highly peaked, as is the case for ocean waters (see e.g. Figure 9.7), then the average cosine is high – typically 0.9

or greater. Expressing the diffusion length in terms of attenuation lengths, we find the number of attenuation lengths that the light must travel before it deviates significantly from the main axis due to multiple scattering can be estimated as

$$cz > \frac{10}{\omega_0}. \tag{9.18}$$

We have now developed, through Eqs. (9.16)–(9.18), a way to parameterize the attenuation of multiply scattered light as a function of scattering albedo.

Experimental results of the normalized received power versus attenuation length for three different scattering albedos at a range of 3.66m are shown in Figure 9.11, along with the curve for $\exp(-cz)$ (dotted line). Note that in low scattering albedo waters ($\omega_0 = 0.55$, $\eta = 0.22$), the received power attenuates according to the attenuation coefficient, c , for a wide range of attenuation lengths. In contrast though, for higher albedo waters ($\omega_0 = 0.83$, $\eta = 0.88$ and $\omega_0 = 0.95$, $\eta = 0.95$), we see a similar trend that was previously observed in Figure 9.10, where the received power attenuates according to c for low attenuation lengths ($cz < 10$), then at a slower rate of k_{sys} for higher attenuation lengths. Computing the diffusion length in Eq. 9.18 for each albedo results in $L_D = 10.5$, 12, and 18 for $\omega_0 = 0.95$, 0.83, and 0.55 respectively. The calculated transition points are shown as dashed lines in Figure 9.11. Good agreement is shown between the predicted and actual transition values.

How may we explain the change in slope seen in high albedo waters? For low albedo, the scattering factor, η , is also low. Indeed, in waters that contain few scattering particles, attenuation is dominated by absorption. This means that any light that is scattered has

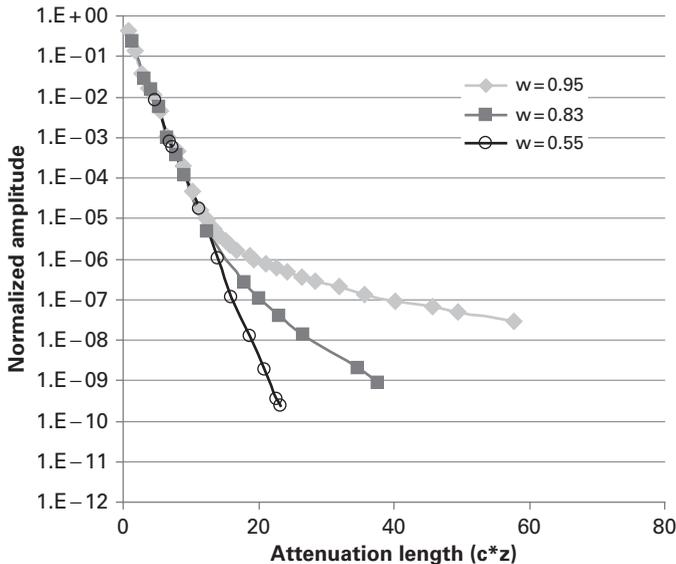


Figure 9.11 The attenuation of light with the receiver on-axis for different values of scattering albedo. The transition between the minimal scattering and multiple scattering regime are indicated by dashed lines. Adapted from [12].

a higher probability of being absorbed, and therefore the chance of multiply scattered light scattering back into the receiver aperture and FOV is low. As such, we expect attenuation by $\exp(-cz)$, even at higher turbidities, because there is little or no contribution of multiply scattered light. Conversely, in high albedo waters (such as a turbid harbor), an increasing η was observed. In this scenario, multiple scattering is quite likely, and the contribution of multiply scattered light at the receiver causes the received power to attenuate more slowly. We see then that albedo effectively “regulates” the degree of multiple scattering. In other words, it is important for the uFSO link designer to consider not only *what* the absorption and scattering coefficients are (in order to determine the attenuation coefficient, $c = a + b$), but also what the *relative* contributions of absorption and scattering are ($\omega_0 = b/c$).

We have seen in the previous two sections how in some sense, multiple scattering and spatial dispersion may actually *help* maintain the uFSO link under a wider range of environments and scenarios. This point has not been lost by other researchers, who have looked at comparing the performance of different types of links under the same environmental scenarios (see for example, [3]).

In summary, we have presented some experimental work that has attempted to validate spatial dispersion models, and examined the impact of multiple scattering on the rate of attenuation with range and/or attenuation length. We find that in cases of multiple scattering, collecting scattered light can be useful, however only if it has not scattered so much that the information encoded on it begins to degrade. In the next section, we finally address this point, and present current work aimed at characterizing the temporal dispersion seen underwater.

9.4.2 Temporal dispersion

9.4.2.1 Methods for measuring temporal dispersion

We previously mentioned that one approach for characterizing the time-dependent properties of light propagation in water is to measure the temporal spread of a short pulse after it travels through water [17]. This information could then be converted to a frequency response via the Fourier transform. The disadvantage of this approach is that it requires a large receiver bandwidth and a correspondingly short integration time, which limits the detection sensitivity. Another option is to use a continuous wave source that is intensity modulated at a range of frequencies and measure the change in modulation depth at each frequency after the signal propagates through water. This technique has the advantage of very narrow receiver bandwidths and long integration times, which drastically improves the detection sensitivity.

The challenge of the intensity modulation approach lies in identifying components required for high time resolution measurements with sufficient sensitivity to overcome the exponential loss of light in water. On the transmitter end, a high-power source is needed in the blue-green wavelength region that can be modulated at rates exceeding 100 MHz. Although electro-optic modulators have been used in the past to externally modulate high-power continuous wave green sources ([35], [8], [9]), they are limited to frequencies < 100 MHz due to the capacitance of the modulator crystal. Blue laser

diodes are beginning to appear commercially with bandwidths of a few hundred MHz, albeit a lower output power (10s mW).

Hanson and Radic [24] reported on a system using an infrared laser diode that was externally modulated by a waveguide electro-optic modulator and amplified by a fiber amplifier before being frequency doubled to 532 nm. Although this source is capable of generating modulated green light at rates up to 20 GHz, the optical power was limited to less than 10 mW. Still, 1 GHz transmission was observed over a 2 m water pipe that simulated coastal ocean waters with 36 dB of attenuation. It was unclear however how the confinement of the narrow water pipe affected multiple scattering dynamics.

On the receiver end, there is a trade-off between aperture and bandwidth. A large active area is desired to optimize the collection of scattered light, but large active areas usually translate into poor time response due to large RC time constants. Although photomultiplier tubes (PMTs) are frequently used due to their high gain and large area, the fastest PMTs are limited to approximately 350 MHz with an 8 mm active area. Avalanche photodiodes (APDs) are another good choice due to their high speed and gain, but the active area is limited to <1 mm at speeds approaching 1 GHz.

An alternative approach has been presented by Mullen *et. al* [37, 38], which is set up capable of measuring the transmission of light modulated at frequencies up to 1 GHz in simulated ocean water up to 55 attenuation lengths. The setup is based on the use of a mode-locked laser and a custom, modified PMT for high-speed applications (the bandwidth of this device (-3 dB) is approximately 1 GHz.). The use of a mode-locked laser allows us to take advantage of the frequency content inherent to a train of mode-locked pulses. These frequency components are also generated simultaneously, so there is no need to vary the modulation frequency of the source. Furthermore, since mode-locked lasers are capable of high average optical powers, this source is an ideal candidate for use in measuring the frequency response of light propagating in water. The frequency content of the received mode-locked signal is examined on a spectrum analyzer. The average DC component is measured with a multimeter. The AC and DC components can then be used to compute a modulation depth (MD), which is equivalent to measuring a frequency response of the underwater channel.

9.4.2.2 Effect of VSF on temporal dispersion

The characteristics (size, shape, composition) of the scattering particles can have significant impact on the way light propagates. Mullen *et. al.* [38] examined this issue by looking at the differences between a variety of experimental scattering agents used to simulate oceanic particles in the laboratory environment, and considered how the VSF of each influences the temporal dispersion. The scattering agents are: Maalox, magnesium hydroxide ($Mg(OH)_2$) and aluminum hydroxide ($Al(OH)_3$), as well as a course grade of Arizona Test Dust (ATD).

Graphs showing the MD (i.e. frequency response) as a function of attenuation length are included in Figure 9.12 for three different modulation frequencies (0.1 GHz, 0.5 GHz, and 1.0 GHz). It is evident that for $cz < 10$, the results for all five scattering agents are the same: the MD remains constant for all three modulation frequencies. However, for $cz > 10$, the MD began to decrease with increasing attenuation length.

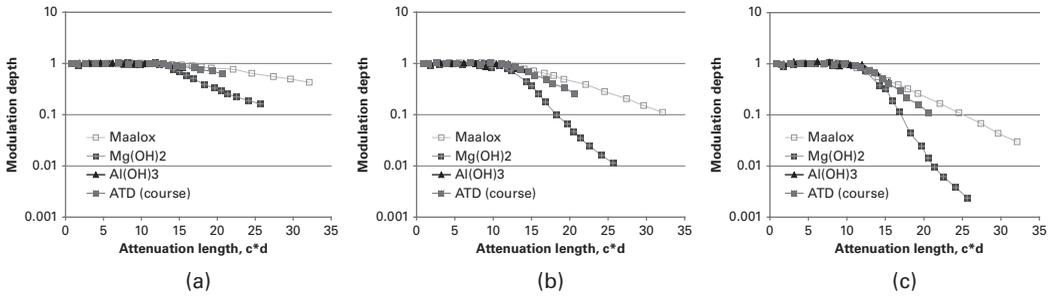


Figure 9.12 Modulation depth of various scattering agents versus attenuation length for (a)100MHz, (b)500MHz, and (c)1GHz. From [38].

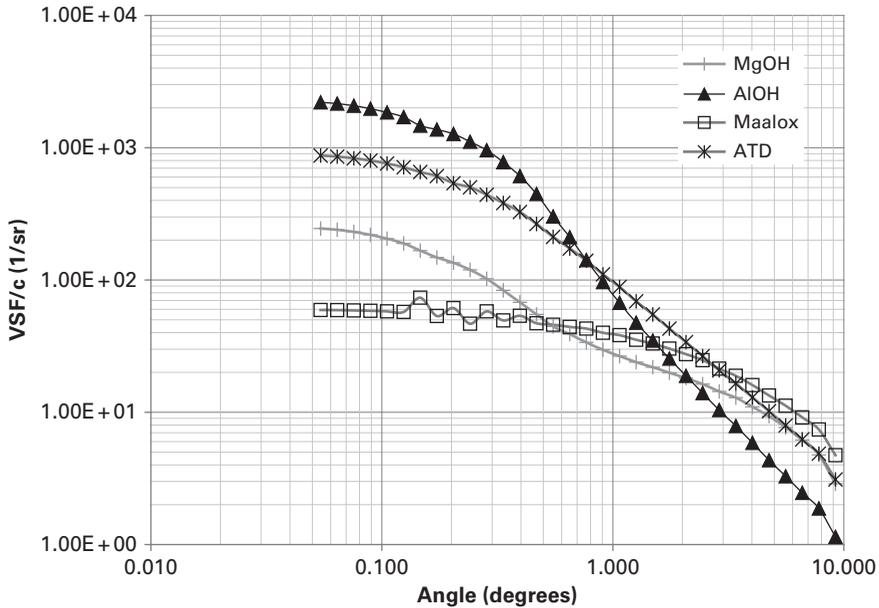


Figure 9.13 VSF of various laboratory scattering agents used to mimic ocean water. From [38].

For all of the scattering agents, the slope of MD vs. attenuation length increased as the modulation frequency increased. However, at each modulation frequency, there were differences between the slopes for each of the scattering agents. The slope was the largest for the $\text{Mg}(\text{OH})_2$ data and the smallest for the Maalox data, with the slopes for the other scattering agents lying somewhere between these two extremes.

We attribute these differences to the VSF. The VSF of each scattering agent is shown in Figure 9.13 where the VSF data has been normalized to the beam attenuation coefficient measured for each data set. The VSF for $\text{Al}(\text{OH})_3$ is the most forward-peaked, followed closely behind by the VSF for the coarse ATD particles. The VSF for $\text{Mg}(\text{OH})_2$ has a flatter response vs. scattering angle for angles greater than one degree. Finally, the

VSF for Maalox follows that of $\text{Mg}(\text{OH})_2$ except that it lacks the small peak observed in the VSF of $\text{Mg}(\text{OH})_2$ at angles less than one degree.

The anticipated correlation between the shape of the VSF at small angles and the MD of forward-scattered light is that the more peaked the VSF, the lower the loss in MD for a given number of attenuation lengths. The reasoning behind this is that the sharper the peak of the VSF at small angles, the lower the probability of large-angle scattering that produces multipath time delays at the receiver. This hypothesis appears to explain the trends for all scattering agents except Maalox. Although the Maalox VSF has the flattest response vs. scattering angle, the MD of the Maalox data is the highest for a given number of attenuation lengths. It is not completely understood why this is, however it may be due to particle coagulation present at high Maalox concentrations in the experiment, that are not present in the VSF measurements (which were made in single scattering conditions).

9.4.2.3 Effect of scattering albedo on temporal dispersion

We have already seen how the scattering albedo impacts the rate of attenuation, and we expect similar influences on the temporal dispersion. Mullen also conducted to evaluate the effect of single scattering albedo on the MD of the detected signal. Graphs showing the MD as a function of attenuation length for two different single scattering albedos – $\omega_0 = 0.95$ and $\omega_0 = 0.70$ – for Maalox is shown in Figure 9.14. Note that the MD for the smaller scattering albedo data (Figure 9.14(b)) remains constant until approximately $cd = 15$ attenuation lengths and then begins to decrease. This is a distinct difference from data with $\omega_0 = 0.95$ (Figure 9.14(a)) where the MD began to decrease for $cz > 10$ attenuation lengths.

We previously saw that Eq. (9.18) accurately predicted the transition between the minimal scattering and multiple scattering regimes for the non-time dependent measurements of attenuation. We may ask ourselves if the same holds true for predicting the transition in the “modulation loss” portion of the frequency response curves in

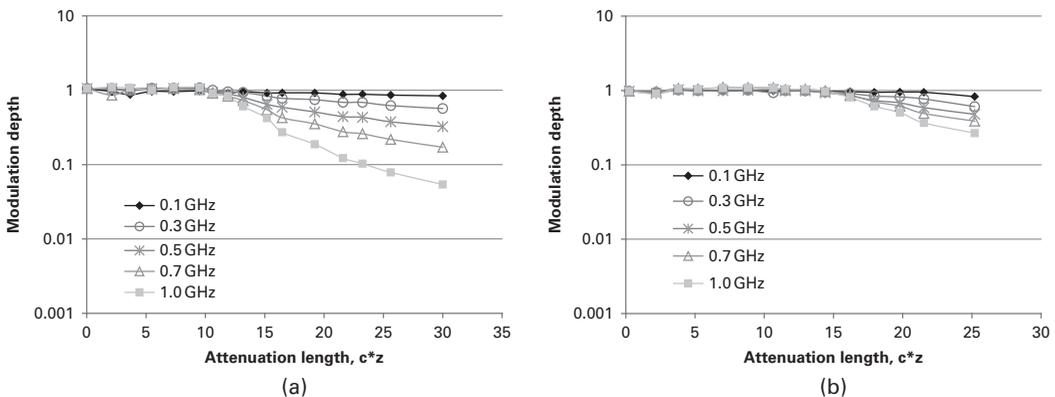


Figure 9.14 Modulation depth versus attenuation length for two different scattering albedos (a) $\omega_0 = 0.95$ and (b) $\omega_0 = 0.70$. Data shown for Maalox scattering agent. From [34].

Figure 9.14. Indeed it does, as calculations show that for $\omega_0 = 0.95$, $L_D = 10.5$ attenuation lengths, and for $\omega_0 = 0.70$, $L_D = 14.3$ attenuation lengths, both of which agree with the transition seen in the data.

9.4.2.4 Impact on link bandwidth

In order to put the preceding discussions in perspective, the frequency response for two different scattering agents is shown in Figure 9.15 for two different albedos at $cd = 20$ attenuation lengths. The 3 dB bandwidth is indicated by the frequency at which the MD is reduced by half. Here it is evident that the bandwidth of a link operating in Maalox-enhanced water increases by nearly a factor of 2 (from 0.5 GHz to 0.9 GHz) when the single scattering albedo is decreased from $\omega_0 = 0.95$ to $\omega_0 = 0.70$. A similar trend is seen for the $\text{Mg}(\text{OH})_2$ data – the link bandwidth increases from 0.3 GHz to 0.5 GHz as the scattering albedo decreases. These results illustrate the importance of evaluating the effect of multiple variables – such as particle type and single scattering albedo – when estimating the bandwidth capacity of an underwater optical link.

In summary, the results presented by [37, 38] illustrate the fact that temporal dispersion is very sensitive to changes in the underwater environment. Specifically, changes in the water optical properties due to variations in both scattering (particle type) and absorption (scattering albedo) were found to significantly influence the frequency response. Altering the type of scattering agent did not affect the number of attenuation lengths that the optical signal propagated before the MD decreased, though changing the scattering albedo does. However, the improved frequency response seen in high absorption waters also reduces the overall signal level at the receiver, limiting physical link range. Fortunately, lower albedo waters are also those that tend to have smaller beam attenuation coefficients. Therefore, modulated light transmitted in “clear ocean” water

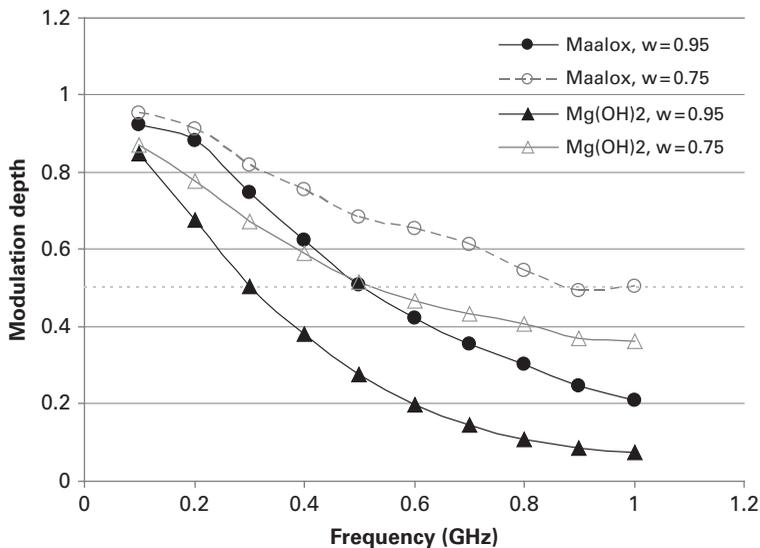


Figure 9.15 The frequency response for Maalox and $\text{Mg}(\text{OH})_2$ with different albedos. From [38].

($\omega_0 = 0.247$, $c = 0.151$ m) can propagate a longer physical range than in “turbid harbor” water ($\omega_0 = 0.833$, 2.19 m) for the same number of attenuation lengths before the modulation depth is reduced.

Once the transition between $MD = 1$ and $MD < 1$ is reached, the type of scattering particle appears to have the largest effect on the propagation of modulated light in water. Specifically, the decay rate of the MD with increasing attenuation length is dependent on the VSF of the scattering particle. To date, there have been no experimental comparisons of the frequency response between various ocean water types (for example, Jerlov water types). Such an investigation would be largely beneficial.

9.5 System design for uFSO links

As the physics of the underwater channel become more clearly understood, researchers are beginning to look at “system-level” issues and testing uFSO links in real ocean environments. Now that we have discussed some of the theoretical and experimental efforts underway to describe the ocean channel, we shift focus in this section to discuss some of the system trade-offs that should be considered by the uFSO link designer, as well as review some related applications by researchers in this area.

9.5.1 Transceiver components

Due to the near ubiquitous presence of fiber optics in today’s world, transceiver components in the IR are relatively mature. Atmospheric FSO links have taken advantage that a certain degree of technological maturity already exists for the wavelengths of interest. However, this cannot always be said for the blue/green portion of the spectrum where uFSO links will operate.

A comparison of transmitter technology is summarized in Table 9.3. While solid-state lasers (often frequency doubled Nd:YAG) offers the highest output powers, LED ([45], [50], [18], [19]) and laser diode sources ([14], [28], [6]) in the blue/green are also gaining popularity due to their small size and flexibility compared to the large size of a solid-state laser. It should be noted that fiber and fiber amplified lasers represent an exciting new area in this field, as they have the potential to provide high powers, flexible modulation schemes, and low size and weight requirements ([41]). LED and laser diode sources are best suited for small platforms like underwater sensor nodes, where size and power consumption must be minimized. Solid-state and some fiber lasers are more suited for large platforms (such as a vehicle) where size and efficiency are less of a concern and can be traded off for higher laser output powers. Indeed, the “holy grail” of transmitters is a small, high-efficiency, high-power source that would allow for closing links at long ranges or high turbidities. However, until such a source arrives, it is necessary to make trade-offs depending on the particular platform and/or application.

On the receiver end, PIN-photodiodes may be attractive due to their fast response (>1 GHz) and generally low cost. However photomultiplier tubes (PMTs) and avalanche

Table 9.3 Comparison of laser transmitter sources.

| Source | Solid-state laser | LED | Laser diode |
|--------|--|--|---|
| Pros | <ul style="list-style-type: none"> • High power (> 1W) • Directionality • Narrow spectral BW • Best for long-range applications | <ul style="list-style-type: none"> • Small, flexible, cheap • Easy to array (for high power) • 1Mbps easily achieved • Best for short/medium range links | <ul style="list-style-type: none"> • Small, flexible • Can be arrayed • Fast modulation (100's MHz depending on wavelength and power) • Best for short/medium range links |
| Cons | <ul style="list-style-type: none"> • Size, weight, and power • Cost • Possible challenges in blue/green wavelengths | <ul style="list-style-type: none"> • Wide optical bandwidth • Low degree of directionality • Low Power (unless arrayed) | <ul style="list-style-type: none"> • Technical maturity • Low power (unless arrayed) |

photodiodes (APDs) are popular due to their larger apertures and higher gains. This is typically desirable given the large amount of attenuation undergone by the optical signal. PMTs offer the highest gain ($>10^5$) as well as low noise. However, their bandwidths are typically limited to <350 MHz, and custom designs are needed to approach 1 GHz. Additionally, there is a trade-off between detector aperture and bandwidth. Photodetector bandwidths can be improved by reducing the active area (and thereby the device capacitance), however at the cost of collecting less photons. Large-aperture (>25 mm), high-speed (>1 GHz), high-gain ($>10^3$) photodetectors are highly desired, albeit elusive.

One must also consider detector FOV. While a large FOV may be desirable for collecting as many scattered photons as possible, it may also degrade the quality of the modulated information signal, depending on the frequency of modulation and turbidity of the water. In general, the choice of photodetector (type, aperture size, FOV, etc.) is largely dependent on the application (short range vs. long range, clear waters vs. turbid waters, etc.), and the system engineer should take into consideration all of these variables when choosing an appropriate photodetector.

Also, on the receiver end it may be useful to use an optical filter, particularly in environments where solar ambient light is expected to be significant. The effect of solar ambient is two-fold. First, it can dominate the precious dynamic range available from the photoreceiver. We wish to reserve as much dynamic range as possible for detecting a small amplitude laser communication signal. In addition, solar ambient light may contribute to shot noise in the receiver, meaning that noise is created by an optical signal that is not the one we wish to detect. Furthermore, if the solar ambient light level fluctuates over time, it may create additional sources of noise that interfere with the desired communications signal. Using an optical filter centered at the transmitter wavelength can reduce the contribution of solar ambient light. The bandwidth of the filter, and hence its ability to reduce solar ambient light, is largely driven by the characteristics of the laser

source. For solid state (and most fiber) lasers, the bandwidth of the optical carrier can often be <1 nm. For LED sources, the bandwidth may be 10s of nm. Narrow optical filter bandwidths may reduce the field of view (depending on the implementation) and may be expensive depending upon the aperture size. These are all trade-offs that must be considered.

9.5.2 Modulation and coding schemes

A related issue to our discussion on laser sources and receivers is the choice of modulation scheme. In general, there are two options. One is to use a pulsed laser (typically solid-state, or fiber-amplified), and to implement a modulation scheme like pulse position modulation (PPM). This is a popular choice for long-range links, or links that traverse the air–water interface (for instance, from an aircraft to a submarine). In this scenario, we are interested in putting as much peak power into the water as possible in order to combat attenuations. As such, pulsed lasers are an ideal source for this. There is a bandwidth trade-off however, as in order to achieve higher peak power pulses, the laser repetition rate often must be reduced, meaning pulses enter the water at a slower rate. This may be offset by using a higher-order PPM encoding, however overall data rate is usually not preserved. As such, the trade-off between link range (or peak power) and data rate still exists. Furthermore, high peak power lasers may be rather large and power hungry, and are therefore not well suited for applications such as sensor nodes or unmanned vehicles. While baseband schemes like PPM are ideal for long-range/high peak power applications due to the inherent nature of pulsed laser action, short-range links using lower power devices, like LEDs, may also implement some flavor of baseband modulation. The motivation in this case is to minimize power consumption per bit by limiting LED on-time, yet still providing sufficient information bandwidth. (cf., [50], [19])

The other alternative is to use a bandpass modulation scheme like M-PSK or QAM [8, 9]. Because of the continuous wave nature of these modulation formats, the peak power available is often lower than solid-state pulsed laser sources, which may lead to shorter range of operation (or conversely, restrict operation to clearer waters). However, higher data rates may be achieved relative to PPM. A further benefit of bandpass modulation schemes may appear in real environments where noises around DC may be significant (for instance, a low-frequency solar ambient fluctuation). Bandpass systems are also beneficial in full-duplex LOS links operating at the same wavelength, or in retro-reflector links [16]. In turbid LOS links, a receiver must be able to discriminate another transmitter's signal from the backscatter caused by its own transmitter's signal. In retro links, the receiver must discriminate the modulated retro return from the backscatter of the non-modulated interrogator. It has been shown that in underwater laser imaging systems, using intensity modulation can reduce the impact of backscattered light. Due to the distributed nature of backscattered light along the beam path, backscatter has a lowpass response [35]. By using high-frequency modulation, the incoming information signal (centered at high frequency) can be separated from a "washed out" backscatter (located near DC). There are practical downsides to bandpass modulation schemes, depending

on the laser source chosen. For example, in order to modulate a high-power CW source (>1 W), large power-hungry external electro-optic modulators must be used. Furthermore, these bulk devices limit carrier frequencies to <100 MHz. Lower power sources like LEDs or laser diodes may be used to achieve some of the above benefits, however due to the CW nature of laser transmission, may be less power efficient than a baseband scheme. As always, trade-offs exist.

While there have been few published studies comparing the two flavors of modulation side by side, unpublished work from SA Photonics (San Francisco, CA) has demonstrated both QPSK and PPM modulation schemes via a custom semi-conductor/fiber amplified laser in laboratory test tank environments. The laser wavelength was 532 nm, and had an output power of 1–3 W depending on the modulation scheme used. An APD was used at the receiver. 50 Mbps QPSK data (140 MHz RF carrier) was transmitted error free ($<10^{-7}$) over 13.2 attenuation lengths. 15 Mbps was demonstrated error free over ~ 21 attenuation lengths using an 8-PPM modulation format. Note that the higher peak power of the PPM pulses allowed for the error-free range to be extended an additional ~ 8 attenuation lengths, albeit at the cost of a reduced information bandwidth.

New studies are looking at error control coding for the underwater channel. Using forward error correction (FEC) allows link operation at lower SNRs due to the error correcting features of the coding. Cox *et. al.* [14], presented a 3.66m uFSO link that implemented (255,129) Reed–Solomon coding. The transmitter was an 80 mW laser diode at 405 nm, and an amplified photodiode was used at the receiver. The link used an on–off keying (OOK) modulation format at 500 kbps. They showed that for a BER of 10^{-4} , a 2.5 dB E_B/N_0 improvement could be gained by using FEC. In their laboratory experiment, the FEC link could achieve the same BER as an uncoded link even though the attenuation coefficient was 1/4 larger. Everett [21] has shown coding gains of 6.8 dB to 9.5 dB E_B/N_0 using state-of-the-art Turbo Codes with rates ranging from $r = 1/2$ to $r = 1/6$. The same work also showed a coding gain of 7.7 dB to 9.2 dB E_B/N_0 using LDPC codes with code rates varying from $r = 1/2$ to $r = 1/4$. All improvements are relative to an uncoded system with a BER of 10^{-4} . An 80 mW laser diode at 405 nm was used at the transmitter, with a PMT used at the receiver.

9.5.3 Examples of uFSO links

9.5.3.1 Line-of-sight (LOS) links

The simplest uFSO link geometry is the direct point-to-point link (Figure 9.16(a)). In this case, each transmitter/receiver pair must maintain accurate line-of-sight (LOS) between them. For static links where the transmitter and receiver are not expected to move relative to each other, this link is fairly easy to implement. An example of such a link is two sensor nodes, fixed to the bottom of the seafloor. For mobile platforms, such as communication between autonomous underwater vehicles (AUVs), sophisticated pointing-and-tracking (PAT) technologies must be implemented in order to maintain the maximum amount of optical power density incident upon the optical receiver. We

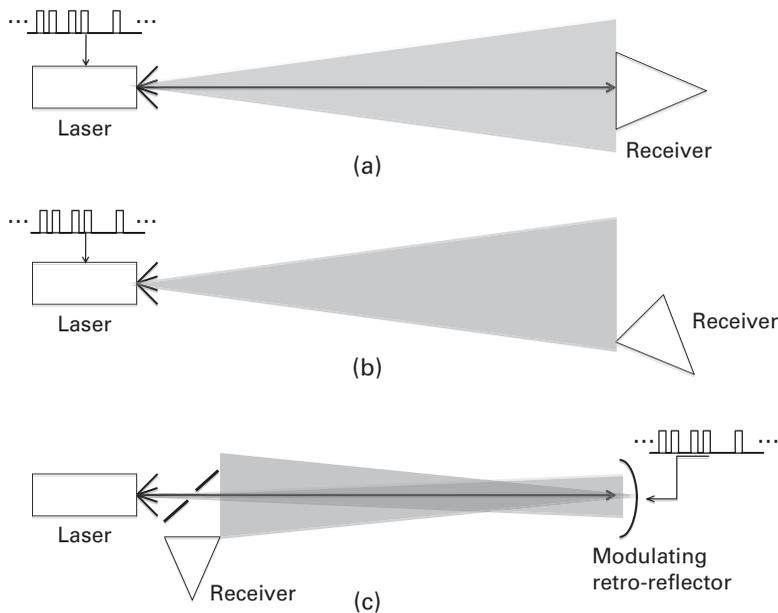


Figure 9.16 uFSO link geometries. (a) LOS, (b) NLOS, and (c) retro-reflector.

have already seen how spatial dispersion may actually help reduce PAT complexity, by making the link less sensitive to changes in transmitter/receiver pointing.

Commercial transceivers for LOS links are just starting to gain visibility. Ambalux (Tucson, AZ) offers a simplex and full-duplex transceiver that transmits 10 Mbps, Ethernet compatible, data. This LED-based system has been operationally demonstrated in an 8 m link ($c = 0.4/\text{m}$). A version of this system has also been demonstrated at longer ranges ($z = 42\text{m}$), and included active pointing and tracking [27].

Laboratory LOS links using bandpass modulation have been demonstrated by [9] which implemented M-PSK and QAM modulation schemes using a 70 MHz carrier with a 3W solid state CW laser and an external modulator. A PMT was used at the receiver. Results showed that even in turbid waters ($c = 3.0/\text{m}$ and $cz = 11$), no loss of modulation depth was observed as a result of temporal dispersion due to forward scattering, even with a large receiver FOV of ~ 100 degrees. As such, the only channel impairment was due to attenuation. Therefore, with enough optical power (in this case, $\sim 3\text{ W}$), a 32-QAM link achieving 5 Mbps could still be closed. Previous studies [8] suggest that at low carrier frequencies ($< 100\text{ MHz}$), the link is relatively immune to temporal dispersion even up to $c = 24/\text{m}$ ($cz = 90$).

Cox *et al.* [15] investigated backscatter suppression in LOS links using polarization shift keying (PolSK). These links also employed UMTS Turbo Codes to minimize data errors. PolSK links can be used to help infer information about the uFSO channel itself, due to the “extra” information inherent in the polarization sensitive signalling scheme. Such information could be used by a transceiver to optimize its configuration for a channel whose IOPs change as a function of time.

9.5.3.2 Non-line-of-sight (NLOS) links

NLOS links are also of interest due to a decreased dependence on pointing and tracking (Figure 9.16(b)). In clear waters, NLOS links are achieved by either intentionally diverging the transmit beam, or by using arrays of high divergence sources (like LEDs). In turbid waters, it is possible that spatial dispersion of a collimated laser source due to scattering may have the same effect, though this is not often considered in analysis.

Researchers at Woods Hole Oceanographic Institute have begun to investigate NLOS transceivers for the application of AUV data mining. Farr *et al.* [22] have performed link budget analysis of NLOS geometries (as well as variations of point-to-point links). While their analysis includes attenuation by the attenuation coefficient, it does not consider spatial dispersion and multiple scattering, and as such is expected to be valid only in clearer waters. Indeed, this is likely to be the case in the open ocean application targeted in their work. It was found that one disadvantage of the NLOS link is the increase of laser power necessary due to the intentional divergence of the transmit beam. In a clear water scenario, analysis showed that a LOS link could achieve 10 Mbps at 100 m with 6 mW of transmit power. However for the NLOS link, a 50-fold increase in transmit power is needed to achieve the same data rates at similar range due to use of a broader transmit beam. That said, laser diodes of 300 mW (or arrays of LEDs totaling the same power) are available in the blue/green, making such a trade-off reasonable. However, in turbid waters, the increase in transmit power is expected to increase exponentially with attenuation coefficient in order to maintain the same data rate or link range. Indeed, follow-on NLOS experiments by [42] showed 5 Mbps at 200 m in clear waters with $c = 0.025/\text{m}$. However operation in turbid waters ($c = 1.25/\text{m}$) necessitated a reduction in both data rate (1 Mbps) and range (30 m). While trade-off studies were performed in [22], exact details on the experimental system, other than the transmitter operating at 470 nm, are hard to come by (i.e. choice of transmitter/receiver, transmitter power, receiver gain, etc.).

Penguin Automated Systems (Ontario, CAN) has developed a NLOS transceiver for tele-robotic operations consisting of 100s of LEDs in a spherical configuration to achieve a 120 degree transmitter field of view. Field tests with this system demonstrated video transfer at 1.5 Mbps at 15 m in turbid lake water [5]. The exact water clarity was not reported, nor has any details regarding the transceiver hardware.

Finally, a unique method of NLOS communications has been proposed by Arnon and Kedar [2, 3]. In this geometry, the transmitted signal is reflected off the air–water interface. Naturally, under normal circumstances the LOS geometry will outperform the “reflective” NLOS link. However the real power in this unique reflective geometry lies in its ability to mitigate obstructions between transceiver nodes.

9.5.3.3 Retro-reflector links

In the links described above, light that is modulated with information is transmitted to a remote optical receiver. In a retro-reflector link, the laser and receiver are situated on one end of the link where there is sufficient power and payload capacity. On the other end of the link is a small passive optical retro-reflector that is coupled to an optical modulator (Figure 9.16(c)). When the laser interrogates a modulated retro-reflector (MRR),

the light is encoded with information and reflected back towards the source where the co-located receiver recovers the information signal. Retro-reflector links are ideal for bottom dwelling sensor nodes, where maximizing battery life is essential to prolonging the lifetime of the sensor. Replacing power-hungry items like lasers and photoreceiver electronics for small, efficient retro-reflectors is an attractive proposal to the underwater communications link designer.

MRR links can be seriously limited by backscatter, similar to the issue previously discussed in full-duplex links. Mullen *et. al.* [36] investigated the use of polarization discrimination to suppress the backscatter signal. This approach relies on the fact that backscattered light tends to retain its original polarization state. However, the polarization state of the light at the retro end can be intentionally rotated so that it is opposite to that of the backscattered light. The receiver polarizer is oriented to accept the retro-reflected light while simultaneously rejecting the backscattered light. Laboratory experimental results showed that the addition of polarization-sensitive components increased the useful range of the MRR link by at least two attenuation lengths.

The challenge now lies in developing an MRR device that can support modulation speeds on the order of > 1 Mbps. While semiconductor multiple quantum well modulators are typically used for atmospheric links in the IR, similar devices are difficult to fabricate in the blue/green. Until this technology becomes more mature, micro-electro-mechanical (MEMS) modulators appear to be the technology of choice. While these modulators have lower bandwidths than the semiconductor modulators, they have the advantage of operating at any wavelength. Two different MEMS approaches have been studied. The first implements a Fabry–Perot cavity with two partially reflective aluminum mirrors; one fixed on a glass substrate and the other on a flexible silicon nitride bridge. The Fabry–Perot cavity is designed such that when interrogated with blue/green light, an applied voltage to the silicon nitride membrane will result in modulation. This approach was recently tested in a laboratory test tank by [16]. The transceiver used a small, compact, 20 mW solid state laser and a PMT receiver. QPSK modulation was implemented at the retro. Error free transmission ($< 10^{-4}$) at 500 kbps was observed up to 5 attenuation lengths, and extended to 6.47 attenuation lengths using rate $r = 1/2$. Reed Solomon forward error control coding. 1 Mbps was achieved out to 2.77 attenuation lengths, which could be extended to 3.76 attenuation lengths with FEC.

In the second MEMS approach, a MEMS fabricated mirror was used as one of the faces of a corner cube. By deflecting the mirror down using electrostatic actuation the retro-reflectivity of the corner cube could be spoiled, thus modulating the light. This MEMS structure was capable of bandwidths exceeding 100 kHz [55].

9.5.3.4 Networking

Networking underwater is an up and coming area of study. With the advent of AUVs, ROVs, and other underwater robots, ocean exploration is poised to grow rapidly in the future. A popular targeted application for underwater communication networks is data mining of scientific sensor nodes. Typically, physical retrieval of the nodes is undesirable

(or in some locations, impossible). In order to limit the frequency of costly physical retrievals, the nodes must either have large memories, complex data processing, or low sampling rates; none of which is particularly desirable. Acoustics could be used as a method of wireless data transfer between the node and, say, an AUV. However due to the slow data rates and high latency associated with acoustics, extra burden is placed on both the sensor node and the AUV.

Researchers are now considering optics as an alternative due to the higher bandwidths they support. On the sensor side, sampling frequency can increase now that higher data rates are supported between the sensor and the AUV. On the AUV side, mission time is reduced as less time is spent visiting each node since data transfer time is significantly decreased (as compared to acoustics).

As a step towards autonomous data mining, a series of tests have been reported by [23] using the human operated vehicle *Alvin*. In these tests, a compact, real-time, system was developed to eventually fit on a small unmanned vehicle (a similar system is reported in [42]). Results in clear water ($c = 0.05/\text{m}$) show error free detection at 1 Mbps out to 100 m. This study also showed full-duplex communications with the ROV *Nereus*. Real-time video transfer from the ROV shows the potential for enabling real-time control of underwater vehicles in addition to data transfer from the vehicle to a controlling platform.

Researchers at MIT have developed prototype sensor nodes and AUVs in order to demonstrate the potential of wireless data collection underwater (cf. [49, 50]). These studies have used a blend of both acoustics and optics in order to accomplish autonomous data collection. By using a blend of acoustics and optics, the system designer can tailor the transceiver based on the expected mission. For example, it is anticipated the acoustics can be used for location of sensor nodes at long range. This low bandwidth information can be used to guide the vehicle within range of the optical link. Once close enough, the optical link will be used for high-speed data transfer. Should the optical link be broken, or the vehicle move out of range, the acoustic link can still be used to transmit data, albeit at a slower rate. In addition, the low bandwidth acoustic link can be used for network activities such as medium access and node discovery. These items can occur in the background during a UUV mission, leaving the optics only for the high-speed data transfer at mid to close range. The most recent version of the optical transceiver has been shown to operate a 2.28 Mbps over 50 m in clear pool tests [19]. The transmitter is an array of 470 nm LEDs, resulting with a total output of 10 W. An APD is used as the photoreceiver. While yet reported, the FPGA-based hardware can support several different modulation schemes, providing a software radio approach to underwater communications. Previous versions of the optical transceiver have been used for wireless underwater robot control in a clear pool 13.7m \times 7.3m in dimension [18].

A slightly different application is considered by [1], opting for an all-optical approach to networking. Here, the authors considered small sensor nodes oriented in a star-like topology, where several nodes all transmit information to a larger “cluster head” node. The cluster node must implement some method of medium access control (MAC) in order to handle contention between its “slave” nodes. Unlike RF protocols, which can

attempt to, “sense” the medium to determine whether or not another node is transmitting, the directional nature of FSO communications makes it largely impossible to know if a nearby slave node is attempting communication with the master node. As such, it is the job of the master node to (1) determine which nodes are trying to send data, and (2) adaptively regulate the medium between those nodes in order to avoid data stream collisions. The custom protocol developed also runs adaptively, meaning that it can handle slave nodes that move in and out of the master node’s range. Such a capability would be highly desired, for instance, in an underwater network with swarms of AUVs.

9.6 Summary

Presented in this chapter is the unique and quickly growing area of wireless optical communications underwater. The underwater channel is a complex environment that has many challenges not seen in other FSO links. Researchers are currently employing both theoretical and experimental tools to characterize the spatial dispersion (beam spreading) seen underwater due to scattering.

In our discussion of spatial dispersion, we find that for attenuation lengths less than a diffusion length (as given by Lerner and Summers in Eq. (9.18)), the only effect of the water on the optical signal is an exponential reduction in power, as given by the attenuation coefficient. Furthermore, the channel appears to support frequencies up to 1 GHz (and possibly beyond). We can broadly assume then that in this “minimally scattered” regime, that the only “loss” is due to the exponential attenuation given by the IOPs. Theoretical modeling has shown that in this regime, performance may also be sensitive to changes in system parameters like FOV.

Beyond a diffusion length however, both spatial and temporal dispersion begins to have a more marked effect. We have denoted this as the “multiple scattering” regime. Here, the rate of attenuation is given by k_{sys} , due to the contribution of scattered light. While it may be beneficial that the intensity decreases at a slower rate in this regime, we have also seen that the frequency response supported by the water begins to decay as well. This rate of decay is largely influenced by the shape of the VSF. While it has yet to be developed, we foresee a similar “ k_{sys} ” term for describing the rate of modulation loss versus attenuation length or range. Regardless, it should be obvious that in the multiple scattering regime, the “loss term” can be dependent on different variables and in different ways than the minimally scattered regime.

The point we wish to make here is that while experiment and modeling are beginning to help the link designer understand how both the environment and system impacts the link, we have barely scratched the surface in achieving an exhaustive parameterization of all possible regimes. Furthermore, characterizations will most certainly vary for different links (LOS, NLOS, retro, etc..). Continued investigation remains a significant, non-trivial task both in experiment and simulation. Regardless, we see that as the physics of propagation is becoming better understood, “real-world” systems will also improve in terms of technical maturity.

References

- [1] Agrawal, N., Davis, C. C., Milner, S. D. Free space optical sensor networking for underwater sensing applications. *Intelligent Sensors, Sensor Networks, and Information Processing (ISSNIP), 5th International Conference on*, 475–480, 2009.
- [2] Arnon, S and Kedar, D. Non-line-of-sight underwater optical wireless communication network. *Journal of the Optical Society of America A*, **26** (3), 530–539, 2009.
- [3] Arnon, S. Underwater optical wireless communication network. *Optical Engineering*, **49**(1), 2010.
- [4] Austin, R. and Halikas, G. *The Index of Refraction in Seawater*, Scripps Institution of Oceanography Visibility Laboratory, 1976.
- [5] Baiden, G., Bissiri, Y. High bandwidth optical networking for underwater untethered telerobotic operation. *IEEE Proc. OCEANS 07 Conf.*, 1–9, 2007.
- [6] Bartolini, L., De Dominicis, L., Ferri de Collibus, M. *et al.* Experimental evidence of signal-optical noise interferencelike effect in underwater amplitude-modulated laser optical radar systems. *Optics Letters* **33**, 2584–2586, 2008.
- [7] Bohren, C. F. and Huffman, D. R. *Absorption and Scattering of Light by Small Particles*. Wiley, 1998.
- [8] Cochenour, B. M., Curran, T., Mullen, L. J., Laux, A. E. Effects of multiple scattering on the implementation of an underwater wireless optical communications link. *IEEE Proc. OCEANS 06 Conf.*, 1–6, 2006.
- [9] Cochenour, B. M., Mullen, L. J., Laux, A. E. Phase coherent digital communications for wireless optical links in turbid underwater environments. *IEEE Proc. OCEANS 07 Conf.*, 1–5, 2007.
- [10] Cochenour, B. M., Mullen, L. J., Laux, A. E. Characterization of the beam-spread function for underwater wireless optical communications Links. *IEEE Journal of Oceanic Engineering*. **33** (4) 513–521, 2008.
- [11] Cochenour, B. M., Mullen, L. J., Rabinovich, W. S., Mahon, R. Underwater optical communications with a modulating retro-reflector. *Proc. SPIE*, **7317**, 2009.
- [12] Cochenour, B. M., Mullen, L. J., Muth, J. F. Effect of scattering albedo on attenuation and polarization of light underwater. *Optics Letters*, **35** (12), 2088–2090, 2010.
- [13] Concannon, B. M. and Davis, J. P Results of a Monte Carlo investigation of the diffuse attenuation coefficient. *Applied Optics*, **38**(24), 5104–5107, 1999.
- [14] Cox, W. C. A 1Mbps underwater communication system using a 405nm laser diode and photomultiplier tube, M.S. thesis, North Carolina State University, Raleigh NC, 2007.
- [15] Cox, W. C., Hughes, B. L., Muth, J. F. A polarization shift keying system for underwater optical communications. *IEEE Proc. OCEANS 09 Conf.*, 1–4, 2009.
- [16] Cox, W. C., Gray, K. F., Simpson, J. A. *et al.* A MEMS Blue/Green Retroreflecting Modulation for Underwater Optical Communication. *IEEE Proc. OCEANS 10 Conf.* Accepted paper, 2010.
- [17] Dagleish, F. R., Caimi, F. M., Vuorenkoski, A. K. *et al.* Efficient laser pulse dispersion codes for turbid undersea imaging and communications applications. *Proc. SPIE*, vol. 7678, 2010.
- [18] Doniec, M., Vasilescu, I., Detweiller, C. *et al.* AquaOptical: A lightweight device for high-rate long-range underwater point-to-point communication. *Marine Technical Society Journal*, **44**(4), 55–65, 2010.

- [19] Doniec, M., Rus, D. BiDirectional optical communication with AquaOptical II. *International Conference on Communication Systems (ICCS)*. Accepted paper, 2010.
- [20] Duntley, S. *Underwater Lighting by Submerged Lasers and Incandescent Sources*. Scripps Institution of Oceanography Visibility Laboratory, ch. 7, 1971.
- [21] Everett, J. S. Forward-error correction coding for underwater free-space optical communication. M. S. thesis, North Carolina State University, Raleigh, North Carolina, 2009.
- [22] Farr, N., Chave, A., Freitag, L. *et al.* Optical modem technology for seafloor observatories. *IEEE Proc. OCEANS 05 Conf.*, **1**, 928–934, 2006.
- [23] Farr, N., Bowen, A., Ware, J., Pontbriand, C., Tivey, M. An integrated, underwater optical/acoustic communications system. *IEEE Proc. OCEANS 10 Conf.*, 1–6, 2010.
- [24] Hanson, R., and Radic, S. High bandwidth underwater optical communication, *Applied Optics* **47**, 277–283, 2008.
- [25] Jaruwatanadilok, S. Underwater wireless optical communication channel modeling and performance evaluation using vector radiative transfer theory. *IEEE Journal on Selected Areas in Communications* **26** (9) 1620–1627, 2008.
- [26] Jerlov, N.G. *Marine Optics*, 1st edn. Elsevier Science, 1976.
- [27] Lacovara, P. Fiber optic interconnect technology SBIR phase II final report: Test. Defense Technical Information Center, Alexandria, VA, 2006.
- [28] Lacovara, P. High-bandwidth underwater communications. *Marine Technology Society Journal*. **42** (1), 93–102, 2008.
- [29] Lerner, R. M., and Summers, J. D. Monte Carlo description of time-and space-resolved multiple forward scatter in natural water. *Applied Optics*, **21** 861–869, 1982.
- [30] Lutomirski, R. F., Ciervo, A. P., Hall, G. J. Moments of multiple scattering. *Applied Optics* **34**(30), 7125–7136, 1995.
- [31] McLean, J. W., Crawford, D. R., Hindman, C. L. Limits of small angle scattering theory. *Applied Optics* **26** (11), 2053–2054, 1987.
- [32] McLean, J. W., Freeman, J. D., Walker, R. E. Beam spread function with time dispersion. *Applied Optics* **37**(21), 4701–4711, 1998.
- [33] Mobley, C. *Light and Water*. Academic Press/Elsevier Science, 1994.
- [34] Morel, A. Optical properties of pure water and pure sea water. In: *Optical Aspects of Oceanography*, N. Jerlov and E.S. Nielson, eds., 1–23, Academic Press.
- [35] Mullen, L.J., Laux, A.E., Cochenour *et al.* Demodulation techniques for the amplitude modulated laser imager. *Applied Optics*, **46**, 7374–7383, 2007.
- [36] Mullen, L., Cochenour, B., Rabinovich, W., Mahon, R., Muth, J. Backscatter suppression for underwater modulating retroreflector links using polarization discrimination. *Applied Optics*, **48** (2), 328–337, 2009.
- [37] Mullen, L. J., Laux, A. E., Cochenour, B. M. Propagation of modulated light in water: implications for imaging and communications systems. *Applied Optics* **48** (14), 2607–2612, 2009.
- [38] Mullen, L., Alley, D., Cochenour, B. Investigation of the effect of scattering agent and scattering abedo on modulated light propagation in water. *Applied Optics* **50** (10), 1396–1404, 2011.
- [39] Pegau, W. S., Gray, D., Zaneveld, J. R. V Absorption and attenuation of visible and near-infrared light in water: dependence on temperature and salinity. *Applied Optics* **36** (24) 6035–6046, 1997.
- [40] Petzold, T. J. *Volume Scattering Functions for Selected Ocean Waters*. Scripps Institution of Oceanography Visibility Laboratory, 1972.

-
- [41] Polynkin, P., Roussev, R., Fejer, M. M., Peyghambarian, N., Moloney, J. Laser transmitter for undersea communications using third-harmonic generation of fiber-laser system at 1.5 μm . *IEEE Photonics Technology Letters*, **19** (17), 1328–1330, 2007.
- [42] Pontbriand, C., Farr, N., Ware, J., Preisig, J., Popenoe, H. Diffuse, high bandwidth optical communications. *IEEE Proc. OCEANS 08 Conf.* 1–4, 2008.
- [43] Pope, R. M., and Fry, E. S. Absorption spectrum (380–700nm) of pure water. II. Integrating cavity measurements. *Applied Optics*, **36** (33), 8710–8723, 1997.
- [44] Prieur, L. and Sathyendranath, S. An optical classification of coastal and oceanic waters based on the specific spectral absorption curves of phytoplankton pigments, dissolved organic matter, and other particulate materials. *Limnology and Oceanography*, **26**(4), 671–689, 1981.
- [45] Simpson, J. A. A 1Mbps underwater communications system using LEDs and photodiodes with signal processing capability. M.S. thesis, North Carolina State University, Raleigh, NC, 2007.
- [46] Simpson, J. A., Hughes, B. L., Muth, F. A spatial diversity system to measure optical fading in an underwater communications channel. *Proc IEEE OCEANS 09 Conf.*, 1–6, 2009.
- [47] Sogandares, F. M., and Fry, E. S. Absorption spectrum (340–640 nm) of pure water. I. Photothermal measurements. *Applied Optics*, **36** (33), 8699–8709, 1997.
- [48] Stotts, L. B. Closed form expression for optical pulse broadening in multiple-scattering media. *Applied Optics* **17** (4), 504–505, 1978.
- [49] Vasilescu, I., Kotay, K., Rus, D., Dunbabin, M., Corke, P. Data collection, storage, and retrieval with an underwater sensor network. *ACM Proc. 3rd International Conference on Embedded Networked Sensor Systems (SenSys)*, 154–165, 2005.
- [50] Vasilescu, I. Using light underwater: devices, algorithms and systems for maritime persistent surveillance, PhD thesis, Massachusetts Institute of Technology, Cambridge MA, 2009.
- [51] Wells, W. H. Loss of resolution in water as a result of multiple small-angle scattering. *Journal of the Optical Society of America*, **59** (6), 686–691, 1969.
- [52] Wells, W. H. Theory of small angle scattering. In: *Optics of the Sea*. AGARD Lect.. Brussels, Belgium: NATO, vol. 61, ch. 3.3, 1973.
- [53] Zege, E. P., Ivanov, A. P., Katsev, I. L. *Image Transfer Through a Scattering Medium*. Springer-Verlag, 1991.
- [54] Zege, E. P., Katsev, I. L., Prikhach, A. S., Ludbrook, G. D., Brusaglioni, P. Analytical computer modeling of the ocean lidar performance, *SPIE Proceedings*, **5059** 189–199, 2002.
- [55] Ziph-Schatzberg, L., Bifano, T., Cornelissen, S., Stewart, J., Bleier, Z. Deformable MEMS mirrors in secure optical communications system. *Proc. SPIE*, **7318**, 2009.

10 The optical wireless channel

Roger Green and Mark Leeson

10.1 Introduction

Infrared (IR) indoor optical wireless (OW) potentially combines the high bandwidth availability of optical communications with the mobility found in radio frequency (RF) wireless communication systems. So although IR is currently overshadowed by a multitude of home and office RF wireless networking schemes, it has significant potential when bandwidth demand is high. Compared to an RF system, OW offers the advantageous opportunity for high-speed medium- to short-range communications operating within a virtually unlimited and unregulated bandwidth spectrum using lower-cost components. Furthermore, OW systems may be securely deployed with immunity to adjacent communication cell interference because the signal radiation cannot penetrate opaque barriers such as walls. Low transceiver costs, coupled with rapid deployment and compatibility with the existing optical fiber communication systems mean that OW is an attractive alternative when fiber deployment is difficult [1], [2]. Moreover, bandwidth congestion in the RF domain has led to the installation of optical hotspots in public buildings and application-specific OW hotspots can be integrated into future office designs [3]. Optical wireless communication has attracted considerable attention from the academic community. From its short-distance, low-speed origins [4], OW has become a viable addition to communication systems with promising prospects, having passed the 155 Mbps indoor landmark during the 1990s [5]. In recent years, techniques such as multispot diffusion (MSD) [6], angular diversity [7], and rate adaptive modulation [8] have been adopted as will be discussed later in this chapter. Furthermore, the conventional RF community has now accepted that optical wireless is part of the electromagnetic communication spectrum thanks to work by the Wireless World Research Forum [9] and the continual work of the commercial Infrared Data Association (IrDA) [10]. Current diverse application requirements suggest that the future communication framework will benefit from a combined RF and optical infrastructure [11]. Nevertheless, the technology is not without its drawbacks compared to RF techniques. A base station needs to be situated in every room and the signal is susceptible to the intense ambient background light from the sun, incandescent and fluorescent lights. Thus, the first sections of the chapter provide a brief overview of the system configurations,

sources, detectors, and filters used for OW followed by consideration of bit error rate (BER) performance in typical indoor scenarios.

10.2 System configurations

Indoor OW network topologies are classified according to the degree of directionality of the transmitter and the receiver, and on the existence of a direct path between them. Based on these factors, optical wireless links are classified as: directed, non-directed, or hybrid. Depending on the existence of a direct path between the transmitter and the receiver, wireless IR links can be classified as line-of-sight (LOS) or non line-of-sight (non-LOS) [12]. Directed links do not suffer from multipath distortion. They also improve power efficiency because, in the case of the transmitter, the emitted energy is concentrated into a narrow beam (which results in more power per unit area at the transmitter). In the case of the receiver, the detector can incorporate a directive optical element of high gain (which increases its sensitivity). A narrow emission beam transmitter and a narrow field of view (FOV) receiver allow the transmitted optical power to be reduced (for a given distance). This improves the power budget of the system as well as eye safety. LOS links improve power efficiency and minimize multipath distortion. Non-LOS links, on the other hand, increase link robustness as they base their operation on the use of reflective surfaces, which allow the system to operate even when obstacles exist between the transmitter and the receiver. This means that alignment between the transmitter and the receiver is not necessarily required. The main disadvantage of the LOS configuration is the possibility of blocking from persons and objects and the fact that it sometimes requires careful alignment of the transmitter and the receiver.

Taking into account the individual characteristics of the directed and of the LOS links it is not difficult to realize that the configuration that provides better power efficiency and minimum multipath distortion is directed-LOS. The main drawbacks of this configuration are its limited coverage and its susceptibility to blocking. A special case of this configuration is the tracking system, which benefits from the advantages offered by directed-LOS links, but provides a wide coverage at the same time. Unfortunately, mechanical tracking systems are expensive and increase system complexity. Other options developed to minimize multipath dispersion and to optimize power consumption include the use of quasi-diffuse transmitters and angle diversity receivers [13].

A potentially attractive configuration is the non-directed non-LOS or *diffuse link*. In this topology, careful alignment between the transmitter and the receiver is not required because the FOV of the receiver and the emission beam of the transmitter are wide. This, and the fact that this configuration makes use of the reflective properties of ceilings and walls to distribute the energy as uniformly as possible in a close environment, implies that this topology also provides maximum coverage and mobility. Furthermore, this configuration presents the advantage of being able to operate even when persons and objects are blocking the direct path between the transmitter and the receiver. The main disadvantage of the diffuse configuration is that it presents maximum multipath

distortion spreading the transmitted pulses, introducing intersymbol interference (ISI) and causing loss of pulse amplitude. This means that a maximum transmission speed can be calculated depending on the size of the room where the communication link is operating.

10.3 Optical sources

One of the advantages of optical wireless is that it leverages the transmitter and receiver developments in optical fiber systems to offer inexpensive solutions. There are two basic source choices, namely the light emitting diode (LED) and the laser diode (LD). The common range of wavelengths for operation is 780 nm to 950 nm [14] since there are readily available off-the-shelf devices in that band. Utilization of longer wavelengths (typically 1300 nm and 1550 nm) precludes the use of cheap silicon detectors and GaAs laser diodes [15] thus adding to the cost. The choice between LEDs and LDs depends on the application and the configuration where they are to be used. LEDs, for instance, are preferred for hybrid LOS or non-LOS short-distance indoor applications that require some degree of mobility. This is because, as they emit a broad beam, careful alignment of the receiver and the transmitter is not required. Also, LEDs are generally cheaper and harder to damage than laser diodes, and they are allowed to emit higher optical powers without damaging the human eye. Furthermore, LED driver circuits are simpler and LEDs do not require stabilization against temperature variations. LDs, on the other hand, are often preferred for high-speed directed-LOS links in outdoor applications as the emission beam of these sources is narrow. This means that the optical power per unit area is higher allowing for longer transmission distances. Also, LDs can be used at higher modulation rates than LEDs, allowing for high-speed data communication transmission. Unfortunately, the driver circuit of LDs is more complicated, and the emitted power is limited due to eye safety considerations. Furthermore, LDs are sensitive to temperature variations that modify their spectral emission. The major differences between LEDs and LDs for optical wireless applications are summarized in Table 10.1.

Table 10.1 Comparison of LEDs and LDs (adapted from [12] and [14]).

| Characteristic | LED | LD |
|-----------------------------|----------------------------------|--|
| Optical spectral width (nm) | 25 – 100 | $<10^{-5} - 5$ |
| Directionality | Broad (divergence $> 15^\circ$) | Narrow ($< 10^\circ$) |
| Modulation bandwidth | \sim kHz to \sim MHz | \sim kHz to \sim GHz |
| Special circuitry required | None | Threshold and temperature compensation |
| Eye safety | Considered eye safe | May need to be rendered eye safe |
| Reliability | High | Moderate |
| E/O conversion efficiency | 10-20% | 30-70% |
| Cost | Low | Moderate to high |

10.3.1 Eye safety

In addition to any cost implications, LDs may have issues with eye safety, particularly for indoor systems where the emitted optical power has to be strictly restricted. The International Electrotechnical Commission (IEC) document IEC 60825 defines the maximum exposure limits [16]. This standard does not distinguish between the laser and LED emission level and formulas have also been established by the American National Standards Institute as a guideline for the safe use of lasers in terms of the maximum permissible exposure (MPE) values of intrabeam viewing for near point sources [17]. The MPE values for the major indoor OW wavelength range of 700 nm to 1400 nm are shown in Figure 10.1 for exposures of one second, one minute, and one hour. It may be observed that there is a significant increase in the MPE threshold because water absorption in the cornea prevents the laser power from reaching the retina.

The decreased safety hazards at longer wavelengths offer the use of high-power light sources and therewith higher bit rates if one is prepared to stand the increased component cost. Nevertheless, shorter-wavelength light sources are perfectly adequate for low to medium data rates up to 100 Mbps where LEDs may well be employed offering bandwidths only in the MHz range [18]. To round off this part, we consider the impact of the beam divergence. The IEC standard maximum power in the range 700 nm to 1050 nm is found by considering exposure over a ten-second period using an empirical formula and a defined standard configuration. It is thus possible to quantify the allowed emitted power for Class 1 operation for any source size and divergence half angle. For illustration, Figure 10.2 shows a plot of the allowed Class 1 power versus beam divergence for several source sizes operating at 850 nm.

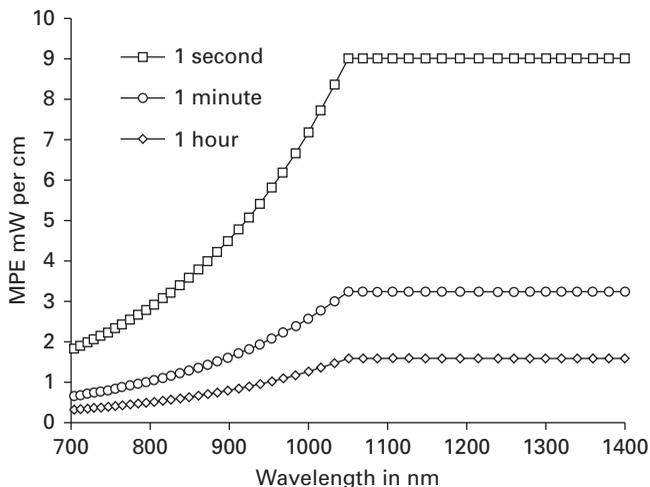


Figure 10.1 MPE variation with wavelength for intrabeam different viewing exposure times.

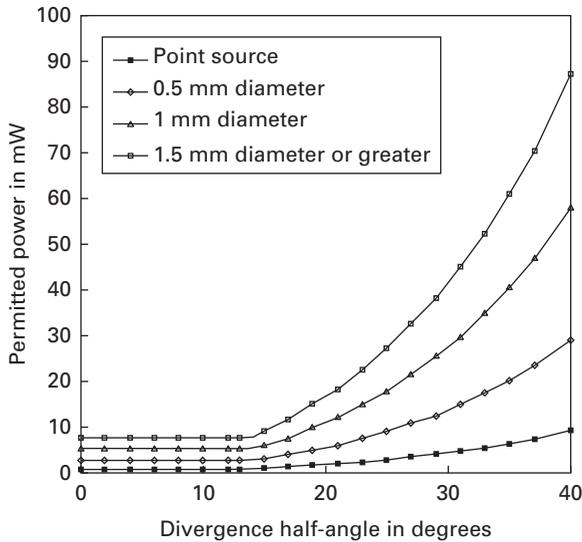


Figure 10.2 Maximum Class 1 power as a function of beam divergence.

10.4 Optical detectors

Ideally, the detector should offer high sensitivity and fidelity, with high optical to electrical efficiency conversion, a large detection area, minimum noise, and short response times at low cost and high reliability. Needless to say, many of these requirements conflict and thus some compromise is necessary. The choice of a photodetector for optical communications is generally between pin photodiodes, avalanche photodiodes (APDs), and metal-semiconductor-metal (MSM) photodiodes [18]. The first two are usually the detectors of choice for OW due to their widespread commercial availability. A pin photodetector is simple both in structure and to employ but is about 10–15 dB less sensitive than an APD [12]. The increased power margin offered by APDs delivers a system that is more robust to pointing inaccuracy and other losses. However, they need high bias voltages and exhibit highly nonlinear current gain and sensitivity to reverse bias voltage giving rise to strong modulation effects from ambient light interference [19].

The performance of a pin photodiode may be quantified via the absorption coefficient α and the structure of the device [18]. The power absorbed in a length x of semiconductor is proportional to $\exp(-\alpha x)$, and the power that is absorbed creates the photocurrent. Thus, for a pin diode with a front face reflectivity R_f and an intrinsic region that stretched from x_1 to x_2 the responsivity is given by

$$\Re = \frac{q\lambda}{hc} (1 - R_f) \{ \exp(-\alpha x_1) - \exp(-\alpha x_2) \} \quad (10.1)$$

where q is the electronic charge, h is Planck's constant, c is the speed of light, and λ is the wavelength of the light.

The other consideration is that of detector bandwidth, determined by the capacitance of the device that limits the maximum speed of operation in conjunction with the resistance presented by the receiver. However, to increase the amount of light in indoor OW systems, large area detectors are to be preferred but this increases the capacitance. The trade-off implied and methods to ameliorate the effect of the capacitance have been the subject of some discussion in the literature [20] and will also be considered later in this chapter. For the illustrative calculations that follow, a pin detector is assumed to be employed as this is the simplest and cheapest option.

10.5 Optical filters

Optical filtering is widely employed to minimize or eliminate the noise introduced in the detector by different sources of background illumination. The two most common optical filtering techniques used to eliminate radiation at unwanted wavelengths are: (a) creating a filter–detector combination and (b) using a thin-film optical filter before the photodetector, and these are now briefly described.

10.5.1 Filter–detector combination

This technique involves using a material with long-pass characteristics (like a GaAs substrate or a colored plastic or glass) in conjunction with a silicon detector. This combination behaves effectively as a bandpass filter, with one of the edges of the pass-band created by the roll-off of the photodetector's frequency response, and the other by the edge of the filtering material's transmittance.

This process can be illustrated using (10.1) for the response of the photodiode and the transmittance of an appropriate filter. Multiplying these together gives the overall responsivity and the results are shown in Figure 10.3 for a Silicon pin diode with a front face reflectivity of 0.2, $x_1 = 2 \mu\text{m}$, $x_2 = 500 \mu\text{m}$, with absorption data taken from Aspens [21] with the characteristics of the Schott RG 780 filter [22]. It can be seen that, since the photodiode only responds to wavelengths below 1100 nm and the filter passes light at wavelengths above 780 nm, an optical pass-band of approximately 320 nm results.

10.5.2 Thin film filter

Unfortunately, this pass-band is too wide for applications that require maximum background noise rejection or for high-speed applications using transmitters with narrow optical spectrum. In these cases, the second technique, which consists in using a thin-film optical filter at the entrance of the photodetector, gives better results [12]. Thin-film optical filters consist of multiple layers of thin dielectric film, which can be combined to provide very narrow optical pass-bands (sometimes below 1 nm). Their main drawback is that their response varies with the angle of the incident light. This effect has to be taken into account when designing the optical front end of the receiver, especially

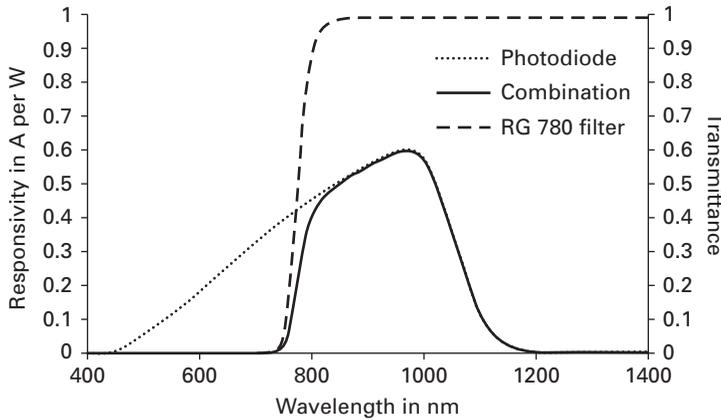


Figure 10.3 Calculated responsivity of silicon p-i-n photodiode, transmittance of Schott RG-780 filter shown with the overall responsivity.

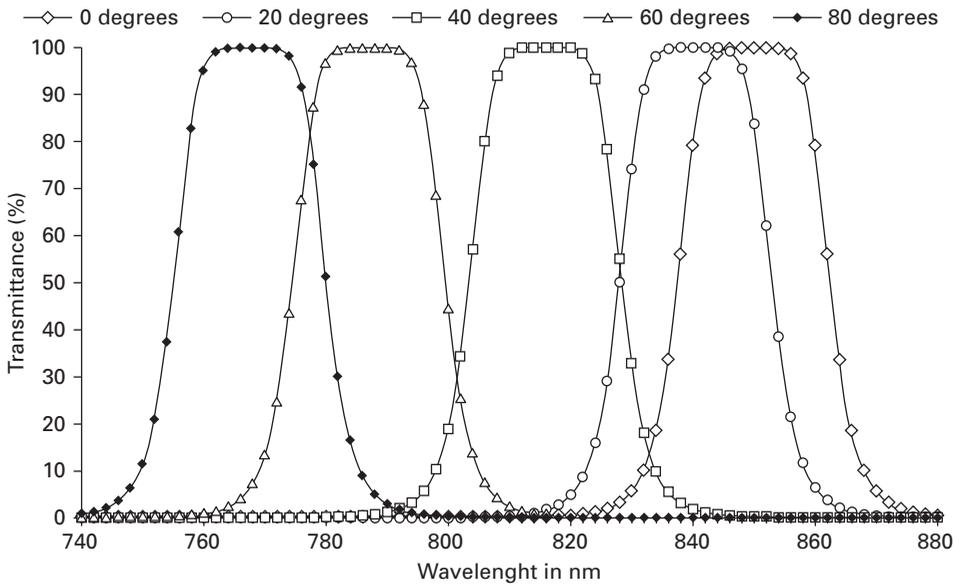


Figure 10.4 Angular transmission characteristics of a thin-film bandpass filter calculated using the Butterworth approximation for a range of incident angles.

in mobile applications and in systems where the filter is used in conjunction with an optical concentrator. Figure 10.4 shows the transmission spectrum of a thin-film optical bandpass filter (using a Butterworth approximation [12]) and its angular dependence with angle of incidence. When using optical filters it is important to remember that, in order to maximize the signal-to-noise ratio (SNR) at the receiver, the optical spectrum of the transmitter must lie within the pass band of the filter.

10.6 Nature of the optical wireless channel

Intensity modulation direct detection (IM/DD) is the only practical transmission technique in virtually all short-range, indoor applications [23]. An IR source emits an instantaneous optical power $X(t)$ that propagates through the channel and is incident upon the optical detector at the receiver. The received signal, $Y(t)$, is then the integration of the detected optical power over the area of the detectors surface, multiplied by the responsivity of the photodetector \mathfrak{R} [24]. At IR wavelengths, a typical photodetector will be in excess of 10^6 wavelengths across, resulting in optically incoherent reception without multipath fading [25]. The channel can be modeled as baseband linear time invariant (LTI) [26], with impulse response $h(t)$, that only changes significantly when the source, receiver, or reflectors are moved by distances in the order of centimeters. In the majority of indoor applications, transmission takes place in the presence of intense infrared and visible background radiation, $N(t)$. This can be modeled as additive white Gaussian, independent of $X(t)$, giving the simplified channel model for an IM/DD link [27]

$$Y(t) = \mathfrak{R}X(t) * h(t) + N(t) \quad (10.2)$$

where $*$ denotes the convolution operation. Through necessity, as $X(t)$ represents power, it cannot be negative, and the average optical power, P_{av} , is proportional to the time integral of $X(t)$ [28]

$$P_{av} = \lim_{T \rightarrow \infty} \int_{-T}^T X(t) dt; \quad X(t) \geq 0, \forall t. \quad (10.3)$$

This contrasts with the situation found in the electrical domain systems, when $X(t)$ represents amplitude (which may be negative) resulting in a SNR proportional to $|X(t)|^2$. This square law dependence means that it is challenging to produce a high SNR at useful data rates since a higher mean optical power level is needed compared to the electrical domain and the transmitted power cannot exceed predefined eye safety limits set by the IEC 60825 regulations previously discussed in Section 10.3.1.

Furthermore, the optical channel path loss is given by the integral of, $h(t)$, the channel impulse response

$$H(0) = \int_{-\infty}^{\infty} h(t) dt \quad (10.4)$$

meaning that the received power is given by

$$P_{rec} = H(0)P_{av} \text{ watts.} \quad (10.5)$$

Therefore, in addition to the eye safety regulation limit, the specific impulse response between a given source and receiver is also a major influence on the achievement of a suitable SNR at the receiver. The actual impulse response itself is the final channel challenge for an OW system designer and will be examined further in Section 10.9 below.

10.6.1 Channel noise

In addition to the distortion introduced by propagation through the dispersive channel, the signal is also degraded by noise, which in general is a term that represents a broad range of possible real world manifestations that interfere with the desired system performance. In the context of OW communications, however, noise is considered to be a combination of two major factors, shot noise and interference noise [29], [30], [31]. Furthermore, the noise itself is subject to the same channel as the signal and so the quantity of noise arriving at the receiver is dependent upon the room configuration and reflectivity properties, receiver properties such as photodiode area, FOV, position and orientation plus the original distribution of the noise signal from its source. The accurate modeling of this scenario requires detailed room models and often simulation. As an introduction, a simplified analysis will be presented to give a feel for the quantities involved and this will be followed by an overview of the necessary steps to construct a full model.

The steady or very slow time varying background irradiance produced by natural and artificial light sources is usually characterized by the DC current it induces in the receiver photodiode since the shot noise power density, N_0 , is directly proportional to that current, I_b , or, alternatively the incident background noise power, P_b , and is given by [32]

$$N_0 = qI_b = q\Re P_b \quad (10.6)$$

where \Re is the photodiode responsivity and q is the electronic charge.

Thus a baseline calculation takes account just of the shot noise and assumes that non-return to zero on-off keying (NRZ-OOK) is employed. With equal probabilities of transmitting a zero or a one, the probability of error is given in general by [31], [33]

$$P_e = \frac{1}{2} \operatorname{erfc} \left(P_{avg} \Re \sqrt{\frac{T_b}{2N_0}} \right) \quad (10.7)$$

where erfc is the complementary error function, P_{avg} is the average transmitted power and T_b is the bit rate. The argument of the complementary error function is the square root of the optical SNR, so it is straightforward to plot the behavior of P_e as a function of the optical SNR, as shown in Figure 10.5

From the figure, it is apparent that an optical SNR of just over 6.7 dB is needed to achieve the commonly used OW benchmark of $P_e = 10^{-6}$.

10.7 Interference sources

The estimate above is highly optimistic since all indoor OW systems suffer from ambient light interference, which acts both to reduce the effective contrast ratio and to increase shot noise. When dealing with the effects of the channel characteristics on the signal, it can be assumed that the emission wavelengths are known. Furthermore, in the case

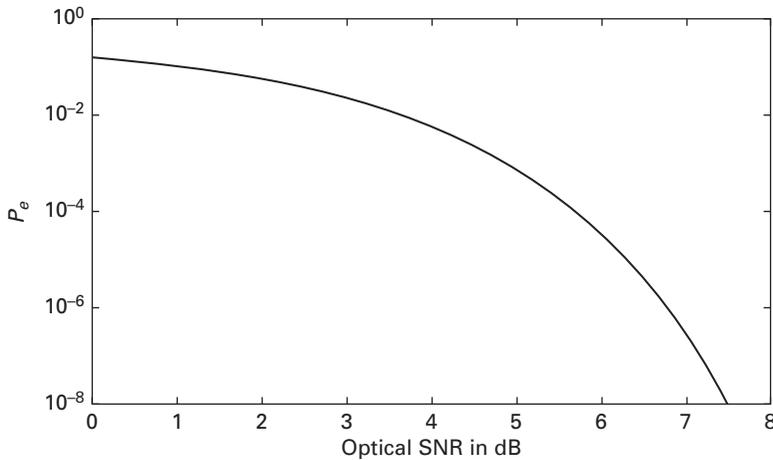


Figure 10.5 General dependence of P_e on optical SNR for an idealized OW channel.

of infrared communications, signals will occupy a fairly narrow band operating at the typical center wavelengths of 780 nm, 870 nm, or 970 nm [29] implying that a given source power is contained within a known range of radiation wavelengths that is well matched with photodiode responsivities. It is more common to find such infrared devices characterized by their radiometric properties [34] whereas illumination light sources are normally characterized by their interaction with the human eye (their photometric properties) or their electrical properties [35].

10.7.1 Sunlight

The effects of the sun on OW links have been apparent since the beginning of work in the area [4] and without filtering, daylight can overpower OW IR links [29]. For modeling purposes, the sun is a relatively broadband source of approximately constant spectral density over the bandwidth of an OW receiver. It thus induces a constant shot noise element in the receiver that can be some hundreds of μA or more [32].

10.7.2 Artificial sources

A 60 W light source does not emit all its power at the wavelengths of interest to optical wireless communications. All illumination light sources, including the sun, emit radiation to varying degree within a silicon photodiode responsivity range. Commonly employed daylight filters with a cut off wavelength < 780 nm remove some of the background radiation but above that wavelength, the light sources still induce a background photocurrent [29]. The effects of background illumination have most commonly been included in the modeling of indoor OW by way of practical measurements such as those in [36], [37] and [38]. Each of the measurement sets in these publications was carried out

in a different environment using different receivers with different properties so although they give very good approximations or expectations of the level of received background noise power that, is present, their results are not totally comparable.

The work by Moreira *et al.* [31] and [32] has provided through measurement, an accurate but nevertheless general model for interference from three types of light bulb: incandescent lamps, fluorescent lamps with conventional ballasts, and fluorescent lamps with electronic ballasts. Each model is based upon the measurements from a wide-ranging collection of lamps from many manufacturers. The resultant Fourier series formulation can be used to represent to a good approximation the level of interference a given type of lamp will produce.

10.7.2.1 Incandescent lamps

Incandescent, or Tungsten filament lamps are set to become less common under the new European Parliament and Council Directive 2005/32/EC [39], under which they are to disappear for energy usage reasons. Nevertheless, they are still very common in offices and homes so continue to contribute to the noise in current indoor optical wireless systems. The interference signal for incandescent lamps may be represented as [32]:

$$i_{\text{incand}}(t) = \frac{I_B}{F_1 A_1} \sum_{k=1}^{\infty} a_k \cos(200\pi kt + \phi_k) \quad (10.8)$$

The magnitude factors, a_k , and the phase factors, ϕ_k , of the 100 Hz harmonics were determined empirically in [36] and the first 10 are significant. The multiplying factor contains I_B , the level of the background radiation, F_1 , the filter attenuation factor and A_1 , which relates the interference amplitude to I_B [31], [32].

10.7.2.2 Fluorescent lamps driven by conventional ballasts

Offices commonly contain fluorescent lamps driven by conventional ballasts and these have interference effects up to around 2 kHz. The interfering signal can be described by

$$i_{\text{fluor}}(t) = \frac{I_B}{F_2 A_2} \sum_{k=1}^{\infty} [b_k \cos(100\pi \{2k - 1\} t + \phi_k) + c_k \cos(100\pi \{2k\} t + \zeta_k)] \quad (10.9)$$

The phase factors, ϕ_k and ζ_k may again be taken from [36] with the first 20 being significant, and F_2 and A_2 are the corresponding values for this interferer. The magnitudes, b_k and c_k , are those of the odd and even Fourier harmonics.

10.7.2.3 Fluorescent lamps driven by electronic ballasts

In recent years, fluorescent lamps driven by electronic ballasts (sometimes known as high-frequency fluorescent lamps) have developed and are likely to become the dominant type in future since this is the technology employed by energy saving lights [39]. Unfortunately, they are a worst case scenario for indoor optical wireless communication systems for two reasons. First, they exhibit settling times following initial switch on

when the output radiation spectrum shifts from one dominated by the infrared emission to visible light after approximately six minutes [29]. This is commonly observed with energy saving bulbs that appear initially dim to the human eye, and is of particular relevance in residential settings where users move between rooms tuning lights on and off. Here, we follow [40] and assume that the bulbs are in their settled state whilst noting that incorporating the effects of bulb switch on would form an interesting topic for future work. The second problem is that the harmonic components of their interference extend to several MHz. The interference signal can be modeled as the sum of two components (low and high frequency) thus:

$$i_{\text{elec}}(t) = i_{\text{low}}(t) + i_{\text{high}}(t) \quad (10.10)$$

where, similar to Equation (10.9)

$$i_{\text{low}}(t) = \frac{I_B}{F_3 A_3} \sum_{k=1}^{\infty} [b_k \cos(100\pi \{2k - 1\}t + \phi_k) + c_k \cos(100\pi \{2k\}t + \zeta_k)] \quad (10.11)$$

$$i_{\text{high}}(t) = \frac{I_B}{F_4 A_4} \sum_{k=1}^{\infty} d_k \cos(2\pi f_{\text{high}} k t + \theta_k). \quad (10.12)$$

The parameters in (10.11) above are those for conventional fluorescent lamps with the appropriate filter and conversion factors; values for d_k and θ_k may be taken from [36] with 22 harmonics making significant contributions and f_{high} equal to 37.5 kHz [32].

10.8 Impact of interference on BER

The work in [31] and [32] also provides the basis for calculations of the effect of ambient interference on the BER. Inclusion of the artificial background sources may be achieved by recognizing that they are periodic and integrating the BER expression over the period. This approach modifies (BER) to give [31]

$$P_e = \frac{1}{T_i} \int_0^{T_i} \left[\frac{1}{2} \operatorname{erfc} \left(\frac{P_{\text{avg}} \mathfrak{N} T_b - v_i(t)}{\sqrt{2N_0 T_b}} \right) + \frac{1}{4} \left\{ 1 + \operatorname{erf} \left(\frac{-P_{\text{avg}} \mathfrak{N} T_b - v_i(t)}{\sqrt{2N_0 T_b}} \right) \right\} \right] dt. \quad (10.13)$$

The period of the integral T_i is the period of the interference signal and $v_i(t)$ is the energy of the interference within the period of a pulse

$$v_i(t) = \int_t^{t+T} i_{\text{int}}(t) dt \quad (10.14)$$

where $i_{\text{int}}(t)$ is the appropriate expression for the type of periodic interferer. When the only source of noise is the shot element from the signal itself, the SNR is

$$\text{SNR}_{\text{max}} = \frac{P_{\text{avg}} \mathfrak{N}}{q} \quad (10.15)$$

where (10.6) has been used. It is straightforward from this to produce signal powers necessary to obtain Figure 10.5, which will typically be in the nW range. In this work, a responsivity of 0.6 AW^{-1} was used [31]. To incorporate interference, (10.15) must be modified so that the current is the sum of the background current levels for the ambient sources. Moreover, the integral (10.14) must be obtained numerically for use in the BER calculation.

As an illustrative example, standard fluorescent lights are considered here in the presence of indirect sunlight. It is assumed that optical filtering is used at the receiver. A value of $200 \text{ }\mu\text{A}$ is used for the sunlight background current, P_{sun} , [31], [32] and the mean background power level for the fluorescent lighting is parameterized by $P_{\text{fluor}} = KP_{\text{avg}}$. This approach makes the optical SNR degradation depend only on the background power level and not on the bit rate, which is chosen to avoid ISI (see Section 10.9 below).

Figure 10.6 shows the results of the calculations for values of $K = 0, 2$, and 4 . The first of these represents only sunlight and (10.15) must be modified since the shot noise comprises that from the signal and that from the sunlight thus

$$SNR_{\text{max}} = \frac{P_{\text{avg}}^2 \mathfrak{R}^2 T_b}{q \mathfrak{R} (P_{\text{avg}} + P_{\text{sun}})}. \quad (10.16)$$

From this equation, one can easily form a quadratic expression for the signal power with no interference. It should be noted that although the shape of the $K = 0$ curve is the same as in Figure 10.7, the actual signal levels differ greatly (with sunlight hundreds of nanowatts are needed whereas in the dark, less than 1 nW). When the level of

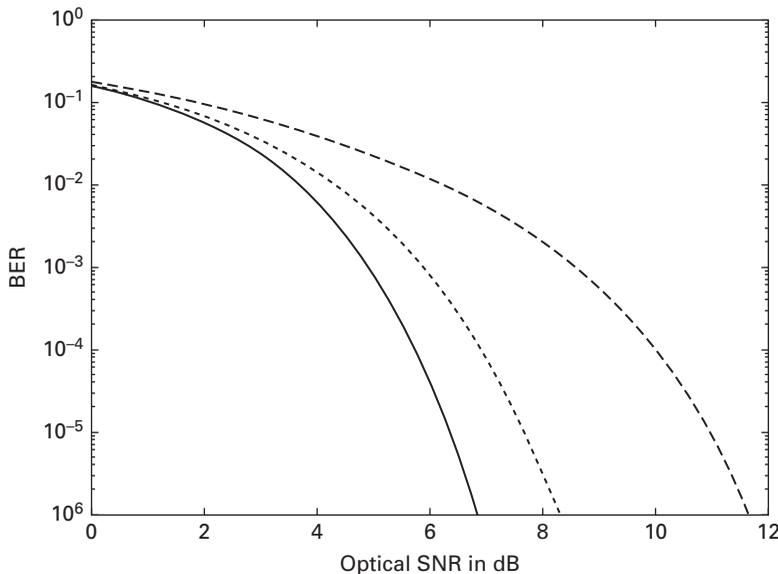


Figure 10.6 SNR versus optical SNR with background sunlight and $K = 0$ (solid line), $K = 2$ (dotted line), and $K = 4$ (dashed line).

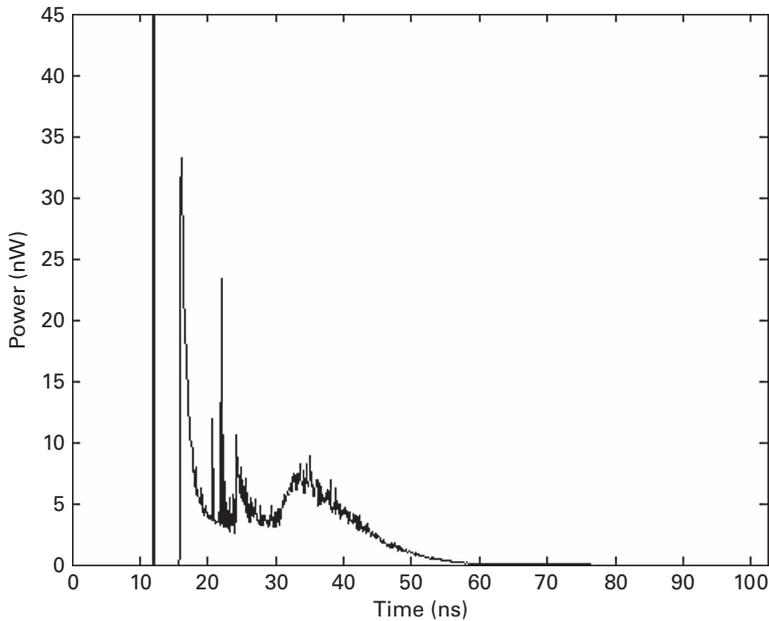


Figure 10.7 Calculated three bounce impulse response.

the fluorescent background light becomes large, power penalties of several dBs rapidly develop.

10.9 Channel impulse response

The impulse response, $h(t)$ is unique to a given source and receiver configuration within the environment in which they are deployed. Hence, the position and orientation of the source and transmitter, together with source properties such as the radiation emission pattern and receiver properties such as FOV and active collection area, are factors contributing to $h(t)$. Coupling these with the dimensions of the environment and any enclosed objects, along with their respective reflectivity coefficients, gives an essentially infinite number of possible impulse responses. Furthermore, the impulse response not only dictates the path loss but also the degree to which ISI will affect the transmitted signal as it is reflected around the room [41], [42].

The closed-form approximations originally employed in early work [4] offer relatively simple ways to model basic configurations and conduct rudimentary analyses of factors such as material reflectivity and the source intensity profile but they are too complex when considering multiple reflections. At the other extreme, experimental characterization [26], [43], [44] is expensive and time consuming and must be performed on a channel-by-channel basis [45].

Although other analytical methods such as the ceiling bounce approach [46] may be employed to add sophistication, it falls to simulation to provide insight into the optical

wireless channel impulse response. A general simulation method was first proposed by Barry *et al.* [47] using techniques borrowed from ray tracing [48]. This offered the ability to determine, with relatively straightforward implementation, the impulse response for a system where the signal underwent any number of reflections for arbitrary source–receiver configurations inside an arbitrary empty rectangular room. However, the algorithm was recursive meaning that the computational time rose exponentially and the memory requirements were impractical beyond three reflections. The memory requirements were addressed by López-Hernández and Betancor in [49] using a method of dividing the simulation into time rather than reflections. Carruthers and Kannan [46] applied the algorithm iteratively allowing for the computational time to be proportional to the square of the number reflections required. Moreover, their method also introduced the possibility of simulating scenarios with multiple transmitters and receivers without considerable time penalties.

An alternative simulation approach, based on a statistical model, was proposed by Pérez-Jiménez, Berges, and Betancor [50]. In this method the root-mean-square (RMS) delay spread and mean excess delay were estimated based purely upon the known geometric factors of the system configuration such as transmitter and receiver positions and orientation. The initial impulse response obtained was then adjusted until it fitted either a Rayleigh or a Gamma function but the method has the disadvantage that it required an initial impulse response. Another related method based upon a mixed deterministic Monte Carlo ray tracing algorithm was beneficial to computational time by excluding rays that would never impinge on the receiver [51]. Finally another statistically based simulation technique was presented in [52] based on a random walk. A ray was generated with a pseudo random direction and traced around the environment with a new pseudo random direction when it hit a surface until it eventually reached the receiver. The method worked well for small spherical diffusers and spherical reflective environments but is not generally applicable to random environments.

Here an outline is given of the methodology employed to incorporate the ISI from multiple reflections. Fuller explanations are contained in [53] and [40] in addition to [47]. Independent of which technology is used in the transmitter, the radiation it emits, if incident upon the boundaries of the environment or the surfaces of objects, will be reflected with an intensity distribution that is predominately determined by the angle of incidence of the radiation, and the roughness of the reflection surface.

Surfaces that are considered rough under the Rayleigh criterion [54] reflect radiation with a intensity distribution function, $R(\phi)$, that can be correctly approximated by Lamberts generalized cosine law [55]

$$R(\phi) = \rho \frac{n+2}{2\pi} P_i \cos^n(\phi) \quad (\text{Wsr}^{-1}) \quad (10.17)$$

where ρ is the reflectivity coefficient of the material, P_i is power of the incident wave, n is the Lambertian order and ϕ is the angle of observation. An ideal Lambertian diffuser, where the surface will look equally bright when observed from any angle, is found by setting $n = 1$. Surfaces that are smooth according to the Rayleigh criterion require consideration of diffuse and specular components, and the interested reader is referred

to the work of Phong for a suitable model [56]. Using the assumption that surfaces within a room reflect radiation in accordance with Lambert's model, the path of rays from the transmitter may be followed to obtain the response of the room to an impulse. It is known [26] that for an intensity modulated direct detection scenario, where the movement of the sources, receivers and any mobile objects within the environment are slow compared to the bit rate of the system, no multipath fading occurs. Furthermore, since the channel can be deemed LTI, it can be completely characterized by its impulse response $h(t)$ [57]. We thus have all that is needed to implement a fuller model once the impulse response has been obtained.

The calculations in Section 10.7 above were made at 10 Mbps so that there was no ISI from pulse to pulse. This statement may be made based on the impulse response that will now be presented [40]. The scenario considered is the room used by Wong *et al.* [33], which had the same dimensions as configuration A in [47]. The room width and depth were 5 m, and the height was 3 m. The room was empty and the reflectivity of all surfaces was set to $\rho = 0.8$; a single Lambertian emission profile ($n = 1$) transmitter was placed centrally upon the ceiling, orientated towards the floor. A receiver with a FOV of 90° , active area 1 cm^2 and unity photodiode responsivity was positioned in the middle of the floor, vertically orientated towards the ceiling. Running a program to implement the method outlined above with three reflections [40] gives an impulse response as shown in Figure 10.7.

The channel impulse response clearly shows a spike for the LOS path followed by a long tail from the reflections that will cause ISI. This is extremely apparent when a simulation is run to include the bit combinations that can occur in transmission using the highest bit rate from [33] of 80 Mbps. Figure 10.8 shows the large penalty incurred at this bit rate for $K = 10$ with 10 Mbps shown for comparison (not impacted by ISI).

10.10 Hardware aspects of the receiver-amplifier in the indoor channel environment

10.10.1 Bootstrapped transimpedance amplifier (BTA)

This configuration combines the well-known benefits of the transimpedance amplifier with those of the bootstrapping approach – which itself will be reviewed and further developed later on in this chapter. The transimpedance amplifier uses positive feedback to gain an effective capacitance reduction – bandwidth enhancement – using a high gain amplifier within a feedback configuration of the general form shown in Figure 10.9.

It can readily be shown that :

$$V_0 = \frac{I_p \cdot R_F}{[1 + 1/A] - j\omega(C_d/A) \cdot R_F}$$

in which I_p is the photocurrent, and other symbols are as defined in the figure above. Thus the bandwidth of the system is enhanced by the gain of the amplifier A . Clearly, the larger the value of A , the greater is the apparent bandwidth enhancement. However,

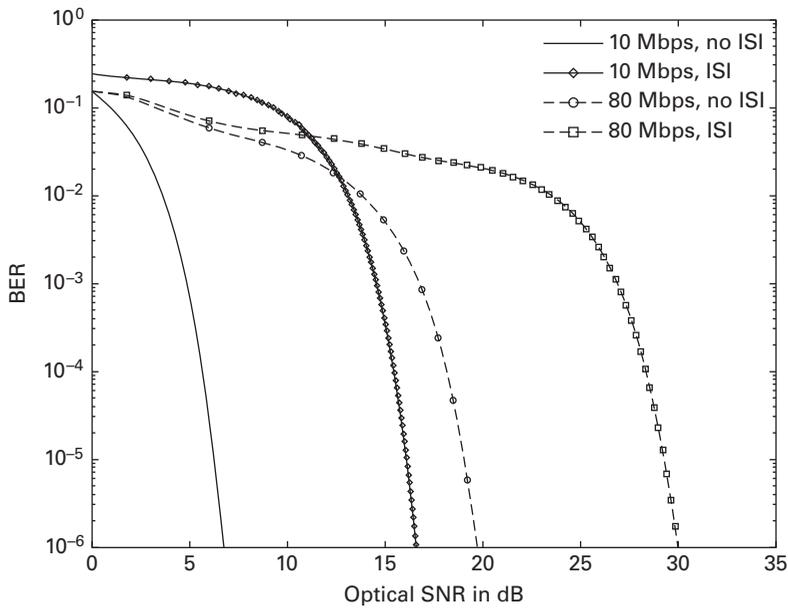


Figure 10.8 BER versus optical SNR with interference.

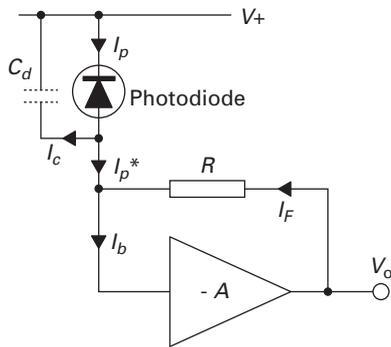


Figure 10.9 Basic transimpedance amplifier configuration.

it can be shown that instability easily occurs for large values, so another method has been devised which requires lower values of A and yet which offers the same or better bandwidth enhancements – therefore with less likelihood of instability, especially at the higher frequencies. The bootstrapping technique is well known as one which also offers bandwidth enhancement – albeit with low gain (< 1), which is not the optimum situation for a front end amplifier in terms of noise figure, as a consequence of the Friis formula:

$$F = F_1 + \frac{[F_2 - 1]}{G_1} + \frac{[F_3 - 1]}{G_1 G_2} \dots$$

where F is the overall noise figure, F_1 is the noise figure of stage 1, and G_1 is the gain of stage 1, and so on. Ideally, of course, G_1 should be set to as high a value as possible to minimize the effect of F_2 , and this therefore makes the first stage crucial.

A further development which was investigated by Green and McNeil [58] combined the transimpedance amplifier with the bootstrapping technique, also later investigated by Idrus *et al.* [59]. Bootstrapping is another effective bandwidth enhancement method in which the effective capacitance of a photodetector, C_d can be reduced to a value:

$$C_d \rightarrow C_d \cdot (1 - A)$$

where A is the gain of the stage, adjusted to be slightly less than unity. The bandwidth of the front end including bootstrapping is thus increased by the factor $1/(1 - A)$. Conceptually, the arrangement can be set up as shown in Figure 10.10:

In this case, the bootstrap function is a loop outside of the transimpedance amplifier, and the available bandwidth is many times greater than that available by either technique alone. In fact, it facilitates a gain–bandwidth trade-off overall which permits better stability in the transimpedance stage. Using the terms for gain in Figure 10, the apparent bandwidth of the whole system is thus increased by the term:

$$f_0 \rightarrow f_0 \cdot A2/(1 - A1)$$

which can be very large. An alternative configuration is also shown in Figure 10.11. In this case, the high gain of the transimpedance amplifier and its bandwidth enhancement capabilities are included in the feedback loop for bootstrapping. Care is needed to ensure that the $[1/A2]$ term compensates properly for the transimpedance gain $A2$. Even though the transimpedance stage is within the bootstrap loop, similar benefits are possible but the whole setup is more critical, therefore. In the example of Figure 10.11, the bootstrap stage is usually an emitter or source follower, where the gain can be generally set to be less than unity, so that the gain of the stage – to eliminate the possibility of instability and oscillation – is never an issue.

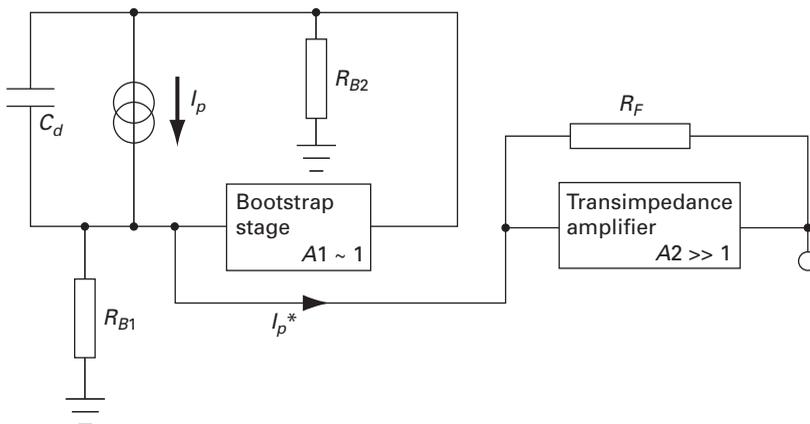


Figure 10.10 One possible bootstrapped transimpedance amplifier configuration.

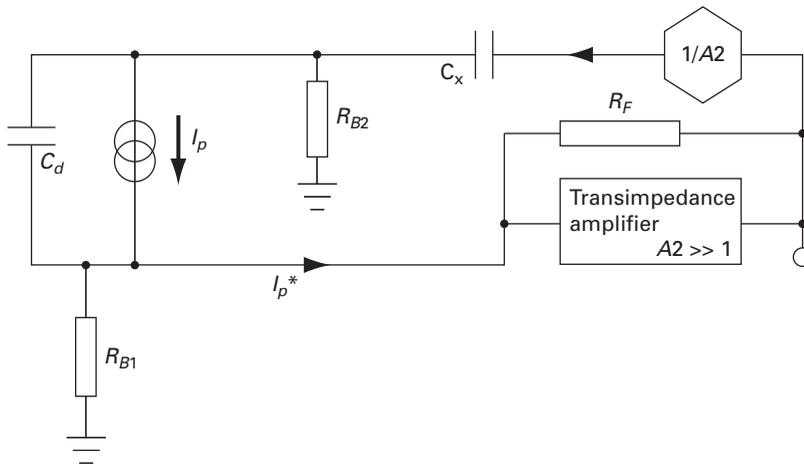


Figure 10.11 Another version of the bootstrapped transimpedance amplifier.

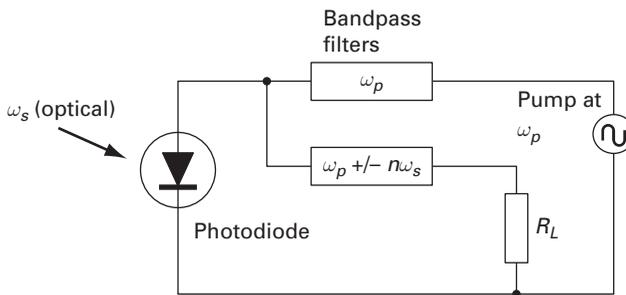


Figure 10.12 Basic PPA configuration.

10.10.2 Photoparametric amplifier (PPA)

This configuration takes advantage of the phenomenon of parametric amplification which has been exploited in RF microwave communications systems but only relatively recently in optical communications. Essentially, the parametric properties of a device are varied in sympathy with a control signal, usually called the pump, such that a signal at another frequency is amplified by energy transfer. Frequency relationships can be established, and an excellent text book on the subject of all electronic parametric amplification was produced by Howson and Smith [60]. Moving forward, the concept of parametric amplification in an optoelectronic device was then investigated by many authors, including Green and Khanifar [61]. This approach is quite radically different to the former, in that, instead of the incoming signal being electronic, it is an optical one modulated at some frequency ω_s , and optoelectronic conversion (via standard photodetection principles) and parametric amplification both take place within a photodiode, as shown in Figure 10.12:

This configuration gives gain according to both the pump frequency and also the signal frequency. The particular benefit of the parametric approach is that the mixing process gives gain at a very low noise figure – less than 3 dB in the parametric amplifier and similarly for the photoparametric version. Realizable gains of 20 dB or more have been achieved, where the gain is defined by:

$$\text{Gain} = \frac{\text{Output voltage across load with pump at frequency } (\omega_p \pm n\omega_s)}{\text{Output voltage across load without pump at frequency } \omega_s}$$

In the configuration shown the gain is theoretically given by the expression:

$$\text{Gain} = (\beta/2).(\omega_i/\omega_s)$$

where β is a term related to device parameters such as doping densities in the N and P regions (assuming a PN junction), and the idler frequency ω_i is $\omega_p \pm n\omega_s$ with n an integer. A similar expression can be used for a PIN diode except that the gain term is smaller. In practice, interestingly, it has been found that, as long as the C-V characteristic of the diode used is steep – and the diode reverse biased so as to operate in this region – then the gain term is quite near to that given above. The concept can be used to implement a very neat fiber–radio interface, as demonstrated by Idrus and Green [62]. The basic principles were further demonstrated in a paper by Leeson *et al* [63].

One of the points associated with the photoparametric amplifier as described is that the output frequency is higher than that of the input, i.e. it is an up-converter. Whereas this higher frequency is quite usable via standard RF detection techniques, it remains, in certain circumstances, inconvenient. It is for this reason that attention has moved to the double version of the configuration, involving a second mixing stage.

10.10.3 Double heterodyne photoparametric amplifier (DHPPA).

One of the key points about any communication system is that the noise performance of the system is primarily determined by the noise figure and gain of the first stage as indicated earlier using the standard Friis formula. Consequently, with a PPA as the front end, it becomes relatively feasible to configure a front end involving also a second stage of mixing in order to translate the up-converted signal from the PPA action to another frequency such as baseband, to permit a direct detection approach overall yet with more sensitivity than a basic photodiode detector alone offers, at the relatively low noise penalty given by parametric amplification. It is called the double heterodyne PPA, or DHPPA [64]. It is particularly beneficial in optical wireless configurations due to the generally adverse signal-to-noise environment found there. An example of a DHPPA is shown in Figure 10.13 for an application in a radio over fiber situation, but which can be equally used in optical wireless in which it has been evaluated recently [62].

The particular merits of the double conversion approach can be quite similar to those of a conventional double superheterodyne configuration in RF systems, as the technique in general of heterodyning is very useful for obtaining selectivity and sensitivity. Several subcarriers can be combined at the transmitter end and used collectively to modulate the output of a laser diode operating in its most linear region above threshold. The output

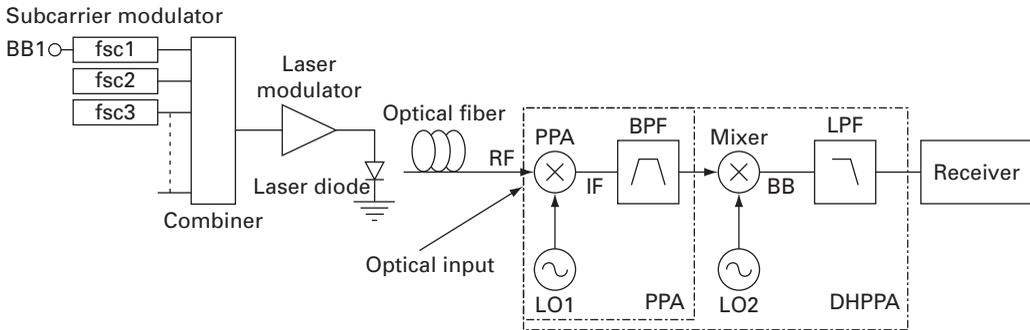


Figure 10.13 DHPPA configuration in a radio over fibre application. (After [62].)

of the laser diode can then either go via an optical fiber to some remote receiving end, such as a fiber hub or other optical fiber termination, or alternatively the output could pass through free space on an optical wireless link to an optical antenna [66] coupled to a photodetector/receiver-amplifier combination.

On reception, in either case, the photodiode is used not as a direct detector, but in the photoparametric mode to create a spectrum of diverse frequency components in the manner described above for the PPA. After this point and suitable bandpass filtering to extract the up-converted signal, the resulted higher-frequency signal passes to a second stage mixer, which has another input signal, designated LO2 in Figure 10.13. If LO2 is the same frequency as LO1 (the PPA pump frequency), then, if the output is passed through a lowpass filter, the result is a baseband signal which exactly (in theory at least) matches the original input signal in spectrum, the benefit being that the configuration is amplifying and at low noise penalty (~ 3 dB or less in theory). Additionally, changes to other frequencies than baseband can be implemented. The technique is thus quite versatile, and can be used for optical-to-RF interface systems overall, and work has gone on at Warwick University examining the options for integration within a mixed optical-RF infrastructure using RF, optical fiber- and optical wireless links [65].

10.10.4 Enhanced bandwidth receiver

One of the main issues in optical wireless is obtaining bandwidth at the front end in spite of relatively high photodiode capacitance, in comparison to values in optical fiber communications. Of course, as discussed above, this capacitance can be turned from a disadvantage to an advantage when using parametric pumping of it to achieve gain. However, there are many circumstances when bandwidth is required, especially as the effect of the capacitance is to reduce either the gain or bandwidth because of the gain-bandwidth product rule. It is interesting to note that many new ideas for circuit configurations are actually reinventions of past ones, updated to new application areas. One case in point is with regard to reducing the effect of input capacitance in a front end optical wireless receiver amplifier. In the past, video systems engineers struggled considerably to extract sufficient resolution from video cameras based on vacuum tube

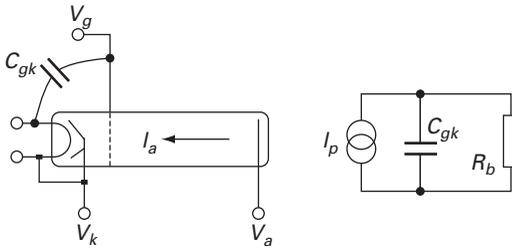


Figure 10.14 The input aspects of a vacuum tube imaging device and equivalent circuit.

devices such as the vidicon and equivalents. The principal limitations to the performance lay in the grid-cathode capacitance, as it effectively shunted the charge-, and hence, ultimately, current-generating mechanisms increasingly at higher frequencies, as shown in Figure 10.14.

In Figure 10.14, C_{gk} represents the grid-cathode capacitance, I_a the anode current, V_g the grid voltage supplied via a bias/load resistor R_b , and V_a and V_k the anode and cathode voltages respectively. I_p represents the photocurrent in the equivalent circuit. Reducing the effect of this capacitance was brought about using Percival coils (after the inventor), in which an inductance was placed in series with the input of the subsequent head amplifier in order to resonate associated capacitance and hence create a peak in the frequency response to reduce the effect of high-frequency rolloff.

Clearly, the equivalent circuit of such a vacuum tube and its head amplifier are very similar to the input stage of an optical receiver-amplifier connected to a photodiode in an optical wireless or optical fiber receiver in its simplest possible form. Therefore, we can take similar remedial action in order to extend the high-frequency performance of front ends for optical wireless. In the circuit in Figure 10.15, there is a basic in bootstrap configuration as discussed in an earlier section, in a BJT implementation:

A typical transistor for Q1 might be the Philips BFR540, which offers a high gain-bandwidth product. The 2K resistors effectively form the load. The output is taken from the emitter of Q1, and positive feedback is from the emitter to the lower side of the photodiode to provide the standard bootstrap action. The issue is concerning that which limits the bandwidth of the bootstrap action itself: Q1 has three aspects which do so, being the base-collector, base-emitter, and emitter-collector capacitances. The base-emitter capacitance is itself bootstrapped, so its effects are reduced. The emitter-collector capacitance does not restrict bandwidth so much as it shunts the relatively low output impedance of the stage, whereas the base-collector capacitance bypasses the input signal current at a high impedance point of the circuit. Clearly, anything which reduces this effect is desirable, so, using the ideas of the Percival coil, input tuning can alleviate the situation. One possible configuration for a double implementation of this [68] is shown in Figure 10.16.

The output is generally taken from the emitter of Q1. In terms of bootstrapping, this output is also used to feed a second stage using Q2 which drives the collector of Q1, thus ensuring that the base-collector voltage of Q1 remains essentially constant and thus

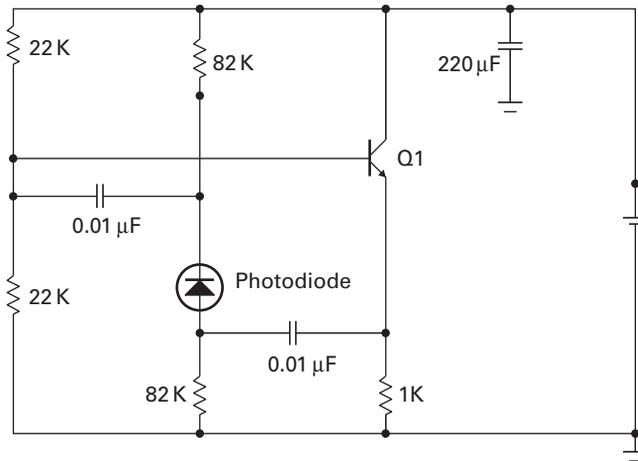


Figure 10.15 Bipolar implementation of a bootstrap amplifier, usable in an optical wireless front end. (After [68].)

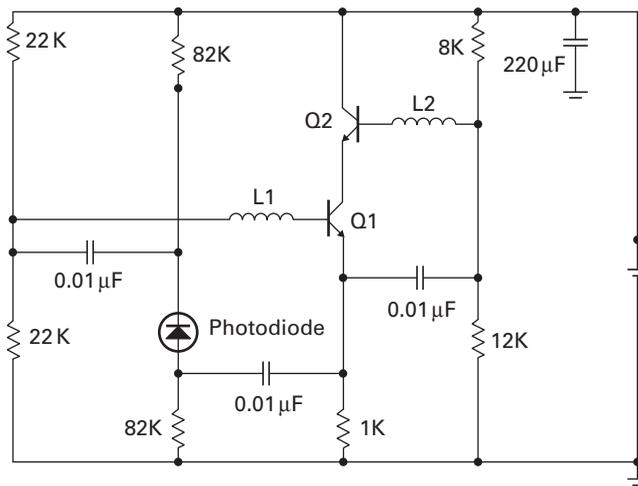


Figure 10.16 Double bootstrap configuration for enhanced high frequency performance. (After [68].)

reducing the capacitive current which would have gone to an AC ground in the form of the supply rail. The situation is further improved by the peaking coils in the base of Q1 (in a similar way to the use in the video systems of old) and also in the base of Q2, in each case to resonate with the effective input capacitances in order to extend the bandwidth further.

Note that the process of bootstrapping combined with Percival coil peaking can extend the bandwidth of a high impedance front end by factors of 10 or more, depending on the required flatness of response and also taking account of the noise performance. Under shot-noise-limited conditions, the SNR degradation caused by the

process corresponds with the increase in bandwidth for AWGN. However, this must be compared with obtaining the same bandwidth using equalization, where this time the noise power is a function of bandwidth cubed for frequencies well above the corner frequency.

10.11 Modulation schemes for optical wireless

10.11.1 Pulse position modulation pulse width modulation

Combining L-level pulse-position modulation (L-PPM) and L-level pulse-width modulation (L-PWM), this section describes two new hybrid modulation schemes. L-level Pulse-Position-Pulse-Width Modulation (PPMPWM) is one such scheme, which has a relatively low bit error rate, and improved power and bandwidth efficiencies compared to PPM [10]. The properties of the PPMPWM scheme can be demonstrated effectively through simulation to show that the scheme is effective in these respects. In early manifestations of optical wireless, OOK i.e. on-off keying was used but this was found to be much less than optimum, and more commonly L-PPM has been used, and extensively adopted by bodies such as the IrDA. L-PPM itself has a low Bit Error Rate (BER) and high power efficiency, and is a relatively simple scheme, and so is widely used in optical wireless communications. However, with low bandwidth efficiency, and as the width of its pulse can be narrow, this is not good from the point of view of required receiver performance. Other schemes exist, and a range of modulation schemes commonly used for optical wireless is shown in Figure 10.17. OOK assigns a binary code to the data directly, and is thus simple to implement. L-PPM (designated simply as PPM in the figure), assigns a pulse position within a time window. The number of slots in which the pulse can be positioned corresponds directly to the data value.

Thus, if a pulse is 1/16 of the available time window, as shown, then it can occupy any one of the 16 time slots. Demodulation therefore requires accurate time synchronization between the two ends of the communication system, in order to detect the pulse position. Another modulation scheme is also Pulse Width Modulation (PWM), commonly also used in speed control circuits. Another variation is Pulse Position Modulation Pulse Width Modulation (PPMPWM), combining the latter two schemes. All of these are illustrated in Figure 10.17 above. In the latter case, two such possibilities are shown, as now there are two independent variables available. In the first example above (sub-figure (d), parameter $r = 1$), this means that, if we consider pulse positions, there are $(r + 1)$, i.e. 2 positions as shown in the example. For 16 possible data words, then this means that, for each of these 2 positions there are 8 pulse widths, as shown. In sub-figure (e) above in Figure 10.17, then the result for $r = 3$ is also shown, so that there are 4 pulse positions and 4 pulse widths. The modulation schemes can be compared in several ways, and one way is in terms of power efficiency, as shown in Figure 10.18, where the reference signal is OOK.

Another comparison is via bandwidth efficiencies, and this is shown in Figure 10.19.

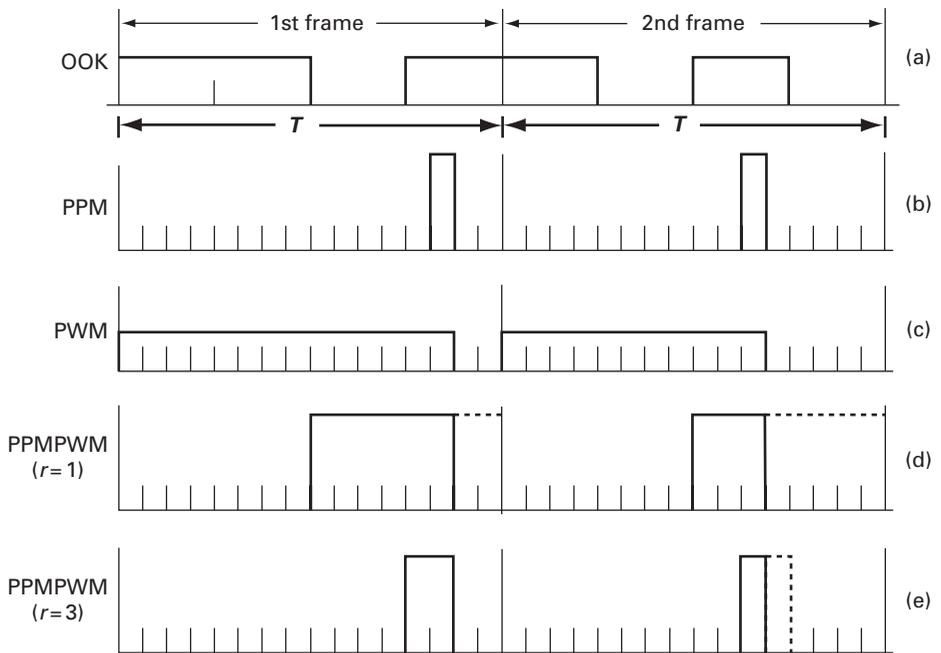


Figure 10.17 An example range of modulation schemes for optical wireless communications. (After [69].)

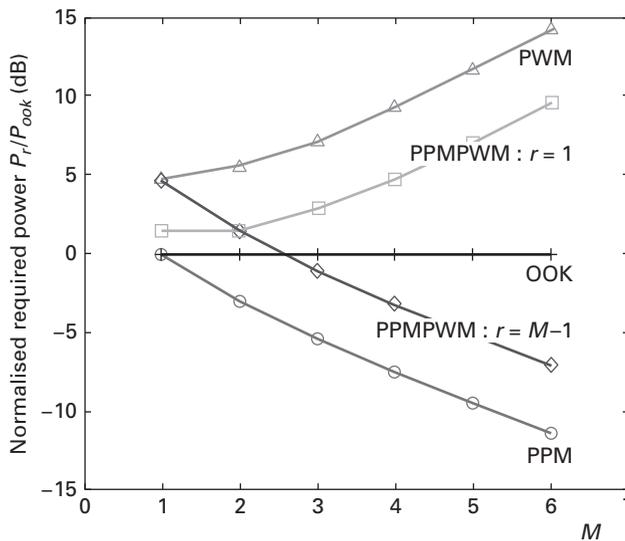


Figure 10.18 Relative power efficiencies of OOK, PWM, PPM, and PPMPWM. (After [70].)

From Figure 10.18 it is easy to see why PPM has been extensively used, especially when the relatively meager power supply of portable equipment using optical wireless (PDAs, cellphones, etc.) is considered. However, this is clearly at the expense of bandwidth efficiency, shown in Figure 10.19. The PPMPWM hybrid scheme is one

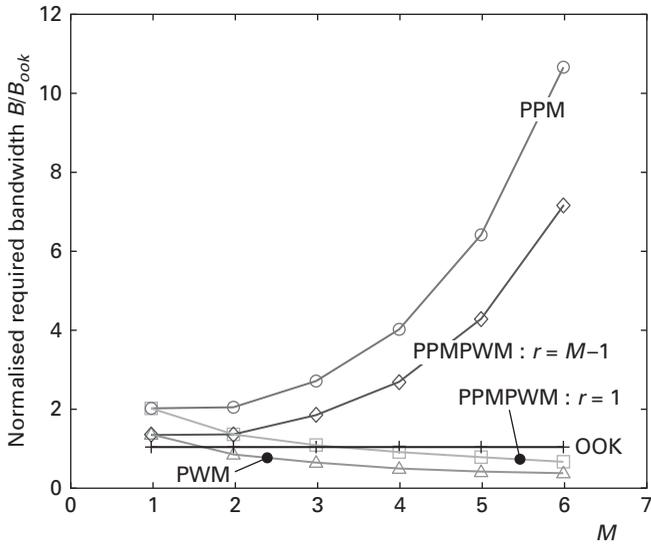


Figure 10.19 Bandwidth efficiencies of various modulation schemes. (After [70].)

option where these aspects can be traded off, and this is clearly shown. The relationships between these parameters for the schemes illustrated are detailed in reference [67], which discusses some collaborative work undertaken between Warwick University in the UK and the Northwestern Polytechnical University in Xi'an, China.

10.11.2 Pulse amplitude pulse position modulation (PAPM)

Another possibility is to consider amplitude modulation in conjunction with pulse position, and this has been studied by Zeng and Green [69]. The objective of the scheme is to combine each modulation type in order to compensate for the shortcomings in either bandwidth efficiency or power efficiency, in a parallel way as was discussed for PPMPWM. In Figure 10.20 a typical PAPM sequence of possibilities is shown.

PAPM is thus a multi-level modulation scheme. It can be expressed as **M-n-PAPM**, where **M** is the number of amplitude levels, and **n** is the pulse numbers within a clock cycle. The bandwidth and power requirements of M-n-PAPM are given, respectively, by:

$$B_{M-n-PAPM} = \frac{n}{\log_2 nM} B_{OOK}$$

and

$$P_{M-n-PAPM} = \sqrt{\frac{2M^2}{n \log_2 nM}} P_{OOK}.$$

Such a hybrid scheme may be made adaptive dynamically [70]. The merits of M-n-PAPM can be summarized graphically in Figure 10.21.

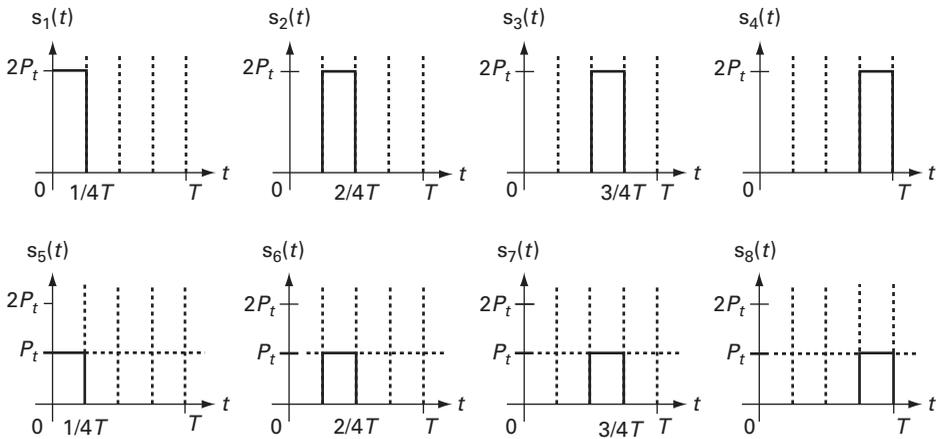


Figure 10.20 Pulse amplitude pulse position modulation options for the case of 2 amplitudes and 4 pulse positions. (2-4-PAPM, after [69].)

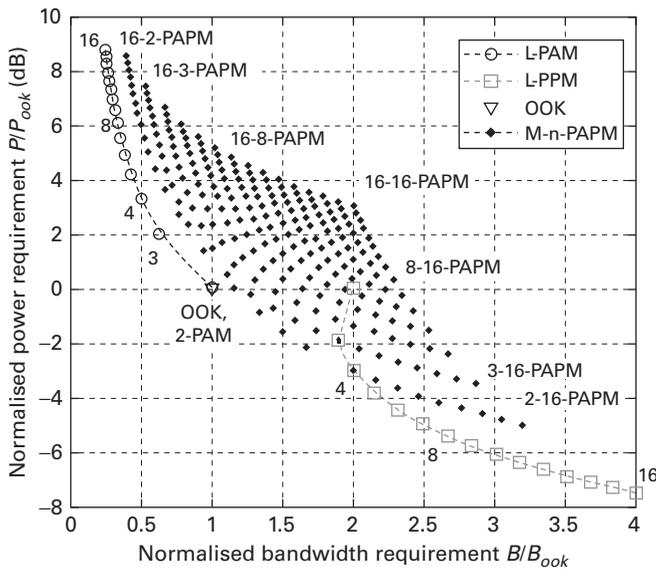


Figure 10.21 Figure 10.21: M-n-PAPM in terms of both bandwidth- and power efficiencies relative to OOK. (After [70].)

It can be seen that various parameters (i.e. M , n values) can provide less bandwidth requirements than OOK or less power requirements, but not both. However, the M-n-PAPM does fill in the gap in the characteristic above whereby one can compare PPM and PAM with OOK, as these are also plotted in Figure 10.21 for clarification. Thus, a continuum of possibilities exists to permit the system designer to use a modulation scheme which is suitable for a particular set of circumstances and requirements.

10.12 Optics for optical wireless

10.12.1 The optical antenna

Optical collection and distribution optics are crucial in optical wireless because of the adverse environment in terms of the ambient illumination and its effect on both signal masking and receiver shot noise as discussed earlier. In general, the requirements for reception are, if anything, more critical than those for transmission as they have to help the receiver optoelectronics extract as much as possible from what is usually a relatively adverse illumination environment. If the indoor channel is considered, then room lighting is usually quite high in level, and can contain a rich supply of the same range of wavelengths as are used for optical wireless communication. In Sections 10.4 and 10.5 previously, the optics and especially filter characteristics were introduced, and their importance in terms of the illumination environment, sources, and detectors. In particular it was shown how the combination of the photodiode responsivity and high pass infrared filter characteristics could be combined to give a bandpass function. The point was also made that narrow band filters are sensitive to illumination angle, so that one has to be careful, when designing the optical side of the indoor channel, that the receiver is blind to the source at angles greater than those defining the optical passband corresponding to the source (i.e LED or laser) wavelengths. Nevertheless, it is possible to use optical structures to help the channel operate effectively, and a range of concentrators or lenses can be used for this purpose. In general, the Infrared Data Association (IrDA) has defined standards for high data rate links (eg. Gigabit IR) which use hemispherical concentrators, as shown in Figure 10.22. These concentrators have the merit of very wide acceptance angle – typically up to 180° but at low concentrations, the latter being defined in general as the ratio of the input area to the output area.

The detector does not cover the whole of the base of the concentrator because, although this would mean more light was collected (rays B in Figure 10.22 would not be lost), the device capacitance would, pro rata, be higher and the receiver bandwidth

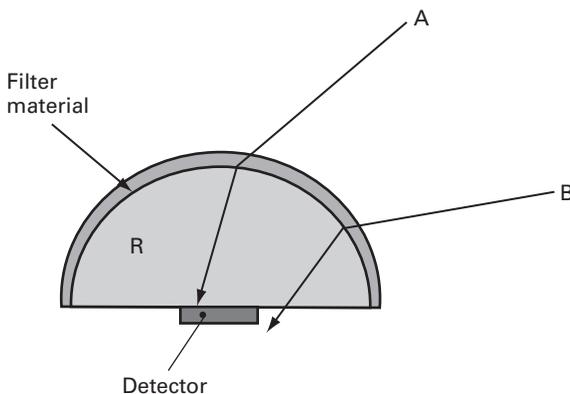


Figure 10.22 Hemispherical concentrator with filter. (After [66].)

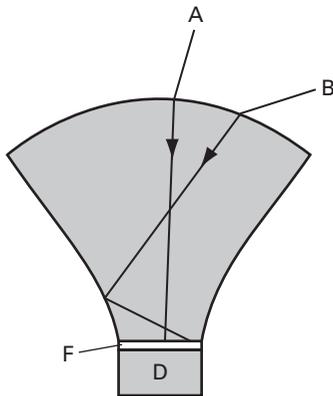


Figure 10.23 Optical antenna. (After [66].)

reduced correspondingly. The filter material provides a high degree of ambient spectrum suppression, but has to be on a curved surface, which makes it more expensive to produce. The refractive material, R , is often dyed with a visible wavelength blocking material as a cheap and cheerful filter, but a broad range of infrared can pass. A better solution to ambient wavelength suppression can be achieved by using a narrowband filter of only sufficient bandwidth to allow source (i.e. communication) wavelengths through over the angles of probable illumination. One way of doing this is demonstrated in the Optical Antenna (= DTIRC = Dielectric Totally Internally Reflecting Concentrator) shown in Figure 10.23 [66].

The essence of its action is that rays from several directions (such as from A and B) can, using the process of total internal reflection, all strike the detector D in a way not possible with a simple lens. A convex lens itself provides a very high degree of concentration, but its capture angle (the angle over which rays striking it hit the focus) is very small, which makes it only suitable for point-to-point links. However, the Optical Antenna permits a capture angle range from that of a lens (a few degrees) to $\pm 60^\circ$ typically. Therefore, the system designer can trade off capture angle for optical concentration for a given collection surface radius of curvature and arc angle. Note also that the optical filtering is done using a small but flat filter just in front of the detector but on the output lower surface of the concentrator section as shown in Figure 10.23. Clearly, one has to consider what optical bandwidth is necessary for a given application, and temper that with the acceptance angle capability of the filter itself. The narrower the filter optical bandwidth, the smaller the range of acceptance angles for the filter, so that, above some well-defined incident angle, the filter will become blind to the incident illumination – a practical trade-off as mentioned earlier.

10.13 Concluding remarks

This chapter has addressed the indoor optical wireless channel. It has examined the transmitter side, the receiver-amplifier aspects, and also the optics associated with both

ends. There has been discussion of various modulation schemes, which permit permutations of bandwidth and power efficiency, especially important in low-power handheld devices and systems. Noise has also been addressed, based on that caused by ambient light in particular, which is particularly severe in comparison to that obtained in optical fiber communications systems.

References

- [1] R. J. Green, H. Joshi, M. D. Higgins, and M. S. Leeson, Recent developments in indoor optical wireless systems. *IET Communications*, **2**:1 (2008), 3–10.
- [2] A. G. Al-Ghamdi and J. M. H. Elmirghani, Multiple spot diffusing geometries for indoor optical wireless communication systems. *International Journal of Communication Systems*, **16**:10 (2006), 909–922.
- [3] D. O'Brien, G. Parry, and P. Stavrinou, Optical hotspots speed up wireless communication. *Nature Photonics*, **1**:5 (2007), 245–247.
- [4] F. R. Gfeller and U. Bapst, Wireless in-house data communication via diffuse Infrared radiation. *Proceedings of the IEEE*, **67**:11 (1979), 1474–1486.
- [5] M. J. McCullagh and D. R. Wisely, 155 Mbit/s optical wireless link using a bootstrapped silicon APD receiver. *Electronics Letters*, **30**:5 (1994), 430–432.
- [6] S. Jivkova and M. Kavehrad, Receiver designs and channel characterization for multi-spot high-bit-rate wireless infrared communications. *IEEE Transactions on Communications*, **49**:12 (2001), 2145–2153.
- [7] A. G. Al-Ghamdi and J. M. H. Elmirghani, Spot diffusing technique and angle diversity performance for high speed indoor diffuse infra-red wireless transmission. *IEE Proceedings Optoelectronics*, **151**:1 (2004), 46–52.
- [8] A. Garcia-Zambrana and A. Puerta-Notario, Rate-adaptive modulation techniques for indoor wireless infrared links at bit rates of wide dynamic range. *Vehicular Technology Conference*, **53**:4 (2001), 2313–2317.
- [9] D. C. O'Brien, M. Katz, P. Wang *et al.* Short range optical wireless communications. *Wireless World Research Forum*, 2005.
- [10] C. D. Knutson and J. M. Brown, *IrDA Principles and Protocols: The IrDA Library, Vol. 1* (Salem, Utah: MCL Press, 2004).
- [11] R. Becher, M. Dillinger, M. Haardt, and W. Mohr, Broadband wireless access and future communication networks. *Proceedings of the IEEE*, **89**:1 (2001), 58–75.
- [12] J. M. Kahn and J. R. Barry, Wireless infrared communications. *Proceedings of the IEEE*, **85**:2 (1997), 265–298.
- [13] J. M. Kahn, R. You, P. Djahani, A. G. Weisbin, and A. Beh Kian Teik Tang, Imaging diversity receivers for high-speed infrared wireless communication. *IEEE Communications Magazine*, **36**:12 (1998), 88–94.
- [14] S. Hranilovic, *Wireless Optical Communication Systems*. (New York: Springer, 2005).
- [15] S. Bloom, E. Korevaar, J. Schuster, and H. Willebrand, Understanding the performance of free-space optics. *Journal of Optical Networking*, **2**:6 (2003), 178–200.
- [16] British Standards Institute, Safety of laser products, equipment classification, requirements and user's guide, *BS EN 60825-1:2007* (London: BSI, 2007).
- [17] American National Standard for the Safe Use of Lasers, *ANSI Z39.1* (New York: ANS, 1986).
- [18] G. P. Agrawal, *Fiber-Optic Communication Systems*, 3rd edn. (New York: Wiley, 2002).

- [19] N. P. Schmitt, T. Pistner, C. Vassipoulis *et al.*, Diffuse wireless optical link for aircraft intracabin passenger communication. *Communications Systems, Networks and Digital Signal Processing Conference*, **5** (2006), 625–628.
- [20] D. Heatley, D. R. Wisely, and P. Cochrane, Optical wireless: The story so far. *IEEE Communications Magazine*, **36**:12 (1998), 72–82.
- [21] D. E. Aspnes, Properties of silicon. In *Emis Datareviews Series*, vol. 4. (London: INSPEC, Institution of Electrical Engineers, 1988), pp. 59–80.
- [22] www.schott.com
- [23] J. G. Proakis, *Wiley Encyclopedia of Telecommunications* (Hoboken, N.J: Wiley, 2003).
- [24] J. B. Carruthers and J. M. Kahn, Multiple-subcarrier modulation for nondirected wireless infrared communication. *IEEE Journal on Selected Areas in Communications*, **14**:3 (1996), 538–546.
- [25] D. C. M. Lee, J. M. Kahn, and M. D. Audeh, Trellis-coded pulse-position modulation for indoor wireless infrared communications. *IEEE Transactions on Communications*, **45**:9 (1997), 1080–1087.
- [26] J. M. Kahn, W. J. Krause, and J. B. Carruthers, Experimental characterization of nondirected indoor infrared channels. *IEEE Transactions on Communications*, **43**:2-4 (1995), 1613–1623.
- [27] Z. Ghassemlooy and A. R. Hayes, Digital pulse interval modulation for IR communication systems – a review. *International Journal of Communication Systems*, **13**:7-8 (2000), 519–536.
- [28] K. Park, Convolutional coded pulse-position modulation on wireless infrared link. *IEE Proceedings Optoelectronics*, **148**:4 (2001), 199–203.
- [29] A. C. Boucouvalas, Indoor ambient light noise and its effect on wireless optical links. *IEE Proceedings Optoelectronics*, **143**:6 (1996), 334–338.
- [30] R. Narasimhan, M. D. Audeh, and J. M. Kahn, Effect of electronic-ballast fluorescent lighting on wireless infrared links. *IEE Proceedings Optoelectronics*, **143**:6 (1996), 347–354.
- [31] A. J. C. Moreira, R. T. Valadas, and A. M. de Oliveira Duarte, Performance of infrared transmission systems under ambient light interference. *IEE Proceedings Optoelectronics*, **143**:6 (1996) 339–346.
- [32] A. J. C. Moreira, R. T. Valadas, and A. M. de Oliveira Duarte, Characterisation and modelling of artificial light interference in optical wireless communication systems. *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. **6**:1 (1995), 326–331.
- [33] K. K. Wong, T. O’Farrell, and M. Kiatweerasakul, The performance of optical wireless link, 2-ppm and spread spectrum under the effects of multipath dispersion and artificial light interference. *International Journal of Communication Systems*, **13**:7–8 (2000), 551–576.
- [34] F. E. Nicodemus, Radiance. *American Journal of Physics*, **31**:5 (1963), 368–377.
- [35] J. R. Meyer-Arendt, Radiometry and photometry: Units and conversion factors. *Applied Optics*, **7**:10 (1968), 2081–2084.
- [36] A. J. C. Moreira, R. T. Valadas, and A. M. de Oliveira Duarte, Optical interference produced by artificial light. *Wireless Networks*, **3**:2 (1997), 131–140.
- [37] R. T. Valadas, A. R. Tavares, and A. M. de Oliveira Duarte, Angle diversity to combat the ambient noise in indoor optical wireless communication systems. *International Journal of Wireless Information Networks*, **4**:4 (1997), 275–288.

-
- [38] A. G. Al-Ghamdi and J. M. H. Elmirghani, Optimization of a triangular PFDR antenna in a fully diffuse OW system influenced by background noise and multipath propagation. *IEEE Transactions on Communications*, **51**:12 (2003), 2103–2114.
- [39] European Union, Directive 2005/32/EC of the European Union. *Official Journal of the European Union*, **48**: L191 (2005), 29–58.
- [40] M. D. Higgins, Genetic algorithm optimization methods applied to the indoor optical wireless communications channel, PhD Thesis, University of Warwick (2009) [wrap.warwick.ac.uk/2757/].
- [41] V. Jungnickel, V. Pohl, S. Nonnig, and C. von Helmolt, A physical model of the wireless infrared communication channel. *IEEE Journal on Selected Areas in Communications*, **20**:3 (2002), 631–640.
- [42] M. R. Pakravan and M. Kavehrad, Design considerations for broadband indoor infrared wireless communication systems. *International Journal of Wireless Information Networks*, **2**:4 (1995), 223–238.
- [43] H. Hashemi, G. Yun, M. Kavehrad, F. Behbahani, and P. A. Galko, Indoor propagation measurements at infrared frequencies for wireless local area networks applications. *IEEE Transactions on Vehicular Technology*, **43**:3 (1994), 562–576.
- [44] M. R. Pakravan and M. Kavehrad, Indoor wireless infrared channel characterization by measurements. *IEEE Transactions on Vehicular Technology*, **50**:4 (2001), 1053–1073.
- [45] C. R. Lomba, R. T. Valadas, and A. M. de Oliveira Duarte, Efficient simulation of the impulse response of the indoor wireless optical channel. *International Journal of Communication Systems*, **13**:7-8 (2000), 537–549.
- [46] J. B. Carruthers and P. Kannan, Iterative site-based modeling for wireless Infrared channels. *IEEE Transactions on Antennas and Propagation*, **50**:5 (2002), 759–765.
- [47] J. R. Barry, J. M. Kahn, W. J. Krause, E. A. Lee, and D. G. Messerschmitt, Simulation of multipath impulse response for indoor wireless optical channels, *IEEE Journal on Selected Areas in Communications*, **11**:3 (1993), 367–379.
- [48] C. M. Goral, K. E. Torrance, D. P. Greenberg, and B. Battaile, Modeling the interaction of light between diffuse surfaces. *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, **11** (1984), 213–222.
- [49] F. J. López-Hernández and M. J. Betancor, Dustin: algorithm for calculation of impulse response on IR wireless indoor channels. *Electronics Letters*, **33**:21 (1997), 1804–1806.
- [50] R. Pérez-Jiménez, J. Berges, and M. J. Betancor, Statistical model for the impulse response on infrared indoor diffuse channels. *Electronics Letters*, **33**:15 (1997), 1298–1300.
- [51] F. J. López-Hernández, R. Pérez-Jiménez, and A. Santamaría, Ray-tracing algorithms for fast calculation of the channel impulse response on diffuse IR wireless indoor channels. *Journal of Optical Engineering*, **39**:10 (2000), 2775–2780.
- [52] V. Pohl, V. Jungnickel, and C. von Helmolt, Integrating-sphere diffuser for wireless infrared communication. *IEE Proceedings Optoelectronics*, **147**:4 (2000), 281–285.
- [53] M. D. Higgins, R. J. Green, and M. S. Leeson, A genetic algorithm method for optical wireless channel control. *Journal of Lightwave Technology*, **27**:6 (2009), 760–772.
- [54] C. R. Lomba, R. T. Valadas, and A. M. de Oliveira Duarte, Experimental characterization and modeling of the reflection of infrared signals on indoor surfaces. *IEE Proceedings Optoelectronics*, **145**:3 (1998), 191–197.
- [55] E. Uiga, *Optoelectronics* (London : Prentice Hall, 1995).
- [56] B. T. Phong, Illumination for computer generated pictures. *Communications of the ACM*, **18**:6 (1975), 311–317.

- [57] S. S. Haykin and B. V. Veen, *Signals and Systems* (New York: Wiley, 2003).
- [58] R. J. Green and M. G. McNeill, The bootstrap transimpedance amplifier – a new configuration, *IEE Proc. on Circuits & Systems*, **136:2**, (1989), 57–61.
- [59] S. M. Idrus, N. Ngajikin, N. N. N. A. Malik, and S. I. A. Aziz, Performance analysis of bootstrap transimpedance amplifier for large windows optical wireless receiver, *International RF and Microwave Conference Proceedings*, (2006), pp. 416–420.
- [60] D. P. Howson and R. B. Smith, *Parametric Amplifiers*, (London: McGraw-Hill, 1970).
- [61] R. J. Green and A. Khanifar, Photoparametric amplifier/converter in subcarrier-multiplexed lightwave communications systems, *Proceedings of the IEEE International Microwave Symposium MTT-S*, Albuquerque, New Mexico, (1992).
- [62] S. M. Idrus and R. J. Green, Photoparametric up-converter for millimeter-wave fibre-radio system, *Int. IrDA/IEEE/IEEE Seminar*, Warwick University, (2003), pp. 11–16.
- [63] M. S. Leeson, R. J. Green, and M. D. Higgins, Photoparametric amplifier frequency converters. *Proceedings of ICTON 08*, session We.D4.1, Athens, (2008), pp. 197–200. (*invited*).
- [64] S. M. Idrus, Theoretical and practical performance of a photoparametric optical wireless receiver, Ph.D. thesis, University of Warwick, (2004).
- [65] H. Alhaghagi, R. J. Green, and E. L. Hines, Double heterodyne photoparametric amplification techniques for optical wireless communications and sensing applications, *Proceedings of ICTON 2010*, (2010), Munich.
- [66] R. Ramirez-Iniguez and R. J. Green, Optical antenna design for indoor optical wireless communication systems, *Special Issue of the International Journal of Communication Systems – Indoor Wireless Communication Systems and Networks*, **18**, (2005), 229–245.
- [67] F. Yangyu, B. Bo, R. J. Green, PPMPWM: a new modulation format for wireless optical communications, *Proceedings of ICTON 2010*, (2010), Munich.
- [68] R. J. Green, H. Joshi, M. D. Higgins, and M. S. Leeson, Bandwidth extension in optical wireless receiver-amplifiers, *Proceedings of ICTON 08*, session We.D4.2, (2008), pp. 201–204, Athens.
- [69] Y. Zeng, R. J. Green, and M. S. Leeson, Multiple pulse amplitude and position modulation for the optical wireless channel, *Proceedings of ICTON 08*, session We.C4.4, (2008), pp. 193–196, Athens.
- [70] Y. Zeng, Adaptive modulation schemes for optical wireless communication systems, Ph.D. thesis, University of Warwick, (2010).

11 Hybrid RF/FSO communications

Nick Letzepis and Albert Guillén i Fàbregas

11.1 Introduction

In free-space optical (FSO) communication an optical carrier is employed to convey information wirelessly. FSO systems have the potential to provide fiber-like data rates with the advantages of quick deployment times, high security, and no frequency regulations. Unfortunately such links are highly susceptible to atmospheric effects. *Scintillation* induced by atmospheric turbulence causes random fluctuations in the received irradiance of the optical laser beam [1]. Numerous studies have shown that performance degradation caused by scintillation can be significantly reduced through the use of multiple-lasers and multiple-apertures, creating the well-known multiple-input multiple-output (MIMO) channel (see e.g. [2], [3], [4], [5], [6], [7], [8], [9], [10]). However, it is the large attenuating effects of cloud and fog that pose the most formidable challenge. Extreme low-visibility fog can cause signal attenuation on the order of hundreds of decibels per kilometer [11]. One method to improve the reliability in these circumstances is to introduce a radio frequency (RF) link to create a *hybrid FSO/RF* communication system [12], [13], [14], [15], [16], [11]. When the FSO link is blocked by cloud or fog, the RF link maintains reliable communications, albeit at a reduced data rate. Typically a millimeter wavelength carrier is selected for the RF link to achieve data rates comparable to that of the FSO link. At these wavelengths, the RF link is also subject to atmospheric effects, including rain and scintillation [17], [18], [19], [20], [21], but less affected by fog. The two channels are therefore complementary: the FSO signal is severely attenuated by fog, whereas the RF signal is not; and the RF signal is severely attenuated by rain, whereas the FSO is not. Both, however, are affected by scintillation.

Most works on the hybrid channel [12], [13], [14], [16], [11] consider the RF and FSO links as separate channels, i.e. the channels do not aid each other to compensate signal level fluctuations. In these works, the main purpose of the RF link is to act as a backup when the FSO link is down. In [15] a hybrid channel coding scheme is proposed that combines both FSO and RF channels and adapts the code rate to the channel conditions. Multilevel coding schemes have also been proposed in [22] for the hybrid channel.

Lacking so far in the literature on hybrid FSO/RF channels is the development of a suitable channel model and its theoretical analysis to determine the fundamental limits

of communication. This is the central motivation of this chapter. We propose a channel model for hybrid FSO/RF communications based on the well-known parallel channel [23], that takes into account the differences in signalling rate, and the atmospheric fading effects present in both the FSO and RF links. These fading effects are slow compared to typical data rates, and as such, each channel is based on a *block-fading* channel model. Previously in [2], [3], using a block-fading channel model, we examined the *outage probability* of the multiple-input multiple-output (MIMO) FSO channel under the assumption of pulse position modulation (PPM) for several well-known scintillation distributions, i.e. lognormal, exponential, gamma-gamma, lognormal-Rice, and I-K distributed scintillation [1]. In particular, we examined the *outage exponent* [24], [25], which describes the slope of the outage probability curve in the asymptotic high power regime. In this chapter, we extend this analysis to include an RF link to create a hybrid FSO/RF channel. The message to be transmitted is encoded into parallel FSO and RF bit streams which are sent across the FSO and RF channels simultaneously. We examine the case when perfect CSI is known at the receiver only (CSIR case), then we consider the case when CSI is also known at the transmitter (CSIT case), and power allocation is employed to reduce the outage probability subject to power constraints. When CSI is not available at the transmitter, we derive the outage exponents of the hybrid channels for general scintillation distributions in each of the channels. On the other hand, when CSI is available at the transmitter, we derive the optimal power allocation algorithm subject to both (individual) peak and (total) average power constraints. For FSO channels with fixed equiprobable input distributions, the optimal solution involves non-convex optimization, which has prohibitive complexity for implementation in practical systems. To remedy this, in [26], we proposed a low complexity, suboptimal solution. In this chapter we consider an alternative approach. Rather than enforcing equiprobable signalling, we consider two alternative signalling strategies: time-sharing; and signalling with the optimal input distribution. For these strategies, the power allocation problem becomes convex, and hence yields a tractable solution. Moreover, we further calculate the outage exponents with CSIT, showing that for any finite peak-to-average power ratios, the exponent is given by the same exponent as the no-CSIT case. However, the larger the peak-power constraints are, the more reliable transmission becomes, eventually removing outages altogether.

The remainder of this chapter is organized as follows. In Section 11.2 we present our channel model and assumptions. Section 11.3 reviews the required information-theoretic preliminaries. Section 11.4 presents our main results for the CSIR-only case while Section 11.5 discusses power allocation and outage exponents for the CSIT case. Section 11.6 draws final concluding remarks. Proofs of our results can be found in the appendices.

In this chapter random variables and their realizations are denoted by upper and lower case letters respectively, e.g. if X is denotes a random variable, then a random sample drawn (or realisation) from X is denoted by x . The probability density and cumulative distribution of the random variable X is denoted by f_X and F_X respectively. Conditional densities are denoted by $f_{Y|X}(y|x)$, and where necessary, to highlight their importance, we explicitly list certain parameters of a density following a semicolon, e.g. the density

$f_X(x; \theta)$ is parameterized by θ . The expectation operator of a random variable is denoted by \mathbb{E} . These notations carry over to random vectors in the usual way, where vectors are denoted by bold-face characters.

11.2 Channel model

The hybrid communication system considered in this chapter is illustrated in Figure 11.1. First, a binary message sequence is jointly encoded into parallel FSO and RF bit streams. Each bit stream is modulated according to a discrete modulation scheme, up-converted to its respective carrier frequency, and transmitted simultaneously through an atmospheric channel to the hybrid receiver. The hybrid receiver simultaneously down-converts each carrier signal, demodulates, and jointly decodes the received sequences to recover the original message.

The development of a suitable model that captures the essential ingredients of the hybrid system is crucial before we can embark on its information-theoretic study. In practice, the FSO link operates with frequencies in the Tera-Hertz (THz) range, whereas RF link would typically operate at millimeter wavelengths, i.e. the Giga-Hertz (GHz) range. As a consequence of these vastly different carrier frequencies, there are several important physical features that must be captured, which are discussed in the sequel.

11.2.1 Hybrid signalling

Since the FSO carrier frequency is many times higher than the RF, it can therefore support higher signalling rates. This feature is captured through the definition of a *hybrid symbol*, which is composed of N FSO symbols and a single RF symbol in parallel. More formally, we define a *hybrid constellation* as follows.

DEFINITION 11.1 (Hybrid constellation) Let \mathcal{X}_{rf} and \mathcal{X}_{fso} denote discrete RF and FSO constellations respectively, where $|\mathcal{X}_{\text{rf}}| = M = 2^m$ and $|\mathcal{X}_{\text{fso}}| = Q = 2^q$. The hybrid symbol constellation is defined as

$$\mathcal{X} \triangleq \mathcal{X}_{\text{fso}}^N \times \mathcal{X}_{\text{rf}}, \quad (11.1)$$

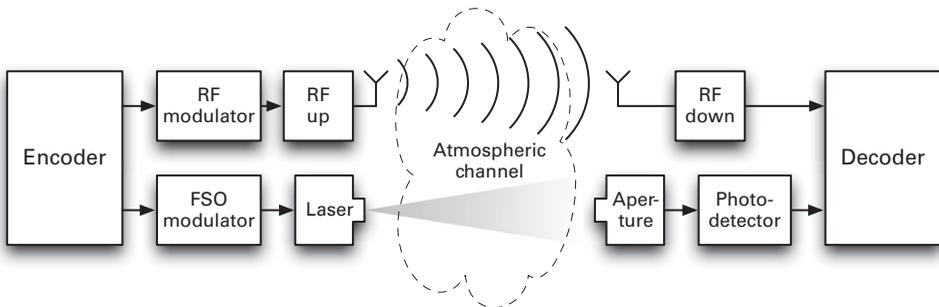


Figure 11.1 Hybrid FSO/RF communication system.

where $N \geq 1$ is a positive integer. The total size of the hybrid constellation is therefore $|\mathcal{X}| = 2^{Nq+m}$. The ratio between the number of FSO bits and the total number of bits is defined as

$$\delta \triangleq \frac{Nq}{Nq+m}. \quad (11.2)$$

Practical RF systems commonly modulate information using both the amplitude and phase of the carrier. At the receiver, a heterodyne operation is employed to down-convert to baseband and recover the in-phase and quadrature components [27]. Hence for the RF link we assume $\mathcal{X}_{\text{rf}} \subset \mathbb{C}$ is some arbitrary discrete complex constellation, e.g. Phase-Shift-Keying (PSK) or Quadrature Amplitude Modulation (QAM), and symbols are drawn independently and identically distributed (i.i.d.) and equi-probable from this constellation. In addition, \mathcal{X}_{rf} is normalized to have unit average energy, i.e. $\frac{1}{M} \sum_{\hat{x} \in \mathcal{X}_{\text{rf}}} |\hat{x}|^2 = 1$.

Whilst it is also possible to modulate the amplitude and phase of the optical carrier, this requires the corresponding heterodyne/homodyne operation to be performed optically before photodetection [28]. This requires complicated receive optics to implement the optical heterodyne as well as to maintain coherence with the optical carrier. Most practical FSO links employ the much simpler intensity modulation and direct detection (IM/DD), i.e. only the power of the optical carrier is modulated. For simplicity we assume IM/DD such that $\mathcal{X}_{\text{fso}} \subset \mathbb{R}_+^J$, i.e. the FSO symbol is divided into J time slots, where in each time slot, a specified amount of optical energy is transmitted. This general constellation scheme covers the special cases, for example, $J = 1$, Q -ary Pulse Amplitude Modulation (PAM) or $J = Q$, Q -ary Pulse Position Modulation (PPM). As in the RF link, we assume the FSO symbols are drawn i.i.d. from \mathcal{X}_{fso} . However, unlike the RF link, as we shall see later, optimizing the input constellation symbol probabilities will be beneficial both in reducing the complexity of the power allocation strategy, and improving outage performance. For all $\mathbf{x} = (x[1], \dots, x[J]) \in \mathcal{X}_{\text{fso}}$, let $f_{\mathbf{X}}(\mathbf{x})$ denote the discrete input probability distribution. We assume the constellation is normalized to have unit energy, i.e. $\sum_{\mathbf{x} \in \mathcal{X}} f_{\mathbf{X}}(\mathbf{x}) \sum_{j=1}^J x[j] = 1$.

11.2.2 Hybrid channel model

There are several important differences in how the atmosphere will affect the two carriers. Both carriers are affected by (scintillation induced) fading, but with different coherence times. Typically the optical carrier experiences much shorter coherence times (10s of msec) compared to the RF carrier (100s msec) [1], [17], [18], [19]. In addition, each carrier is affected by particulate matter comparable to their respective wavelength. Optical wavelengths are adversely affected by fog/cloud and aerosol particles [1]. On the other hand, RF millimeter wavelengths are predominantly affected by rain droplets [17], [18], [19].

It is important to note that in both the FSO and RF channels, the aforementioned fading effects are slow compared to their typical signalling rates. In the information-theoretic literature, slow-fading channels are simplistically modeled by the *block-fading*

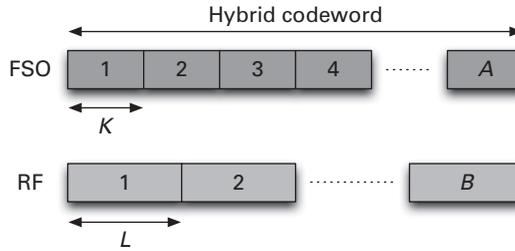


Figure 11.2 Hybrid codeword.

channel [29], [30]. In this setting, the transmitted codeword is divided into a finite number of blocks each consisting of an equal number of channel uses (or symbols). The symbols within a block experience the same fading realization, which is drawn i.i.d. from block to block. Given the slow-varying nature of scintillation, the block-fading model is appropriate for both FSO and RF channels, i.e. we model the overall hybrid channel as a parallel block-fading channel. However, within this framework, we must still model the difference in coherence time between the two channels. To capture this important effect, as shown in Figure 11.2, we assume the FSO and RF components of the hybrid codeword experiences A and B i.i.d. fading realizations respectively. Typically $A \geq B$, since the fading on the FSO link is usually faster than that of the RF link. We assume the FSO component of K hybrid symbols are affected by the same FSO fading realization. Similarly the RF component of L hybrid symbols is affected by the same RF fading realization. Note that since the total number of hybrid symbols must be consistent between the two channels, then we have the requirement $AK = BL$. With these assumptions, we write the parallel FSO and RF block-fading channels as follows.

For the FSO link, we write the received optical irradiance signal as

$$\mathbf{r}_a[k, n] = \rho h_a \frac{p_a}{N} \mathbf{x}_a[k, n] + \lambda, \quad (11.3)$$

where: $\mathbf{r}_a[k, n] = (r_a[k, n, 1], \dots, r_a[k, n, J]) \in \mathbb{R}_+^J$ denotes a J vector representing the received optical energy in each of the J time slots of FSO symbol $n = 1, \dots, N$ of hybrid symbol $k = 1, \dots, K$ in block $a = 1, \dots, A$; $\mathbf{x}_a[k, n] = (x_a[k, n, 1], \dots, x_a[k, n, J]) \in \mathcal{X}_{\text{fso}}$ denotes the n th transmitted FSO symbol vector of hybrid symbol k in block a ; ρ models long-term fading effects; $h_a > 0$ denotes the fading in block a due to scintillation, assumed to be drawn i.i.d. from some arbitrary density $f_H(h)$ with nonnegative support; p_a denotes the power allocated to block a , and $\lambda \geq 0$ models background irradiation/dark current (assumed to be constant). In (11.3) the $1/N$ factor ensures the FSO and RF components of the hybrid symbol have the same energy.

For the RF channel we similarly write the l th received symbol of block b as

$$\hat{y}_b[l] = \sqrt{\hat{\rho} \hat{h}_b p_b} \hat{x}_b[l] + \hat{\sigma} \hat{z}_b[l], \quad (11.4)$$

for $l = 1, \dots, L$ and $b = 1, \dots, B$, where: $\hat{x}_b[l] \in \mathcal{X}_{\text{rf}}$ denotes the transmitted symbol; $\hat{\rho}$ models long-term fading effects; $h_b > 0$ denotes the scintillation of block b , drawn from

some arbitrary density $f_{\hat{h}}(\hat{h})$ with nonnegative support; \hat{p}_b denotes the power allocated to block b ; $\hat{z}_b[l] \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ models zero mean, unit variance complex additive Gaussian white noise (AWGN); and $\hat{\sigma}^2$ denotes the RF noise power. We assume h_a and \hat{h}_b are independent. Note that, as opposed to what happens in the RF channel, in the FSO channel, the received signal scales with p_a rather than $\sqrt{p_a}$.

The parameters $0 < \rho, \hat{\rho} < 1$ in (11.3) and (11.4) model differences in the relative strengths of the two parallel channels, e.g. it reflects long-term fading effects due to rain, fog, or cloud as well as other parameters such as aperture/antenna gains and propagation loss. When $\rho > \hat{\rho}$, the FSO channel is much stronger than the RF, e.g. modelling the effects of severe rain attenuation. On the other hand, if $\rho < \hat{\rho}$, then the RF channel is stronger than the FSO channel, e.g. modelling the effects of severe fog/cloud attenuation. Although in practice ρ and $\hat{\rho}$ are randomly varying with time (and are also most likely correlated random variables), we assume they remain unchanged over many code-word time intervals and therefore are fixed constants. Under this assumption, as we shall see later, these parameters will not affect the asymptotic outage analysis that is to follow.

11.2.3 Channel state information and power adaptation

In wireless optical communications, efficient power adaptation not only extends the battery lifetime in portable applications, but also reduces stress on the semiconductor laser device, extending its lifetime as well. However, in order to adapt the power, we first require knowledge of the fading at the transmitter, commonly referred to as the *channel state information* (CSI). Since the fading is slow, the CSI can be estimated at the receiver and conveyed to the transmitter via a feedback link.

In this chapter, we consider two channel state information (CSI) scenarios. First we will assume only the receiver has perfect CSI and the transmitter allocates power uniformly across all blocks (CSIR case). Then we will consider the case where perfect CSI is also known at the transmitter (CSIT case), i.e. the transmitter has perfect knowledge of h_1, \dots, h_A and $\hat{h}_1, \dots, \hat{h}_B$. The transmitter then performs power allocation to reduce the outage probability subject to power constraints. It is important to note that this assumption is impossible in practice, as the transmitter will only have access to past estimates of the fading (causal CSI), which may also contain imperfections due to estimation errors at the receiver (imperfect CSI) [31], [32]. Nevertheless, the assumption of perfect CSI at the transmitter serves as an important performance benchmark for practical systems.

The transmit power for both channels is ultimately drawn from the same power resource. As such, we assume the long-term average power consumed by the hybrid system is constrained according to

$$\mathbb{E}[\langle \mathbf{p} \rangle] + \mathbb{E}[\langle \hat{\mathbf{p}} \rangle] \leq P_{\text{av}}, \quad (11.5)$$

where $\langle \mathbf{p} \rangle = \frac{1}{A} \sum_{a=1}^A p_a$ and $\langle \hat{\mathbf{p}} \rangle = \frac{1}{B} \sum_{b=1}^B \hat{p}_b$. Practical communication systems have limitations on the short-term power that can be transmitted. Since the FSO and RF links will realistically have different short-term (or peak) power limitations, we assume each is subject to its own individual short-term power constraint, i.e.

$$\langle \mathbf{p} \rangle \leq \alpha_{\text{fso}} P_{\text{av}} \quad \text{and} \quad \langle \hat{\mathbf{p}} \rangle \leq \alpha_{\text{rf}} P_{\text{av}}, \quad (11.6)$$

where α_{fso} and α_{rf} denote the peak-to-average power ratios (PAPRs) for the FSO and RF channels respectively.

11.2.4 Photodetection models

The FSO channel model (11.3) only describes the received optical energy in each of the J time slots of an FSO symbol interval. In order to jointly decode the FSO and RF signals, the optical signal must be converted to an electrical one via photodetection. However, the photodetection process itself introduces many additional noise effects uncommon in RF systems. In the optical communications literature, a plethora of photodetection models have been considered, each dependent on certain photodetector device and channel model assumptions. For a detailed review and comparison of these models the interested reader is referred to [28], [33], [34]. As we shall see later, whilst the photodetection model may affect the power adaptation scheme, it does not affect the asymptotic slope of the outage probability curve (i.e. the outage exponent). However, for the sake of comparison, in the following we summarize some of the commonly employed photodetection models that will be considered in this chapter.

In this section, for ease of exposition and brevity of notation, we drop indices a, k, n in (11.3) and consider non-fading FSO channels of the form

$$\mathbf{r} = \gamma \mathbf{x} + \lambda, \quad (11.7)$$

where $\mathbf{r} = (r[1], \dots, r[J])$ and $\mathbf{x} = (x[1], \dots, x[J]) \in \mathcal{X}_{\text{fso}}$ and γ denotes the transmitted optical power. Conversion back to (11.3) is straightforward by letting $\gamma = \rho h_a p_a / N$ and replacing the appropriate indices a, k, n .

Given the received irradiance vector \mathbf{r} , the photodetector converts it into an electrical signal vector $\mathbf{y} = (y[1], \dots, y[J])$. In this chapter, we will mostly be concerned with the channel from the transmitted signal vector \mathbf{x} to the photodetector output vector \mathbf{y} . In particular, we consider channels for which the transition probability can be written as

$$f_{Y|X}(\mathbf{y}|\mathbf{x}; \gamma, \lambda) = \prod_{j=1}^J f_{Y|X}(y[j]|x[j]; \gamma, \lambda). \quad (11.8)$$

Note that we have made explicit the dependence of the channel transition probabilities on the transmitted optical power γ and background irradiation λ . This will be useful later on, when fading is introduced.

Under ideal photodetection, thermal noise is negligible and the photodetected electrical signal represents discrete photon counts. The photon count over a given time interval is a Poisson random variable whose rate is proportional to the average irradiance over the time interval [28]. Let $y[j]$ denote the photodetector output induced by the received optical energy $r[j]$ as given by (11.7). For the ideal photodetection case, the conditional density of $y[j]$ is the discrete Poisson distribution given by

$$f_{Y|X}(y[j]|x[j]; \gamma, \lambda) = \frac{(\gamma x[j] + \lambda)^{y[j]}}{y[j]!} e^{-(\gamma x[j] + \lambda)} \quad (11.9)$$

for $i = 1, \dots, Q$. This is the well-known Poisson channel, used in a number of works on FSO communications (see e.g. [10], [7], [35]).

As the average irradiance increases due to increased background irradiation or high optical signal power, these statistics can be well approximated by Gaussian statistics [36], [37]. In Appendix A we show that the Kullback–Leibler divergence [23] between the Poisson distribution and a Gaussian distribution with the same mean and variance approaches zero as the mean increases. This implies that, provided λ is sufficiently large, $y[j] \sim \mathcal{N}(\gamma x[j] + \lambda, \gamma x[j] + \lambda)$. Hence we may approximate (11.9) as

$$f_{Y|X}(y[j]|x[j]; \gamma, \lambda) \approx \frac{1}{\sqrt{2\pi(\gamma x[j] + \lambda)}} \exp\left(-\frac{(y[j] - (\gamma x[j] + \lambda))^2}{2(\gamma x[j] + \lambda)}\right). \quad (11.10)$$

In addition to background irradiation (and/or dark current), the photodetected signal will also be corrupted by thermal noise. We assume this noise source is AWGN with zero mean and variance σ^2 . Thus the overall received electrical signal will be the convolution of (11.9) with the Gaussian thermal noise density. For large enough λ , where we may use the approximation (11.10), then we have

$$f_{Y|X}(y[j]|x[j]; \gamma, \lambda, \sigma^2) \approx \frac{1}{\sqrt{2\pi(\gamma x[j] + \lambda + \sigma^2)}} \exp\left(-\frac{(y[j] - (\gamma x[j] + \lambda))^2}{2(\gamma x[j] + \lambda + \sigma^2)}\right). \quad (11.11)$$

It is important to note that (11.10) and (11.11) are signal-dependent Gaussian noise channels, i.e. the noise variance is dependent on the value of $x[j]$. Moreover (11.11) implies we may write the photodetected channel model as

$$y[j] = \gamma x[j] + \lambda + \sqrt{\gamma x[j] + \lambda + \sigma^2} z[j], \quad (11.12)$$

where $z[j] \sim \mathcal{N}(0, 1)$. In addition, we may ignore the constant background irradiation contribution to the mean signal level, firstly because it can be removed straightforwardly by receiver electronics, and secondly it does not affect the input–output mutual information of the channel [23]. Hence we have

$$y[j] \approx \gamma x[j] + \sqrt{\gamma x[j] + \lambda + \sigma^2} z[j]. \quad (11.13)$$

When $\lambda + \sigma^2 \gg \gamma$ then the signal-dependent contribution to the noise variance can be considered negligible and we arrive at the signal-independent Gaussian noise model, i.e.

$$y[j] \approx \gamma x[j] + \sqrt{\lambda + \sigma^2} z[j]. \quad (11.14)$$

Many works on FSO communications have employed the signal-independent photodetection model, see e.g. [38], [39], [40], [41], [4], [42].

Comparing (11.13) and (11.14) with the received RF electrical signal model (11.4) we see an important difference: the received FSO photodetected signal is proportional to the optical power, whereas the received RF signal is proportional to the square root of the transmitted RF power. As we shall see later, this difference in power scaling will have significant repercussions on the optimal power allocation strategy.

11.3 Information-theoretic preliminaries

11.3.1 Mutual information

Consider a particular transmitted hybrid codeword consisting of A and B FSO and RF blocks respectively, with vector power allocations $\mathbf{p} = (p_1, \dots, p_A)$ and $\hat{\mathbf{p}} = (\hat{p}_1, \dots, \hat{p}_B)$. The FSO and RF components of the transmitted codeword experience vector channel realizations $\mathbf{h} = (h_1, \dots, h_A)$ and $\hat{\mathbf{h}} = (\hat{h}_1, \dots, \hat{h}_B)$, respectively. In the limit as $K, L \rightarrow \infty$, the largest achievable rate supported by the hybrid channel is the input–output *mutual information* [23]

$$I_{\text{tot}}(\mathbf{p}, \hat{\mathbf{p}}, \mathbf{h}, \hat{\mathbf{h}}) = \frac{N}{A} \sum_{a=1}^A I_{\mathcal{X}_{\text{fso}}}(h_a \rho p_a / N, \lambda, \sigma^2) + \frac{1}{B} \sum_{b=1}^B I_{\mathcal{X}_{\text{rf}}}(\hat{h}_b \hat{\rho} \hat{p}_b, \hat{\sigma}^2), \quad (11.15)$$

in bits per hybrid channel symbol, where

$$I_{\mathcal{X}}(\boldsymbol{\theta}) = \mathbb{E} \left[\log_2 \frac{f_{Y|X}(\mathbf{Y}|\mathbf{X}; \boldsymbol{\theta})}{f_Y(\mathbf{Y}; \boldsymbol{\theta})} \right] \quad (11.16)$$

denotes the input–output mutual information of the vector input channel with constellation \mathcal{X} , parameterized by the vector $\boldsymbol{\theta}$. Note that the achievable rate (11.15) implicitly assumes joint encoding and decoding across FSO and RF channels.

11.3.1.1 FSO channel mutual information

For the FSO channel, as discussed in Section 11.2.4, the transition probabilities $f_{Y|X}(\mathbf{Y}|\mathbf{X}; \boldsymbol{\theta})$ in (11.16) depend on the actual photodetection model. For the Poisson model (11.9), assuming the input vector symbols are drawn from \mathcal{X}_{fso} using density f_X , we have

$$I_{\mathcal{X}_{\text{fso}}}^{\text{Poisson}}(\gamma, \lambda) = -\mathbb{E} \left[\log_2 \left(\sum_{\mathbf{x}' \in \mathcal{X}_{\text{fso}}} f_X(\mathbf{x}') e^{\sum_{j=1}^J Y_j \log \left(\frac{\lambda + \gamma x'_j}{\lambda + \gamma X_j} \right) - \gamma (x'_j - X_j)} \right) \right], \quad (11.17)$$

where the expectation is over the dependent random vectors (X_1, \dots, X_J) and (Y_1, \dots, Y_J) . For the case when thermal noise and background irradiation are large, the signal-independent noise model (11.14) is valid. Thus,

$$I_{\mathcal{X}_{\text{fso}}}^{\text{awgn}} \left(\frac{\gamma^2}{\lambda + \sigma^2} \right) = -\mathbb{E} \left[\log_2 \left(\sum_{\mathbf{x}' \in \mathcal{X}_{\text{fso}}} f_X(\mathbf{x}') e^{-\sum_{j=1}^J \frac{\gamma^2}{\lambda + \sigma^2} (X_j - x'_j)^2 + 2 \frac{\gamma}{\sqrt{\lambda + \sigma^2}} (X_j - x'_j) Z_j} \right) \right], \quad (11.18)$$

where $Z_j \sim \mathcal{N}(0, 1)$. Figure 11.3(a) compares the mutual information of the Poisson channel (11.17) with that of the signal-independent Gaussian noise channel (11.18) (with $\sigma^2 = 0$, i.e. zero thermal noise) for on–off keying (OOK) with equi-probable symbols. We see that both sets of curves have similar shape. The Gaussian approximation tends to underestimate the mutual information at low powers and over-estimates for high powers.

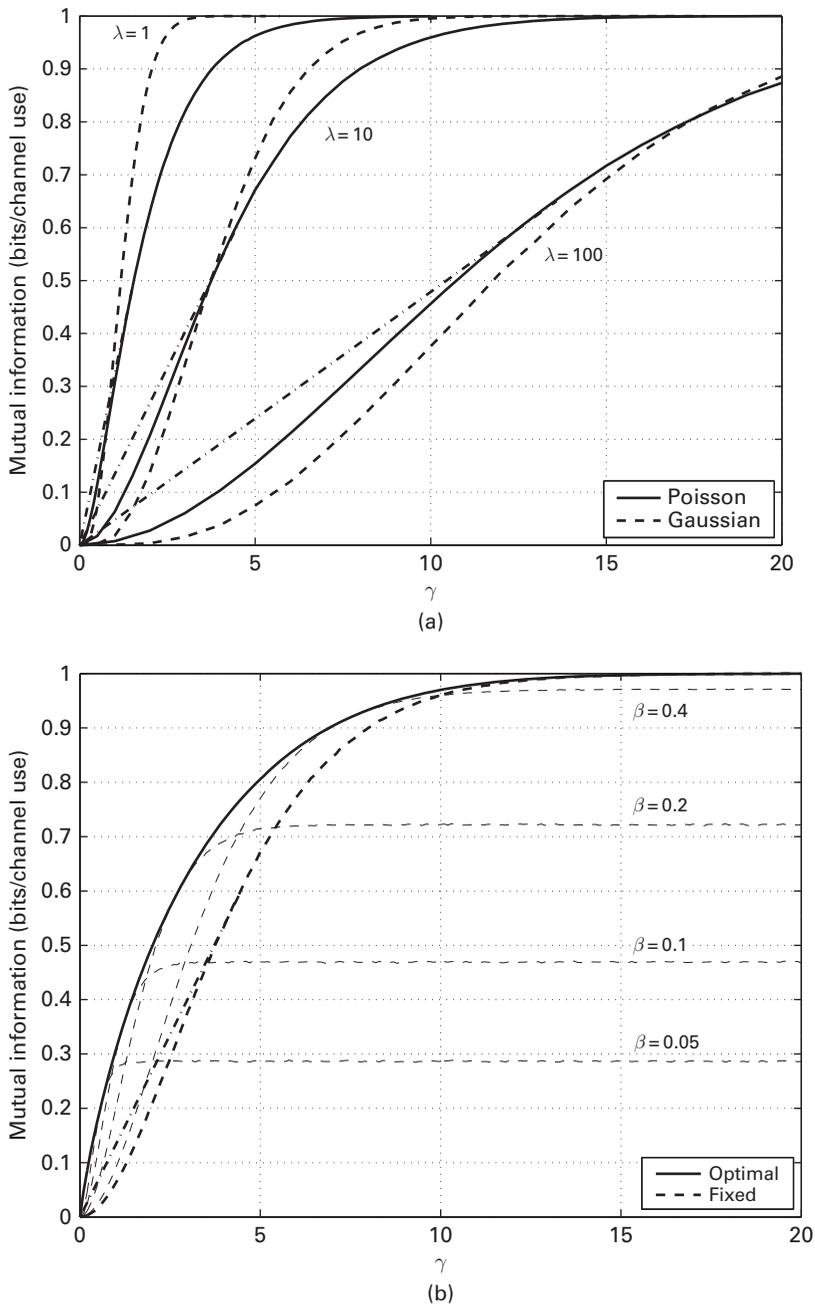


Figure 11.3 Mutual information of OOK: (a) compares Poisson channel (11.17) (solid) to the signal independent Gaussian noise channel (11.18) (dashed), with equi-probable symbols $\sigma^2 = 0$ and increasing λ ; (b) compares (11.22) with $Q = 2$ to OOK with equi-probable signalling (dashed) and $\lambda = 10$, various other input distributions are also shown (thin dashed), where $\beta = \beta_1 = 1 - \beta_0$. Time-sharing is also included for the sake of comparison (dash-dotted).

11.3.1.2 FSO channel, time-sharing mutual information

On closer inspection of Figure 11.3(a) we see that both sets of curves are not concave with γ . Moreover, we see that the curves are convex until a certain threshold power, at which point they become concave. This suggests equi-probable signalling is suboptimal, since, in the convex region, a higher achievable rate is possible via timesharing. More specifically, if the mutual information for equi-probable inputs is convex between $0 \leq \gamma < \gamma_{\text{ts}}$ we can employ time-sharing between points $(0, 0)$ and $(\gamma_{\text{ts}}, I_{\mathcal{X}_{\text{fso}}}(\gamma_{\text{ts}}, \lambda))$, i.e. when $\gamma < \gamma_{\text{ts}}$ we transmit using a peak power of γ_{ts} , but only for a fraction of $\gamma/\gamma_{\text{ts}}$ of the block interval and zero power for the remaining $1 - \gamma/\gamma_{\text{ts}}$ fraction. Hence the average power of the block is γ , and the resulting achievable rate is therefore given by

$$I_{\mathcal{X}_{\text{fso}}}^{\text{ts}}(\gamma, \lambda) = \begin{cases} I_{\mathcal{X}_{\text{fso}}}(\gamma_{\text{ts}}, \lambda) \frac{\gamma}{\gamma_{\text{ts}}} & \gamma < \gamma_{\text{ts}} \\ I_{\mathcal{X}_{\text{fso}}}(\gamma, \lambda) & \text{otherwise.} \end{cases} \quad (11.19)$$

The threshold γ_{ts} is the point at which the tangent to the curve $I_{\mathcal{X}_{\text{fso}}}(\gamma, \lambda)$ passes through the origin. Hence γ_{ts} is the solution to

$$I_{\mathcal{X}_{\text{fso}}}(\gamma, \lambda) = \gamma \phi_{\mathcal{X}_{\text{fso}}}(\gamma, \lambda), \quad \gamma > 0, \quad (11.20)$$

where

$$\phi_{\mathcal{X}}(\gamma, \lambda) \triangleq \frac{\partial}{\partial \gamma} I_{\mathcal{X}_{\text{fso}}}(\gamma, \lambda) \quad (11.21)$$

is the derivative of the mutual information with respect to γ .¹ Note that γ_{ts} is dependent on \mathcal{X}_{fso} and the background irradiation λ . Threshold values for some example modulation schemes are listed in Table 11.1.

Figure 11.3(a) also compares with the resulting time-sharing scheme (dash-dotted line) using the parameters from Table 11.1. As we observe, some improvement is possible, especially for low γ .

For signal-independent Gaussian channels, $\phi_{\mathcal{X}}(\gamma, \lambda)$ is the minimum mean-squared error (MMSE) [43]. Whereas for Poisson channels it is a certain function of the MMSE [44], and we provide further details for its computation in Appendix B.

11.3.1.3 FSO channel, optimal PAM mutual information

Whilst time-sharing obtains a higher achievable rate than fixed equi-probable signalling, the periods of zero power transmission represent an inefficient use of the block interval.

Table 11.1 Time-sharing thresholds for the Poisson channel, $[\gamma_{\text{ts}}, I_{\mathcal{X}_{\text{fso}}}(\gamma_{\text{ts}}, \lambda)]$, where γ_{ts} is measured in Decibels.

| λ (dB) | OOK | 2PPM | 4PPM |
|----------------|--------------|--------------|--------------|
| 0 | [1.80,0.49] | [3.16,0.52] | [3.76,1.09] |
| 10 | [6.50,0.60] | [6.18,0.62] | [8.48,1.29] |
| 20 | [11.33,0.65] | [12.80,0.66] | [13.38,1.38] |

¹ Note that if $I_{\mathcal{X}_{\text{fso}}}(\gamma, \lambda)$ is concave, the only solution to (11.20) is $\gamma = 0$

Time-sharing is in fact a naive way of changing the input distribution from the channel's perspective.² For example, if equi-probable OOK is employed, in the time-sharing interval, $\gamma < \gamma_{\text{ts}}$, from the channel's perspective the probability of a pulse being input to the channel is reduced from $\frac{1}{2}$ to $\frac{\gamma}{2\gamma_{\text{ts}}}$. It turns out we can obtain an even higher achievable rate by optimizing the input distribution. In particular, for a fixed number of PAM mass points, we find the mass point locations and their probabilities that maximizes the mutual information subject to an average (block) power constraint, i.e. for Q -ary PAM, let μ_0, \dots, μ_{Q-1} and $\beta_0, \dots, \beta_{Q-1}$ denote the mass point locations and their associated probabilities, then

$$I_{\mathcal{X}_{\text{PAM}}}^{\text{opt}}(\gamma, \lambda) = \begin{cases} \text{Maximize:} & I_{\mathcal{X}_{\text{PAM}}}(\boldsymbol{\mu}, \boldsymbol{\beta}, \lambda) \\ \text{Subject to:} & \sum_{q=0}^{Q-1} \mu_q \beta_q \leq \gamma \\ & \sum_{q=0}^{Q-1} \mu_q = 1 \\ & 0 \leq \beta_q \leq 1, \quad q = 0, \dots, Q-1, \end{cases} \quad (11.22)$$

where, e.g. for the Poisson channel,

$$I_{\mathcal{X}_{\text{PAM}}}(\boldsymbol{\mu}, \boldsymbol{\beta}, \lambda) = - \sum_{q=0}^{Q-1} \beta_q \log_2 \beta_q - \sum_{q=0}^{Q-1} \beta_q \mathbb{E}_{Y|X=\mu_q} \left[\log_2 \left(1 + \sum_{j \neq q} \frac{\beta_j}{\beta_q} \left(\frac{\lambda + \mu_j}{\lambda + \mu_q} \right)^Y e^{-(\mu_j - \mu_q)} \right) \right]. \quad (11.23)$$

Note that a peak power constraint may also be considered in (11.22). Indeed, for bandwidth constrained Poisson channels with average and peak power constraints, the capacity-achieving input distribution is proven to be discrete, i.e. a PAM scheme whereby the mutual information is maximized over the mass point locations and their associated probabilities [45]. When there is no peak power constraint, the optimal input distribution has only been conjectured to be discrete [45]. For high photon counts it has been shown to be approximately exponential [46].

The solid line in Figure 11.3(b) shows the mutual information $I_{\mathcal{X}_{\text{OOK}}}^{\text{opt}}(\gamma, \lambda)$ in comparison to equi-probable OOK as shown by the dot-dashed curve. From this figure, we see optimizing over the input distribution not only results in a larger achievable rate (for the same average power), but it also results in a concave mutual information curve. As we observe, optimizing the input distribution yields significant gains also when compared to time-sharing.

11.3.1.4 RF channel mutual information

For the RF channel (11.4) with power $\hat{\gamma}$ and noise variance $\hat{\sigma}^2$, the mutual information is [47]

² That is, picking a symbol interval from the entire block uniformly at random.

$$I_{\mathcal{X}_{\text{rf}}}^{\text{awgn}}\left(\frac{\hat{\gamma}}{\hat{\sigma}^2}\right) = -\mathbb{E}\left[\log_2\left(\sum_{\hat{x}' \in \mathcal{X}_{\text{rf}}} f_{\hat{X}}(\hat{x}') e^{-\frac{\hat{\gamma}}{\hat{\sigma}^2}|\hat{X}-\hat{x}'|^2 - 2\sqrt{\frac{\hat{\gamma}}{\hat{\sigma}^2}}\Re\{(\hat{X}-\hat{x}')Z^*\}}\right)\right], \quad (11.24)$$

where the expectation is over the independent random variables \hat{X} and $Z \sim \mathcal{N}_{\mathbb{C}}(0, 1)$. From (11.18) and (11.24) we see that for signal-independent Gaussian channels, the mutual information need only be parameterized by a certain ratio of the various channel parameters. For the RF channel it is the signal-to-noise ratio (SNR), i.e. $\frac{\hat{\gamma}}{\hat{\sigma}^2}$. However, it is important to note that for the FSO case, as a consequence of (11.14), $\frac{\gamma^2}{\lambda + \sigma^2}$ is not the SNR, since it is proportional to the optical power squared.

11.3.2 Outage probability

Suppose the hybrid system is required to operate at a target rate of R bits per hybrid channel symbol. Since A and B are finite, and I_{tot} is a function of $\mathbf{h}, \hat{\mathbf{h}}$, which are realizations of the fading random variables, then I_{tot} is also a random variable.³ Moreover, there is a non-zero probability that $I_{\text{tot}} < R$, i.e. the channel cannot support the desired target rate. Such events are referred to as *outages*, and the probability that they occur,

$$P_{\text{out}}(P_{\text{av}}, R) \triangleq \Pr\left\{I_{\text{tot}}(\mathbf{p}, \hat{\mathbf{p}}, \mathbf{h}, \hat{\mathbf{h}}) < R\right\}, \quad (11.25)$$

is called the *outage probability*. Note that we have written P_{out} as a function of the average power constraint P_{av} as well as the target rate total R . The information outage probability lower bounds the codeword error probability of any coding scheme [29], [30] and has been shown to be the fundamental limit of block-fading channels [49]. In particular, there exist coding schemes of sufficiently large block length whose error probability is arbitrarily close to the outage probability; conversely, the error probability of any sufficiently long code is lower-bounded by the outage probability. In order to simplify the expressions that will follow, we define the rate of the underlying binary code as

$$R_c = \frac{R}{Nq + m}. \quad (11.26)$$

11.3.3 Outage exponent

Analytical computation of the outage probability (11.25) requires a closed-form expression of the distribution of the instantaneous input–output mutual information (11.15). Such a task is intractable in general, and the usual approach is to numerically compute the outage probability via Monte Carlo simulations. Whilst this approach is simple to implement, at low outage probability values, it requires lengthy simulations to count enough outage events for a reliable estimate. In addition, it reveals little insight as to how various channel model and system design parameters affect the outage performance

³ The channel is therefore not information stable and the channel capacity in the strict Shannon sense is zero [48].

of the system. Towards this end, we study the *outage exponent*, which is substantially simpler to compute analytically. The outage exponent is defined as

$$d^{(k)} \triangleq \lim_{P_{\text{av}} \rightarrow \infty} - \frac{\log P_{\text{out}}(P_{\text{av}}, R)}{(\log P_{\text{av}})^k}, \quad (11.27)$$

where $k = 1, 2$. In simple terms, the outage exponent describes the slope of the outage probability curve (when plotted on a log-log scale) in the high power regime. The larger the outage exponent, the faster the outage probability will decrease with increasing power. Note that by including the integer k , (11.27) is more general than the outage exponent normally defined in RF systems [24], [25]. This modified definition is required to allow for scintillation cases where the resulting outage probability curve will not converge to a constant slope when plotted on a log-log scale, but does when plotted on a log-log² scale (most notably under weak turbulence conditions where the scintillation is log-normal distributed [2]).

Clearly, the outage exponent will depend on the distribution of the fading coefficients. Rather than assuming a specific distribution, we characterize the fading via the component channel's single block transmission outage exponent, defined as

$$d_{\text{fso}}^{(i)} \triangleq \lim_{\gamma \rightarrow \infty} - \frac{\log \Pr\{I_{\text{fso}}(h\gamma) < R_{\text{fso}}\}}{(\log \gamma)^i} \quad (11.28)$$

$$d_{\text{rf}}^{(j)} \triangleq \lim_{\gamma \rightarrow \infty} - \frac{\log \Pr\{I_{\text{rf}}(\hat{h}\gamma) < R_{\text{rf}}\}}{(\log \gamma)^j}, \quad (11.29)$$

for given component channel rate constraints $0 < R_{\text{fso}} < Nq$ and $0 < R_{\text{rf}} < m$, where $i, j \in \{1, 2\}$. We see that (11.28) and (11.29) imply

$$F_H \left(\frac{c}{\gamma} \right) \doteq e^{-d_{\text{fso}}^{(i)} (\log \gamma)^i}, \quad F_{\hat{H}} \left(\frac{c}{\gamma} \right) \doteq e^{-d_{\text{rf}}^{(j)} (\log \gamma)^j}, \quad (11.30)$$

where $c > 0$ is some arbitrary constant (which in this case is the inverse of the mutual information function). In other words, (11.28) and (11.29) each define a family of scintillation distributions whose left-sided tail behaves like (11.30). In Table 11.2 we list some single block transmission outage exponents (derived in [2]) for some typical scintillation distributions [1].⁴

Note that the exponents derived in [2] defined the outage exponent in terms of the received electrical SNR. In this chapter, we define the FSO exponent in terms of the transmitted optical power. This results in a factor of 2 in the exponents listed in Table 11.2 compared to those given in [2].

⁴ Note that if MIMO FSO with transmit repetition and equal gain combining is employed, then the exponents listed in Table 11.2 are simply multiplied by $N_t N_r$, where N_t and N_r denote the number of lasers and apertures respectively [2].

Table 11.2 Outage exponent for some typical scintillation distributions.

| Distribution | PDF $f(h)$ | SNR exponent |
|----------------|--|---|
| Exponential | $\exp(-h)$ | $d^{(1)} = 1$ |
| Lognormal | $\frac{1}{h\bar{\sigma}\sqrt{2\pi}} e^{-(\log h - \bar{\mu})^2 / (2\bar{\sigma}^2)}$ | $d^{(2)} = \frac{1}{4\log(1+\bar{\sigma}^2)}$ |
| Gamma-gamma | $\frac{2(\bar{a}\bar{b})^{\frac{\bar{a}+\bar{b}}{2}}}{\Gamma(\bar{a})\Gamma(\bar{b})} h^{\frac{\bar{a}+\bar{b}}{2}-1} K_{\bar{a}-\bar{b}}(2\sqrt{\bar{a}\bar{b}h})$ | $d^{(1)} = \min(\bar{a}, \bar{b})$ |
| Lognormal-Rice | $\frac{(1+\bar{r})e^{-\bar{r}}}{\sqrt{2\pi}\bar{\sigma}} \int_0^\infty \frac{1}{v^2} I_0\left(2\sqrt{\frac{(1+\bar{r})\bar{r}h}{v}}\right) e^{-\frac{(1+\bar{r})h}{v} - \frac{1}{2\bar{\sigma}^2}\left(\log v + \frac{1}{2}\bar{\sigma}^2\right)^2} dv$ | $d^{(1)} = \frac{1}{2}$ |
| I-K | $\begin{cases} 2\bar{\alpha}(1+\bar{\rho}) \left[\frac{(1+\bar{\rho})h}{\bar{\rho}}\right]^{\frac{\bar{\alpha}-1}{2}} K_{\bar{\alpha}-1}(2\sqrt{\bar{\alpha}\bar{\rho}}) I_{\bar{\alpha}-1}(2\sqrt{\bar{\alpha}h(1+\bar{\rho})}) & h < \frac{\bar{\rho}}{1+\bar{\rho}} \\ 2\bar{\alpha}(1+\bar{\rho}) \left[\frac{(1+\bar{\rho})h}{\bar{\rho}}\right]^{\frac{\bar{\alpha}-1}{2}} I_{\bar{\alpha}-1}(2\sqrt{\bar{\alpha}\bar{\rho}}) K_{\bar{\alpha}-1}(2\sqrt{\bar{\alpha}h(1+\bar{\rho})}) & h > \frac{\bar{\rho}}{1+\bar{\rho}} \end{cases}$ | $d^{(1)} = \frac{1}{2}\bar{\alpha}$ |

11.4 Uniform power allocation

First let us assume that perfect CSI is known only at the receiver (CSIR case). The transmitter allocates power uniformly across all blocks, i.e. $p_1 = \dots = p_A = \hat{p}_1 = \dots = \hat{p}_B = p = P_{\text{av}}$.

THEOREM 11.1 *Suppose the mutual information of the FSO and RF non-fading channels are monotonically increasing functions of power, and asymptotically behave as follows,*

$$\lim_{\gamma \rightarrow \infty} I_{\mathcal{X}_{\text{fso}}}(\gamma, \lambda, \sigma^2) = q \quad (11.31)$$

$$\lim_{\gamma \rightarrow 0} I_{\mathcal{X}_{\text{fso}}}(\gamma, \lambda, \sigma^2) = 0 \quad (11.32)$$

$$\lim_{\hat{\gamma} \rightarrow \infty} I_{\mathcal{X}_{\text{rf}}}(\hat{\gamma}, \hat{\sigma}^2) = m \quad (11.33)$$

$$\lim_{\hat{\gamma} \rightarrow 0} I_{\mathcal{X}_{\text{rf}}}(\hat{\gamma}, \hat{\sigma}^2) = 0. \quad (11.34)$$

Define component channel SNR exponents $d_{\text{fso}}^{(i)}$ and $d_{\text{rf}}^{(j)}$ as in (11.28) and (11.29) respectively. Assuming $\rho, \hat{\rho} > 0$ and $i = j = k$, then the overall hybrid outage exponent is given by

$$d^{(k)} = \inf_{\mathcal{K}(\delta, R_c)} \left\{ d_{\text{fso}}^{(k)} \kappa_1 + d_{\text{rf}}^{(k)} \kappa_2 \right\}, \quad (11.35)$$

where

$$\mathcal{K}(\delta, R_c) \triangleq \left\{ \kappa_1, \kappa_2 \in \mathbb{Z} : \delta \frac{\kappa_1}{A} + (1-\delta) \frac{\kappa_2}{B} > 1 - R_c, \quad 0 \leq \kappa_1 \leq A, \quad 0 \leq \kappa_2 \leq B \right\}. \quad (11.36)$$

The generality of conditions (11.31) to (11.34) imply that the theorem is applicable to all of the photodetection models discussed in Section 11.2.4 and most practical discrete modulation schemes of interest. However, it should be noted that the time-sharing and optimal PAM signalling strategies described in Section 11.3.1 cannot be applied as they require knowledge of the CSI at the transmitter.

From Theorem 11.1, we see that the overall outage exponent depends on R_c , δ , A , B and the individual exponents $d_{\text{fso}}^{(k)}$ and $d_{\text{rf}}^{(k)}$ in a non-trivial way. However, if one considers $(\kappa_1, \kappa_2, d^{(k)})$ as a three-dimensional space, we see that the overall outage exponent is dominated by the end points $(A(1 - R_c)/\delta, 0, d_{\text{fso}}^{(k)}A(1 - R_c)/\delta)$ and $(0, B(1 - R_c)/(1 - \delta), d_{\text{rf}}^{(k)}B(1 - R_c)/(1 - \delta))$ of the line of intersection between the planes defined by $d^{(k)} = \kappa_1 d_{\text{fso}}^{(k)} + \kappa_2 d_{\text{rf}}^{(k)}$ and $\delta \frac{\kappa_1}{A} + (1 - \delta) \frac{\kappa_2}{B} = 1 - R_c$. Thus we find, as a rule of thumb, if

$$\frac{d_{\text{fso}}^{(k)}}{d_{\text{rf}}^{(k)}} < \frac{B}{A} \frac{\delta}{1 - \delta} = \frac{B N q}{A m}, \quad (11.37)$$

then the scintillation on the FSO channel will dominate the overall hybrid exponent (and vice versa).

Although in general, the solution to (11.35) can be straightforwardly determined numerically, it is difficult to obtain insight as to how the various system parameters influence the overall SNR exponent. However, for the most basic and interesting scenario, $A = B = 1$, the solution to (11.35) reduces to a simple intuitive form.

COROLLARY 11.1 *Suppose $A = B = 1$. The solution to (11.35) is divided into two cases as follows.*

(1) *If $\delta \leq \frac{1}{2}$, then*

$$d^{(k)} = \begin{cases} d_{\text{fso}}^{(k)} + d_{\text{rf}}^{(k)} & 0 < R_c \leq \delta \\ d_{\text{rf}}^{(k)} & \delta < R_c \leq 1 - \delta \\ \min(d_{\text{fso}}^{(k)}, d_{\text{rf}}^{(k)}) & 1 - \delta < R_c < 1. \end{cases} \quad (11.38)$$

(2) *If $\delta \geq \frac{1}{2}$, then*

$$d^{(k)} = \begin{cases} d_{\text{fso}}^{(k)} + d_{\text{rf}}^{(k)} & 0 < R_c \leq 1 - \delta \\ d_{\text{fso}}^{(k)} & 1 - \delta < R_c \leq \delta \\ \min(d_{\text{fso}}^{(k)}, d_{\text{rf}}^{(k)}) & \delta < R_c < 1. \end{cases} \quad (11.39)$$

From Corollary 11.1 we can see directly how the hybrid system parameters will affect the asymptotic slope of the outage probability curve for the single block transmission case. In particular, we see that the asymptotic slope is affected by the SNR exponent of the individual component channels (which is in turn dependent on the scintillation distribution), the size of the FSO/RF signal set constellation, and the overall binary code rate of the system. In most practical systems, $\delta \geq \frac{1}{2}$, i.e. in a hybrid symbol period, the number of transmitted FSO bits will be greater than the number of RF transmitted

bits. From (11.39), we see that the highest diversity is achieved if the binary code rate R_c is set to be less than $1 - \delta = m/(m + nq)$, i.e. the total information rate is less than the maximum information rate of the stand-alone RF channel. If $1 - \delta < R_c \leq \delta$, the exponent is the same as a single FSO link, i.e. the additional coding over an RF link will not improve the asymptotic slope of the outage probability curve. For high binary code rates, $\delta < R_c < 1$, the asymptotic performance is dominated by the worst of the two exponents. Note that code rates above δ are not achievable with a stand-alone FSO link.

Although we have concentrated on the single block case, given the short coherence time of the optical channel compared to the RF channel, the cases of $A = 2, 3$ and $B = 1$ are also of practical interest and are readily evaluated from Theorem 11.1. In particular, these provide significant outage exponent improvements for lower rates, compared to the single block case. This is illustrated in Figure 11.4, which plots the hybrid outage exponent with, $d_{\text{fso}}^{(k)} = 2$, $d_{\text{rf}}^{(k)} = 1$ and $\delta = 0.8$ for a: $A = 1, B = 1$ system (dash-dotted line); $A = 3, B = 1$ system (solid line); and a stand-alone FSO link with three blocks,⁵ i.e. $A = 3, B = 0$ (dashed line). By coding over more FSO blocks, vast improvements in the outage exponent can be seen, particularly as the code rate decreases. In addition, we see that the $A = 3, B = 1$ exponent exceeds that of the stand-alone FSO system by $d_{\text{rf}}^{(k)}$ for most code rates.

THEOREM 11.2 Define component channel outage exponents $d_{\text{fso}}^{(i)}$ and $d_{\text{rf}}^{(j)}$ as in (11.28) and (11.29) respectively. Suppose $i > j$ then the outage exponent is

$$d^{(i)} = d_{\text{fso}}^{(i)} \left(1 + \left\lfloor \frac{A}{\delta} (\delta - R_c) \right\rfloor \right) \quad 0 < R_c \leq \delta \quad (11.40)$$

$$d^{(j)} = d_{\text{rf}}^{(j)} \left(1 + \left\lfloor \frac{B}{1 - \delta} (1 - R_c) \right\rfloor \right) \quad \delta < R_c < 1. \quad (11.41)$$

Otherwise, if $i < j$ then the outage exponent is

$$d^{(j)} = d_{\text{rf}}^{(j)} \left(1 + \left\lfloor \frac{B}{1 - \delta} (1 - \delta - R_c) \right\rfloor \right) \quad 0 < R_c \leq 1 - \delta. \quad (11.42)$$

$$d^{(i)} = d_{\text{fso}}^{(i)} \left(1 + \left\lfloor \frac{A}{\delta} (1 - R_c) \right\rfloor \right) \quad 1 - \delta < R_c < 1 \quad (11.43)$$

Theorem 11.2 shows how the overall performance of the hybrid channel will be affected when one of the component channels has an asymptotic outage probability that decays with power much faster than the other. In particular, we see that the overall outage exponent will be dominated by the worst of the two component channel outage exponents unless the binary code rate is below a certain threshold dependent on the ratio of FSO bits to total transmitted bits (δ). Note that the exponents in Theorems 11.1 and 11.2 are related to how outage-approaching codes should be designed. In particular, they are a form of the Singleton bound on the blockwise Hamming distance [49], [50], [51], [25].

⁵ Note that for the stand-alone FSO system with A blocks, the exponent is given by $d^{(k)} = d_{\text{fso}}^{(k)}(1 + \lfloor A(1 - R_c/\delta) \rfloor)$ for $0 < R_c < \delta$, since we have defined R_c with respect to the hybrid system.

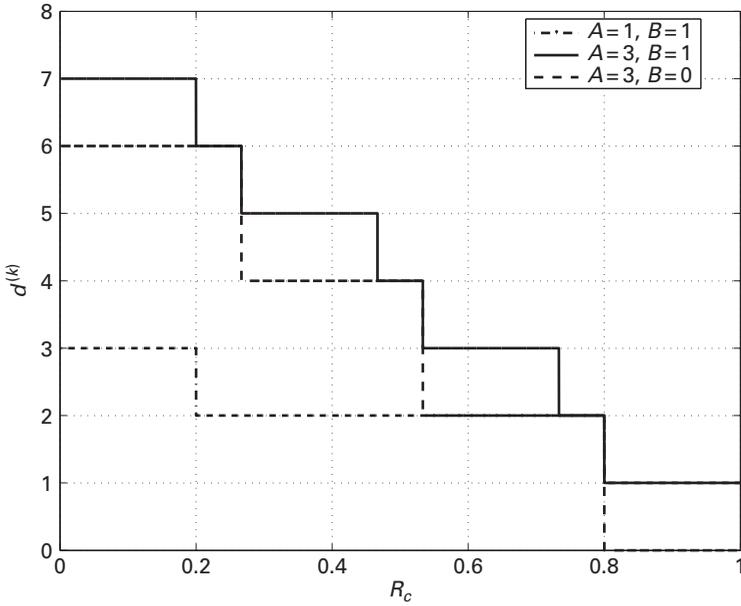


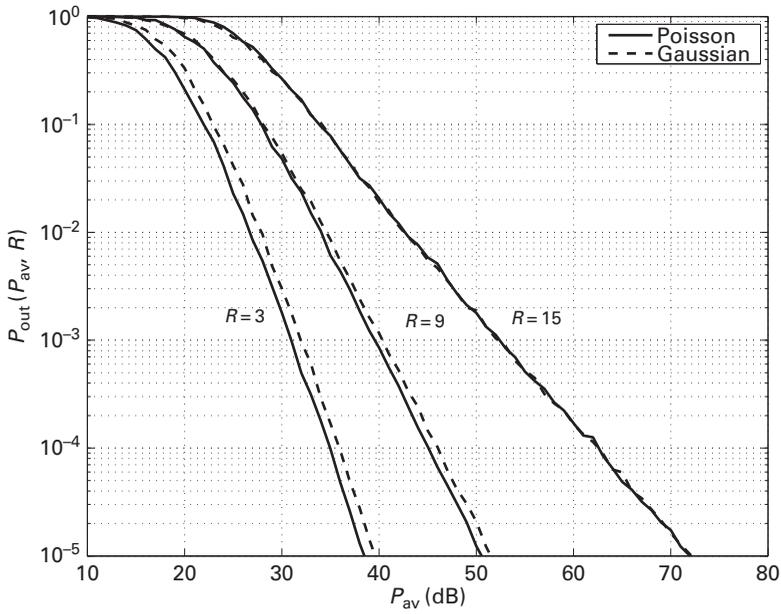
Figure 11.4 Hybrid channel outage exponent (11.35) with $d_{\text{iso}}^{(k)} = 2$, $d_{\text{rf}}^{(k)} = 1$ and $\delta = 0.8$: $A = 1$, $B = 1$ system (dash-dotted line); $A = 3$, $B = 1$ system (solid line); and a stand-alone FSO link with three blocks, i.e. $A = 3$, $B = 0$ (dashed line).

11.4.1 Numerical results

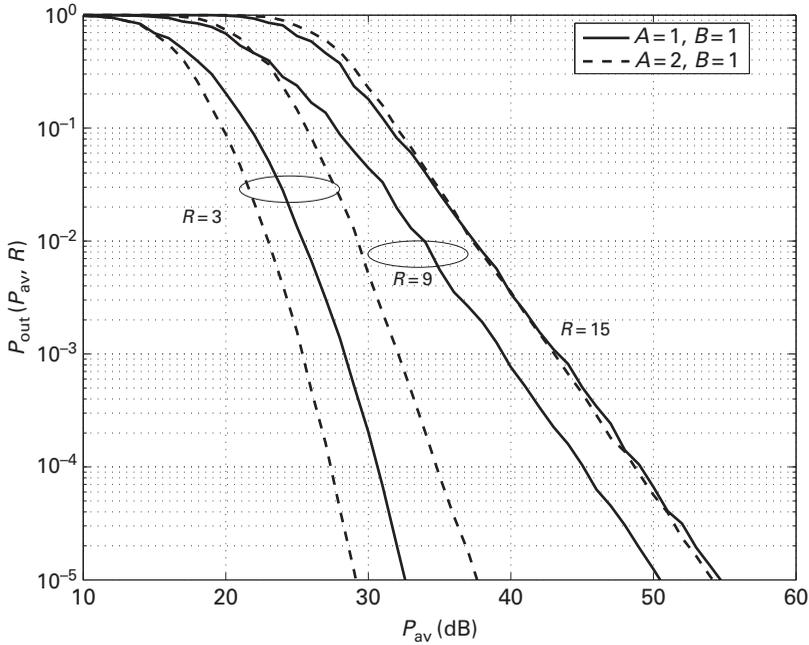
To demonstrate the implications of our asymptotic results, we conducted a number of Monte Carlo simulations. This involved computing the mutual information curves $I_{\mathcal{X}_{\text{iso}}}(\gamma)$ and $I_{\mathcal{X}_{\text{rf}}}(\gamma)$ for a given constellation. These curves are then used as lookup tables to determine the total mutual information (11.15) for a given set of vector fading realizations \mathbf{h} and $\hat{\mathbf{h}}$, which are generated randomly according to one of the distributions listed in Table 11.2.

We note that since all the models under consideration meet the conditions in Eqs. (11.31)–(11.34), they all have the same exponents, provided the coding rates and modulations are the same. Figure 11.5(a) confirms this fact for transmission with $R = 3, 9$ and 15 ($R_c = \frac{1}{6}, \frac{1}{2}$ and $\frac{5}{6}$) over the Poisson channel (solid lines) and the signal-independent Gaussian model (11.14) with $\sigma^2 = 0$ (dashed lines). The FSO link employs OOK modulation while the RF link employs 64-QAM; we assume $N = 12$. For these parameters we have that $\delta = \frac{2}{3}$. Not only the exponents are the same, but we also observe very little quantitative difference between the Poisson and signal-independent Gaussian models. In particular, the difference becomes smaller for high rates. This is due to the fact that for high rates, the mutual information for both Poisson and Gaussian models saturate at the same point.

Figure 11.5(b) shows the outage probability for the same setup as 11.5(a), but assuming a lognormal scintillation distribution in the RF link. The Poisson channel model is



(a) Poisson vs. Gaussian models for FSO with gamma-gamma scintillation ($\bar{a} = \bar{b} = 2$); RF: exponential scintillation.



(b) Poisson FSO with gamma-gamma scintillation ($\bar{a} = \bar{b} = 2$); RF: lognormal scintillation with parameters $\bar{\sigma} = 0.8326$ and $\bar{\mu} = -0.3466$, corresponding to a scintillation index of 1.

Figure 11.5 Outage performance for the hybrid FSO/RF channel with $\rho = \hat{\rho} = 0.5$, $\lambda = \sigma^2 = 10$, $N = 12$ OOK modulation (FSO link) and 64-QAM modulation (RF link).

assumed for the FSO link. The lognormal scintillation distribution is the only one that induces a nonzero second order exponent $d^{(2)}$. Therefore, the results of Theorem 11.2 are relevant in this case, where the first- and second-order exponents $d^{(1)}$ and $d^{(2)}$ are related. In particular, for $A = 1$ we have that when $R = 3$, we have that $d^{(2)} = \frac{1}{4 \log 2}$ (as given in Table 11.2), while for $R = 9, 15$ from Theorem 11.2 we obtain that $d^{(1)} = 2$. Instead, in the case where $A = 2$, for $R = 3$ we have that $d^{(2)} = \frac{1}{4 \log 2}$, while for $R = 9, 15$ we have $d^{(1)} = 4, 2$, respectively. In both cases, for $R = 3$, the dominating exponent is that of the RF link, i.e., $d^{(2)}$, while for higher rates the dominating exponent is that of the FSO link, i.e., $d^{(2)} = 0$ and $d^{(1)}$ is finite.

11.5 Power allocation

Since the fading experienced on both the FSO and RF links is slow compared to typical signalling rates, it is reasonable to assume the receiver is able to estimate the CSI and feedback this information to the transmitter. Using this information, the transmitter can therefore adapt the power (subject to power constraints) to compensate for channel fluctuations and significantly reduce the outage probability. Realistically, since the scintillation is a time-varying process, the transmitter will only have access to the CSI of past blocks of the codeword (causal CSIT). Moreover, the estimation process at the receiver may induce errors resulting in imperfect CSI at the transmitter (imperfect CSIT). Analysis of these effects on the outage diversity in general block-fading channels can be found in [31], [32], [52], [53] and references therein. In this section, we assume the transmitter has perfect a-causal knowledge of the CSI. Whilst this assumption is unrealistic, it provides an important benchmark for practical systems.

11.5.1 Optimal solution

In general, to find the optimal power allocation strategy, we require the solution to the following minimization problem.

$$\begin{cases} \text{Minimize:} & \Pr \left\{ I_{\text{tot}}(\mathbf{p}, \hat{\mathbf{p}}, \mathbf{h}, \hat{\mathbf{h}}) < R \right\} \\ \text{Subject to:} & \mathbb{E}[\langle \mathbf{p} \rangle] + \mathbb{E}[\langle \hat{\mathbf{p}} \rangle] \leq P_{\text{av}}, \\ & \langle \mathbf{p} \rangle \leq \alpha_{\text{fso}} P_{\text{av}}, \quad \langle \hat{\mathbf{p}} \rangle \leq \alpha_{\text{rf}} P_{\text{av}}. \end{cases} \quad (11.44)$$

To solve (11.44) first consider the following lemma.

LEMMA 11.1 *Let \mathbf{p}^* and $\hat{\mathbf{p}}^*$ denote the solution to*

$$\begin{cases} \text{Minimize:} & \langle \mathbf{p} \rangle + \langle \hat{\mathbf{p}} \rangle \\ \text{Subject to:} & I_{\text{tot}}(\mathbf{p}, \hat{\mathbf{p}}, \mathbf{h}, \hat{\mathbf{h}}) \geq R. \end{cases} \quad (11.45)$$

Then the solution to

$$\begin{cases} \text{Minimize:} & \langle \mathbf{p} \rangle + \langle \hat{\mathbf{p}} \rangle \\ \text{Subject to:} & I_{\text{tot}}(\mathbf{p}, \hat{\mathbf{p}}, \mathbf{h}, \hat{\mathbf{h}}) \geq R \\ & \langle \mathbf{p} \rangle \leq \alpha_{\text{fso}} P_{\text{av}}, \langle \hat{\mathbf{p}} \rangle \leq \alpha_{\text{rf}} P_{\text{av}} \\ & \mathbf{p}, \hat{\mathbf{p}} \succeq \mathbf{0}, \end{cases} \quad (11.46)$$

denoted by \wp and $\hat{\wp}$, is separated into four cases as follows.

- (1) If \mathbf{p}^* and $\hat{\mathbf{p}}^*$ both satisfy their respective short-term power constraints in (11.46). Then $\wp = \mathbf{p}^*$ and $\hat{\wp} = \hat{\mathbf{p}}^*$.
- (2) If the power allocation over the FSO link satisfies its short-term power constraint, but the RF link does not, i.e. $\langle \mathbf{p}^* \rangle \leq \alpha_{\text{fso}} P_{\text{av}}$ and $\langle \hat{\mathbf{p}}^* \rangle > \alpha_{\text{rf}} P_{\text{av}}$. Then $\hat{\wp}$ is the solution to

$$\begin{cases} \text{Maximize:} & I_{\text{rf}}(\hat{\mathbf{p}}, \hat{\mathbf{h}}) \\ \text{Subject to:} & \hat{\mathbf{p}} = \alpha_{\text{rf}} P_{\text{av}}, \\ & \hat{\mathbf{p}} \succeq \mathbf{0}, \end{cases} \quad (11.47)$$

and \wp is the solution to

$$\begin{cases} \text{Minimize:} & \langle \mathbf{p} \rangle \\ \text{Subject to:} & I_{\text{fso}}(\mathbf{p}, \mathbf{h}) + I_{\text{rf}}(\hat{\wp}, \hat{\mathbf{h}}) = R \\ & \hat{\mathbf{p}} \succeq \mathbf{0}. \end{cases} \quad (11.48)$$

If $\wp > \alpha_{\text{fso}} P_{\text{av}}$, then the solution to (11.46) is infeasible. Note that if $R - I_{\text{rf}}(\hat{\wp}, \hat{\mathbf{h}}) > Nq$, then the solution to (11.48), and hence (11.46) is also infeasible. In other words, the short-fall in mutual information required to avoid outage exceeds the maximum possible rate supported by the FSO link.

- (3) If $\langle \mathbf{p}^* \rangle > \alpha_{\text{fso}} P_{\text{av}}$ and $\langle \hat{\mathbf{p}}^* \rangle \leq \alpha_{\text{rf}} P_{\text{av}}$. Then the solution to (11.46) is the same as the previous case, with the roles of rf and fso interchanged.
- (4) If $\langle \mathbf{p}^* \rangle > \alpha_{\text{fso}} P_{\text{av}}$ and $\langle \hat{\mathbf{p}}^* \rangle > \alpha_{\text{rf}} P_{\text{av}}$, then the solution to (11.46) is infeasible.

THEOREM 11.3 The solution to problem (11.44) is given by

$$(\wp^*, \hat{\wp}^*) = \begin{cases} (\wp, \hat{\wp}) & \langle \wp \rangle + \langle \hat{\wp} \rangle \leq s^* \\ (\mathbf{0}, \mathbf{0}) & \text{otherwise,} \end{cases} \quad (11.49)$$

where $(\wp, \hat{\wp})$ is the solution to (11.46) and s^* is a threshold determined by

$$s^* = \sup \left\{ s : \mathbb{E}_{(\mathbf{h}, \hat{\mathbf{h}}) \in \mathcal{R}(s)} [\langle \wp \rangle + \langle \hat{\wp} \rangle] \leq P_{\text{av}} \right\}, \quad (11.50)$$

where

$$\mathcal{R}(s) \triangleq \left\{ (\mathbf{h}, \hat{\mathbf{h}}) \in \mathbb{R}^{A+B} : \langle \hat{\rho} \rangle + \langle \hat{\rho} \rangle \leq s \right\}. \quad (11.51)$$

Proof The proof follows similar arguments to those described in [54], [55]. Essentially, (11.46) gives the minimum set of power allocations that satisfy the short-term power constraints such that total mutual information is greater than the rate requirement. The inclusion of threshold s^* ensures that the long-term power constraint is also satisfied. The larger s^* is, the smaller the outage probability will be. Since the solution $(\hat{\rho}, \hat{\rho})$ from (11.46) gives the minimum sum power to satisfy the rate constraint, then this solution also gives the maximum s^* , which in turn minimizes the outage probability. \square

We see that in general, the solution to (11.44) requires the solution to (11.46), which we have further decomposed into a number of smaller optimization sub-problems in Lemma 11.1. Ultimately a tractable solution to the overall problem is only possible if these smaller sub-problems are convex. In this case, these sub-problems are solvable via the Karush–Kahn–Tucker (KKT) conditions [56]. However, this is only possible if and only if the input–output mutual information of each component channel is a concave function of the input power. In this case, the KKT conditions yield a solution that is expressed as a function of the derivative of the mutual information with respect to the input power [57], [54], which is equal to the minimum mean-squared error (MMSE) for the signal-independent Gaussian noise model [43]. Whilst this is the case for the RF channel, as discussed in Section 11.3.1, the mutual information of the FSO channel is not concave in general. Moreover, to enforce concavity, we must either employ time-sharing or optimize the FSO input distribution. Alternatively, we may consider suboptimal power allocation. For example, in [26] a suboptimal scheme was proposed for the signal-independent Gaussian photodetection model (11.14) with an equi-probable signalling, by replacing the resulting non-convex problem (11.46) with a convex one. It was then proven that this suboptimal strategy still achieves the optimal outage diversity. However, in this section, we only consider optimal power allocation and therefore assume either time-sharing or optimal PAM signalling is employed to ensure a concave mutual information function.

11.5.2 Asymptotic Analysis

The asymptotic outage performance of optimal power allocation for discrete-input block-fading AWGN channels was analyzed by Nguyen *et al.* in [54], [58]. In particular, from [54, Prop. 3], if the PAPRs α_{fso} and α_{rf} are finite, then the SNR exponent will be the same as the CSIR case given in Theorems 11.1 and 11.2. When there are no PAPR constraints then the SNR exponent of the optimal power allocation strategy is [58, Th. 1]

$$d_{\text{csit}}^{(1)} = \begin{cases} \infty & d_{\text{csir}}^{(1)} > 1 \\ \frac{d_{\text{csir}}^{(1)}}{1-d_{\text{csir}}^{(1)}} & d_{\text{csir}}^{(1)} < 1, \end{cases} \quad (11.52)$$

where $d_{\text{csir}}^{(1)}$ is the SNR exponent for the CSIR case.

The implications of (11.52) are described as follows. When $d_{\text{csit}}^{(1)} = \infty$, then the outage probability curve will be vertical at a certain threshold of average power, i.e. the hybrid system is able to maintain a constant level of instantaneous input–output mutual information. The threshold at which this occurs is referred to as the *delay-limited capacity* of the system [59]. Note that if $d_{\text{csir}}^{(1)} = 1$ in (11.52) then $d_{\text{csit}}^{(1)} = \infty$, however, the outage curve will not show a vertical behaviour, nor will it converge to a constant slope when plotted on a log-log scale [54]. When the PAPRs are finite, the short-term power constraints introduce an error floor with a slope equal to the CSIR case. The height of the error floor is dependent on α_{fso} and α_{rf} [54].

11.5.3 Numerical results

To demonstrate the benefits of power allocation and its asymptotic behavior, we simulated the optimal power allocation strategies with OOK and gamma-gamma scintillation with parameters $\bar{a} = \bar{b} = 2$ for the FSO link and 64-QAM and exponential scintillation for the RF link. In our simulations, we have chosen the following parameters $\rho = \hat{\rho} = 0.5$, $A = B = 1$ and $N = 12$.

Figure 11.6(a) shows the outage probability with and without power allocation, for several values of the peak-to-average power ratio $\alpha = \alpha_{\text{fso}} = \alpha_{\text{rf}} = 5, 10, 15$, and 20 dB and $R = 9$. Optimal OOK signalling is assumed for power allocation (solid lines) while equiprobable OOK is used for uniform power allocation (dashed line). Note also that, as predicted by Eq. (11.52), since $d_{\text{csir}}^{(1)} = 2$, the outage curve without peak power restrictions will show a vertical behavior. This can be seen in the figure (thick solid curve), for $P_{\text{av}} > 24$ dB outages are completely removed. Remark first the order of magnitude of the gains. We see that there is a power saving of more than 20 dB compared to uniform power allocation to achieve 10^{-5} outage probability. When peak power constraints are introduced, as expected, we see that an error floor is introduced with the same slope as the CSIR case. The floor shifts down in probability as the peak-to-average power ratio increases.

Figure 11.6(b) shows the outage probability for a fixed peak-to-average power ratio $\alpha = 15$ dB and $R = 3, 9$, and 15. The figure compares uniform power allocation and optimal power allocation for time-sharing OOK and optimal OOK. For $R = 3$, we observe nearly 4 dB difference between the optimal scheme and time-sharing. Remark that this difference becomes smaller for higher rates, eventually disappearing. This is due to the fact that the operating point is exactly in a region where the optimal mutual information and the time-sharing one are nearly the same.

11.6 Conclusions and summary

We proposed a simple hybrid FSO/RF channel model based on parallel block fading channels. This hybrid model takes into account differences in signalling rates and fading effects typically experienced by the component channels involved. Under this framework, we examined the information-theoretic limits of the hybrid channel. In particular,

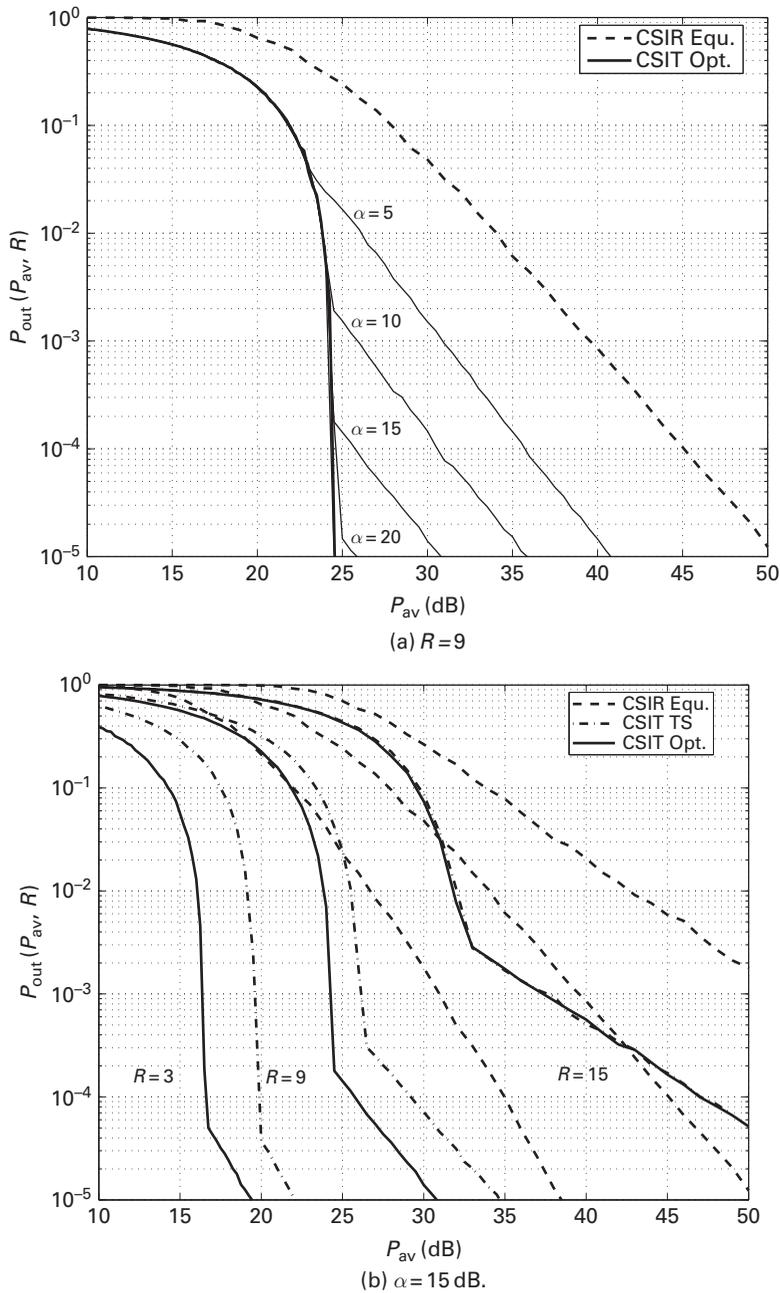


Figure 11.6 Outage performance of the hybrid FSO/RF channel with power adaptation: (a) compares the outage performance of a rate $R = 9$ system with uniform power allocation and equi-probable OOK signalling (dashed) to a system with optimal power allocation and optimal OOK signalling (solid), results for various PAPR are also shown, where $\alpha = \alpha_{\text{fso}} = \alpha_{\text{rf}}$; (b) compares optimal power allocation (with $\alpha = 15$ dB) for optimal OOK signalling (solid), equi-probable OOK with time-sharing (dot-dashed), and uniform power allocation with equi-probable signalling (dashed). Other system parameters included, $\rho = \hat{\rho} = 0.5$, $A = B = 1$, $N = 12$, OOK FSO and 64QAM RF, FSO gamma-gamma scintillation with parameters $a = b = 2$ and RF exponential scintillation.

we studied the outage performance in the asymptotic high-power regime by analyzing the outage exponent. When CSI is only available at the receiver, in the general case, the exponent is not available in closed form. Instead, we derived simple expressions from which it can be computed numerically. These outage exponents are related to how optimal codes should be designed, and therefore are useful for practical code design. When CSI is also available at the transmitter, we derived the optimal power allocation scheme that minimizes the outage probability subject to peak and average power constraints. Our results indicate that significant power savings (on the order of tens of dBs) are achievable using optimal power allocation compared to uniform power allocation.

Appendix A Kullback–Leibler divergence between Poisson and Gaussian distributions

THEOREM 11.4 *Let $p(x) = \sum_{k=0}^{\infty} \frac{\Lambda^k}{k!} e^{-\Lambda} \delta(x-k)$ denote the continuous domain conversion of the discrete Poisson distribution with mean Λ . Let $q(x) = \frac{1}{\sqrt{2\pi\Lambda}} e^{-\frac{1}{2\Lambda}(x-\Lambda)^2}$ be a Gaussian density with mean and variance Λ . Then*

$$\frac{1}{12}\Lambda^{-1} - \frac{5}{24}\Lambda^{-2} - \mathcal{O}(\Lambda^{-3}) \leq D(p||q) \leq +\frac{1}{12}\Lambda^{-1} + \frac{33}{24}\Lambda^{-2} + \mathcal{O}(\Lambda^{-3}) \quad (11.53)$$

Proof The proof follows by first considering [23]

$$D(p||q) = -H(p) - \int_{-\infty}^{\infty} p(x) \log q(x) dx, \quad (11.54)$$

where $H(p)$ is the entropy of a Poisson random variable with mean Λ , which can be bounded as follows [60]

$$-\frac{33}{24}\Lambda^{-2} - \mathcal{O}(\Lambda^{-3}) \leq H(p) - \frac{1}{2} \log(2\pi e\Lambda) + \frac{1}{12}\Lambda^{-1} \leq \frac{5}{24}\Lambda^{-2} + \mathcal{O}(\Lambda^{-3}).$$

Using properties of the mean and variance of a Poisson random variable, it is straightforward to show the cross-entropy term,

$$-\int_{-\infty}^{\infty} p(x) \log q(x) dx = \sum_{k=0}^{\infty} \frac{\Lambda^k}{k!} e^{-\Lambda} \left[\frac{1}{2} \log 2\pi\Lambda - \frac{(k-\Lambda)^2}{2\Lambda} \right] = \frac{1}{2} \log(2\pi e\Lambda). \quad (11.55)$$

The proof of (11.53) then follows. \square

Appendix B Derivative of the mutual information for discrete-input Poisson channels

THEOREM 11.5 ([44]) *Let $I_{\mathcal{X}}(\gamma, \lambda)$ denote the input–output mutual information of the discrete-input Poisson channel (11.7) in bits per channel use. Then*

$$\frac{\partial}{\partial \gamma} I_{\mathcal{X}}(\gamma, \lambda) = \sum_{j=1}^J \mathbb{E} \left[x_j \log_2 \left(\frac{\gamma x_j + \lambda}{\mathbb{E}[\gamma x_j + \lambda | \mathbf{y}]} \right) \right], \quad (11.56)$$

where $\mathbb{E}[\gamma x_j + \lambda | \mathbf{y}]$ is the conditional mean estimate.

LEMMA 11.2 *The conditional mean estimate in (11.56) is given by*

$$\mathbb{E}[\gamma \mathbf{x} + \lambda | \mathbf{y}] = \sum_{\mathbf{x} \in \mathcal{X}} (\gamma \mathbf{x} + \lambda) \left(1 + \sum_{\substack{\mathbf{x}' \in \mathcal{X} \\ \mathbf{x}' \neq \mathbf{x}}} \frac{\Pr\{\mathbf{x}'\}}{\Pr\{\mathbf{x}\}} \prod_{j=1}^J \left(\frac{1 + \frac{\gamma}{\lambda} x'_j}{1 + \frac{\gamma}{\lambda} x_j} \right)^{y_j} e^{-\gamma(x'_j - x_j)} \right)^{-1} \quad (11.57)$$

Proof

$$\mathbb{E}[\gamma \mathbf{x} + \lambda | \mathbf{y}] = \sum_{\mathbf{x} \in \mathcal{X}} (\gamma \mathbf{x} + \lambda) \Pr\{\mathbf{x} | \mathbf{y}\} \quad (11.58)$$

$$= \sum_{\mathbf{x} \in \mathcal{X}} (\gamma \mathbf{x} + \lambda) \frac{\Pr\{\mathbf{y} | \mathbf{x}\} \Pr\{\mathbf{x}\}}{\sum_{\mathbf{x}' \in \mathcal{X}} \Pr\{\mathbf{y} | \mathbf{x}'\} \Pr\{\mathbf{x}'\}} \quad (11.59)$$

$$= \sum_{\mathbf{x} \in \mathcal{X}} \frac{\gamma \mathbf{x} + \lambda}{1 + \sum_{\substack{\mathbf{x}' \in \mathcal{X} \\ \mathbf{x}' \neq \mathbf{x}}} \frac{\Pr\{\mathbf{y} | \mathbf{x}'\} \Pr\{\mathbf{x}'\}}{\Pr\{\mathbf{y} | \mathbf{x}\} \Pr\{\mathbf{x}\}}} \quad (11.60)$$

$$= \sum_{\mathbf{x} \in \mathcal{X}} (\gamma \mathbf{x} + \lambda) \left(1 + \sum_{\substack{\mathbf{x}' \in \mathcal{X} \\ \mathbf{x}' \neq \mathbf{x}}} \frac{\Pr\{\mathbf{x}'\}}{\Pr\{\mathbf{x}\}} \prod_{i=1}^Q \left(\frac{1 + \frac{\gamma}{\lambda} x'_i}{1 + \frac{\gamma}{\lambda} x_i} \right)^{y_i} e^{-\gamma(x'_i - x_i)} \right)^{-1} \quad (11.61)$$

□

THEOREM 11.6 *For the special case of Q -ary PPM with equi-likely input symbols,*

$$\frac{\partial}{\partial \gamma} I_{\mathcal{X}_{\text{PPM}}}(\gamma, \lambda) = \log_2 Q - I_{\mathcal{X}_{\text{PPM}}}(\gamma, \lambda) - \mathbb{E} \log_2 \left(1 + \sum_{q=2}^Q \left(1 + \frac{\gamma}{\lambda} \right)^{y_q - y_1 - 1} \right). \quad (11.62)$$

Proof From Lemma 11.2 we have

$$\mathbb{E}[\gamma x_i + \lambda | \mathbf{y}] = \sum_{q=1}^Q \frac{\gamma e_{q,i} + \lambda}{\sum_{j=1}^Q \left(1 + \frac{\gamma}{\lambda} \right)^{y_j - y_q}} \quad (11.63)$$

$$= \left(\sum_{j=1}^Q \left(1 + \frac{\gamma}{\lambda} \right)^{y_j - y_i} \right)^{-1} \left(\gamma + \lambda \sum_{q=1}^Q \left(1 + \frac{\gamma}{\lambda} \right)^{y_q - y_i} \right). \quad (11.64)$$

Due to the symmetry of PPM, we need only consider $\mathbf{x} = \mathbf{e}_1$ was transmitted. Hence combining (11.64) and (11.56) we have

$$\begin{aligned}
\frac{\partial}{\partial \gamma} I_{\mathcal{X}_{\text{PPM}}}(\gamma, \lambda) &= \mathbb{E} \left[\log_2 \left(\frac{\gamma + \lambda}{\mathbb{E}[\gamma x_1 + \lambda |y]} \right) \right] \\
&= \mathbb{E} \log_2 \left(\sum_{j=1}^Q \left(1 + \frac{\gamma}{\lambda} \right)^{y_j - y_1} \right) - \mathbb{E} \log_2 \left(\frac{\gamma}{\gamma + \lambda} + \frac{\lambda}{\gamma + \lambda} \sum_{q=1}^Q \left(1 + \frac{\gamma}{\lambda} \right)^{y_q - y_1} \right) \\
&= \log_2 Q - I_{\mathcal{X}_{\text{PPM}}}(\gamma, \lambda) - \mathbb{E} \log_2 \left(\frac{\gamma}{\gamma + \lambda} + \frac{\lambda}{\gamma + \lambda} \sum_{q=1}^Q \left(1 + \frac{\gamma}{\lambda} \right)^{y_q - y_1} \right) \\
&= \log_2 Q - I_{\mathcal{X}_{\text{PPM}}}(\gamma, \lambda) + \log_2 \left(1 + \frac{\gamma}{\lambda} \right) - \mathbb{E} \log_2 \left(1 + \frac{\gamma}{\lambda} + \sum_{q=2}^Q \left(1 + \frac{\gamma}{\lambda} \right)^{y_q - y_1} \right) \\
&= \log_2 Q - I_{\mathcal{X}_{\text{PPM}}}(\gamma, \lambda) - \mathbb{E} \log_2 \left(1 + \sum_{q=2}^Q \left(1 + \frac{\gamma}{\lambda} \right)^{y_q - y_1 - 1} \right). \tag{11.65}
\end{aligned}$$

□

Acknowledgments

This work was supported by the Sir Ross and Sir Keith Smith Fund, Cisco Systems, the Australian Research Council under RN0459498, DP0558861, DP088160, and the Royal Society under International Travel Grants 2009/R2 and 2009/R4.

References

- [1] L. C. Andrews and R. L. Phillips, *Laser Beam Propagation through Random Media*, SPIE Press, 2nd edition, 2005.
- [2] N. Letzepis and A. Guillén i Fàbregas, “Outage probability of the Gaussian MIMO free-space optical channel with PPM,” *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3682–3690, Dec. 2009.
- [3] N. Letzepis and A. Guillén i Fàbregas, “Outage probability of the free-space optical channel with doubly stochastic scintillation,” *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 2899–2902, Oct. 2009.
- [4] N. Letzepis, I. Holland, and W. Cowley, “The Gaussian free space optical MIMO channel with Q -ary pulse position modulation,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, May 2008.
- [5] N. Cvijetic, S. G. Wilson, and M. Brandt-Pearce, “Performance bounds for free-space optical MIMO systems with APD receivers in atmospheric turbulence,” *IEEE J. Select. Areas. Commun.*, vol. 26, no. 3, pp. 3–12, April 2008.

- [6] S. M. Navidpour, M. Uysal, and M. Kavehrad, "BER performance of free-space optical transmission with spatial diversity," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2813–2819, August 2007.
- [7] S. G. Wilson, M. Brandt-Pearce, Q. Cao, and J. H. Leveque, "Free-space optical MIMO transmission with Q -ary PPM," *IEEE Trans. Commun.*, vol. 53, no. 8, pp. 1402–1412, Aug. 2005.
- [8] K. Chakraborty, "Capacity of the MIMO optical fading channel," in *Proc. IEEE Int. Symp. Inform. Theory*, Adelaide, Sept. 2005, pp. 530–534.
- [9] E. J. Lee and V. W. S. Chan, "Part 1: optical communication over the clear turbulent atmospheric channel using diversity," *J. Select. Areas Commun.*, vol. 22, no. 9, pp. 1896–1906, Nov. 2005.
- [10] S. M. Haas and J. H. Shapiro, "Capacity of wireless optical communications," *IEEE J. Select. Areas Commun.*, vol. 21, no. 8, pp. 1346–1356, Oct. 2003.
- [11] F. Nadeem, B. Flecker, E. Leitgeb, M. S. Awan, and T. Javornik, "Comparing the fog effects on hybrid network using optical wireless and GHz links," in *Proc. Int. Symp. Commun. Sys., Networks and Digital Signal proc.*, Jun. 2008, pp. 278–282.
- [12] H. Wu, B. Hamzeh, and M. Kavehrad, "Achieving carrier class availability of FSO link via complementary RF link," in *Proc. 38th Asilomar Conf. Signals, Systems and Computers*, Oct. 2004.
- [13] Z. Jia, F. Ao, and Q. Zhu, "BER performance of the hybrid FSO/RF attenuation system," in *Int. Symp. Anten., Prop. & EM Theory*, Sep. 2006.
- [14] T. Kamalakis, I. Neokosmidis, A. Tsipouras, S. Pantazis, and I. Andrikopoulos, "Hybrid free space optical/millimeter wave outdoor links for broadband wireless access networks," in *Proc. Int. Symp. Personal, Indoor and Mobile Radio Commun.*, Aug. 2007.
- [15] S. Vangala and H. Pishro-Nik, "A highly reliable FSO/RF communication system using efficient codes," in *Proc. IEEE Global Commun. Conf.*, 2007.
- [16] H. Wu, B. Hamzeh, and M. Kavehrad, "Availability of airbourne hybrid FSO/RF links," in *Proc. SPIE*, 2005, vol. 5819.
- [17] T. Mouldsley and E. Vilar, "Experimental and theoretical statistics of microwave amplitude scintillations on satellite down-links," *IEEE Trans. Anten. Propag.*, vol. 30, no. 6, pp. 1099–1106, Jan. 1982.
- [18] C. E. Mayer, B. E. Jaeger, R. K. Crane, and X. Wang, "Ka-band scintillations: measurements and model predictions," *Proc. IEEE*, vol. 85, no. 6, pp. 936–945, Jun. 1997.
- [19] M. S. Alouini, S. A. Borgsmiller, and P. G. Steffes, "Channel characterization and modeling for Ka-band very small aperture terminals," *Proc. IEEE*, vol. 85, no. 6, pp. 981–997, Jun. 1997.
- [20] S. A. Khan, A. N. Tawfik, and C. J. Gibbins, "Short-term amplitude scintillations at 97 GHz on 6.5 km urban link," *Electron. Lett.*, vol. 36, no. 19, pp. 1654–1656, Aug. 2000.
- [21] S. A. Khan, A. N. Tawfik, B. C. Gremont, and C. J. Gibbins, "Long-term amplitude scintillations at 97 GHz on 6.5 km urban link," *Electron. Lett.*, vol. 36, no. 16, pp. 1425–1426, Jul. 2000.
- [22] I. B. Djordjevic, B. Vasic, and M. A. Neifeld, "Power efficient LDPC-coded modulation for free-space optical communication over the atmospheric turbulence channel," in *Proc. Conf. Optical Fiber Commun. and Nat. Fiber Optic Engineers Conf.*, Feb. 2007.
- [23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, 1991.
- [24] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, May 2003.

-
- [25] A. Guillén i Fàbregas and G. Caire, “Coded modulation in the block-fading channel: Coding theorems and code construction,” *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 262–271, Jan. 2006.
- [26] N. Letzepis, K. D. Nguyen, A. Guillén i Fàbregas, and W. G. Cowley, “Outage analysis of the hybrid free-space optical and radio-frequency channel,” *IEEE J. Select. Areas Commun. (special issue on optical wireless communications)*, vol. 27, no. 9, pp. 1709–1719, Dec. 2009.
- [27] J. Proakis, *Digital Communications*, McGraw-Hill, 1995.
- [28] R. M. Gagliardi and S. Karp, *Optical communications*, John Wiley & Sons, Inc., 1995.
- [29] L. H. Ozarow, S. Shamai, and A. D. Wyner, “Information theoretic considerations for cellular mobile radio,” *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378, May 1994.
- [30] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: information-theoretic and communications aspects,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [31] A. T. Asyari and A. Guillén i Fàbregas, “Nearest neighbour decoding in block-fading channels with imperfect CSIR,” in *IEEE Inf. Theory Workshop, Taormina, Italy*, 2009.
- [32] K. D. Nguyen, N. Letzepis, A. Guillén i Fàbregas, and L. K. Rasmussen, “Outage diversity of MIMO block-fading channels with causal channel state information,” in *IEEE Int. Symp. Inf. Theory, Austin, TX, USA*, 2010.
- [33] S. Dolinar, D. Divsalar, J. Hamkins, and F. Pollara, “Capacity of pulse-position modulation (PPM) on Gaussian and Webb channels,” *JPL TMO Progress Report 42-142*, Aug. 2000, URL: lasers.jpl.nasa.gov/PAPERS/OSA/142h.pdf.
- [34] S. J. Dolinar, J. Hamkins, B. E. Moision, and V. A. Vilnrotter, *Optical Modulation and Coding*, chapter 4, John Wiley & Sons, Inc., 2006.
- [35] K. Chakraborty, S. Dey, and M. Franceschetti, “Outage capacity of MIMO Poisson fading channels,” *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4887–4907, Nov. 2008.
- [36] E. J. Gumbel, “The limiting form of Poisson’s distribution,” *Phys. Rev.*, vol. 60, p. 689, 1941.
- [37] E. Rodgers, “Probable error for Poisson distributions,” *Phys. Rev.*, no. 57, pp. 735–737, April 1940.
- [38] X. Zhu and J. M. Kahn, “Free-space optical communication through atmospheric turbulence channels,” *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1293–1300, Aug. 2002.
- [39] X. Zhu and J. M. Kahn, “Mitigation of turbulence-induced scintillation noise in free-space optical links using temporal-domain detection techniques,” *IEEE Photon. Tech. Lett.*, vol. 15, no. 4, pp. 623–625, April 2003.
- [40] M. Uysal, J. Li, and M. Yu, “Error rate performance analysis of coded free-space optical links over gamma-gamma atmospheric turbulence channels,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1229–1233, June 2006.
- [41] H. G. Sandalidis and T. A. Tsiftsis, “Outage probability and ergodic capacity of free-space optical links over strong turbulence,” *IET Electron. Lett.*, vol. 44, no. 1, pp. 46–47, Jan. 2008.
- [42] B. He and R. Schober, “Bit-interleaved coded modulation for hybrid RF/FSO systems,” *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3753–3763, Dec. 2009.
- [43] D. Guo, S. Shamai, and S. Verdú, “Mutual information and minimum mean-square error in Gaussian channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
- [44] D. Guo, S. Shamai, and S. Verdú, “Mutual information and conditional mean estimation in Poisson channels,” *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1837–1849, May 2008.

- [45] S. Shamai, "Capacity of a pulse amplitude modulated direct detection photon channel," *IEE Proc.*, vol. 137, no. 6, pp. 424–430, 1990.
- [46] J. P. Gordon, "Quantum effects in communication systems," *Proc. IRE*, vol. 50, pp. 1898–1908, 1962.
- [47] G. Ungerboeck, "Channel coding with multilevel signals," *IEEE Trans. Inf. Theory*, vol. 28, Jan. 1988.
- [48] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.
- [49] E. Malkamaki and H. Leib, "Coded diversity on block-fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 771–781, March 1999.
- [50] R. Knopp and P. Humblet, "On coding for block fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 1643–1646, July 1999.
- [51] E. Malkamaki and H. Leib, "Evaluating the performance of convolutional codes over block fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1643–1646, Jul. 1999.
- [52] T. T. Kim and G. Caire, "Diversity gains of power control with noisy CSIT in MIMO channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1618–1626, Apr. 2009.
- [53] T. T. Kim, K. D. Nguyen, and A. Guillén i Fàbregas, "Coded modulation with mismatched csit over mimo block-fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, Nov. 2010.
- [54] K. D. Nguyen, A. Guillén i Fàbregas, and L. K. Rasmussen, "Power allocation for block-fading channels with arbitrary input constellations," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2514–2523, May 2009.
- [55] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1489, July 1999.
- [56] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [57] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, July 2006.
- [58] K. D. Nguyen, A. Guillén i Fàbregas, and L. K. Rasmussen, "Outage exponents of block-fading channels with power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2373–2381, May 2010.
- [59] S. V. Hanly and D. N. C. Tse, "Multiaccess fading channels. II. delay-limited capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2816–2831, Nov. 1998.
- [60] J. A. Adell, A. Lekuona, and Y. Yu, "Sharp bounds on the entropy of the Poisson law and related quantities," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2299–2306, May 2010.

Part IV

Applications

12 Quantum key distribution

Rupert Ursin, Nathan Langford and Andreas Poppe

12.1 Motivation

The ability to guarantee security and privacy in communication are critical factors in encouraging people to accept and trust new tools and methods for today's information-based society (e.g. in eCommerce or eHealth) and for future services (e.g. eGovernment, eVoting). Cybercrime already hinders these new possibilities by creating widespread mistrust in these new services. Another problem is that a (perhaps unpublicized) breakthrough in mathematics or computer science could completely compromise current state-of-the-art encryption methods overnight, even those which are the basis of all existing internet banking transactions. This is because the security of cutting-edge public-key cryptography (conventional cryptography) relies on the computational difficulty of certain mathematical tasks, meaning that conventional cryptography alone can neither provide any evidence of eavesdropping nor guarantee strong security. The trend towards faster electronics provides the ability to handle longer keys, thus providing better security, but also increases the possibility to break keys. This also has another worrying effect: even using today's most advanced available security technology to protect data during communication with best practice and without making any mistakes, the ongoing improvements in technology (which hence require longer key lengths to ensure security) mean that today's transmitted data will not be secure for more than a few years. Thus, while encrypted data might be safe today, it can still be stored by the adversary now and then broken later on once the future technology has caught up with today's encryption.

Alternatively, modern quantum cryptography has created a new paradigm for cryptographic communication, which provides strong, long-term security and incontrovertible evidence of any attempted eavesdropping which is based on theoretically and experimentally proven laws of nature. This technique, called quantum key distribution (QKD), generates a symmetrical classical bit string using the correlations of measurements on quantum systems and has already developed into a mature technology providing commercial products capable of everyday use [1–3]. The main hurdle for quantum communication is that, with present fiber and detector technologies, terrestrial QKD links are limited to distances of just over 100 km, well within reach of how far someone could travel in a short time to simply deliver the information in person. In the future, however,

it will be possible to extend the distances spanned by individual fiber-based QKD links by using repeater nodes. These individual QKD links could then be combined to create larger and more complex QKD networks which will allow many different combinations of users to be connected over the same infrastructure, as has been shown at the recent SECOQC and UQCC demonstrations [4, 5]. The economic benefits of such an interlinking network approach to QKD will be most apparent in a typical metropolitan scenario, where many potential users are likely to be located in a relatively small area, each wanting to be able to communicate securely with many different partners.

To further extend the distance for quantum communication, these metropolitan areas would have to be connected and an interesting possibility for achieving this is to use optical free-space links between (moving) high-altitude platforms (HAPs; a repeater node located high above the turbulent atmosphere, keeping a roughly stationary position over a metropolitan area). Some QKD systems already make use of free-space links [6–9], and experiments have been performed outside the protected environment of the laboratory, under real-world conditions over distances up to 144 km [10–12], the distance limited mainly by the available terrestrial test locations. Alternatively, in urban areas, HAPs could be used to distribute secret keys over distances potentially more than 500 km, far exceeding the feasible distances for today’s fiber technology. Systems using moving platforms, however, would have to tackle additional challenges to terrestrial demonstrations. Ultimately, a satellite-based quantum communication terminal in a medium-earth orbit (MEO) would even be able to distribute single photons between continents [13–18] and this is also being published in the framework of progress reports of proposals [19–21]. The pointing acquisition and tracking of the optical links required for such a scenario are a well-developed technology in regular use in satellite missions like OISETS (Japan) or TerraSarX (Germany) and ARTEMIS (ESA).

Using a combination of terrestrial, HAP, and satellite-based point-to-point links, global quantum communication is a feasible target for current state-of-the-art technology. Extending quantum communication and QKD to a global scale would have a substantial impact on communication technology of the future. Moreover, the idea of performing QKD from air- or space-based platforms has a particular appeal to organizations which require highest possible security to protect and authenticate communication over global distances. Given the long-term security of messages transmitted using keys distributed by QKD, a global QKD system will be highly marketable.

12.2 Security considerations of QKD

Quantum key distribution has an important advantage over its conventional counterpart of public-key cryptography, namely that its security is based on fundamental laws of quantum physics (e.g., the no-cloning theorem and the superposition principle) [1–3]. The output of the QKD protocol for both communicating parties is a symmetrical (identical) classical key which generally takes the form of a random bit string. Both parties subsequently use the same key to e.g. encrypt data to secure transmission over conventional communication lines. If this key is used only once (a one-time-pad protocol) the

combined procedure of QKD and one-time-pad encryption is proven to be information-theoretically secure (ITS): that is, without knowledge of the key, a ciphertext would provide no more information about the content of the plaintext than randomly guessing. In other words, from an encrypted message the adversary could extract neither the message nor even a fraction of the key, even with unlimited computational power.

Conventional (asymmetrical) public-key exchange depends on unverifiable assumptions that the computational complexity of the mathematical techniques involved will deny an eavesdropper the ability to learn the contents of encrypted messages. QKD, however, is based on the fact that any attempt to intercept and read out the encoded quantum information will either damage or destroy it, decreasing the correlations observed by the communicating parties. As a result, such an eavesdropping attempt introduces errors in the QKD protocol and can thus be detected. Indeed, if initially strong correlations decrease to the point where they can be explained by classical laws, then the communicating parties can assume that their communication has been compromised in some way. If, on the other hand, they still observe correlations beyond the reach of classical physics, then the laws of physics guarantee that they can extract a secret key which will be completely secure from all possible eavesdroppers, even those who might hypothetically have access to unlimited computational power and resources, now and into the far future – even future quantum computers!

Quantum cryptography needs to be compared with the two different types of conventional cryptography: secret-key (also called symmetric) cryptography and public-key (asymmetric) cryptography, both of which can operate over arbitrary distances (a striking advantage of conventional key distribution). Secret-key cryptography (e.g., Advanced Encryption Standard (AES)) typically uses short keys (approved sizes by NIST: 128, 192, and 256 bits) to encrypt nearly arbitrarily large messages for high-speed communication on the fly and with basically no additional latency. From time to time, these secret keys are then changed. Public-key cryptography, on the other hand, provides both a widely used means to generate the secret keys for symmetric cryptography (e.g. Diffie–Hellman key exchange protocol), and the means to generate public keys to encrypt/decrypt messages directly (e.g. RSA, named after authors Rivest, Shamir, and Adleman). Such asymmetric cryptography is not widely used since it requires much more computational power than the symmetric version. By contrast, QKD provides the means to generate secret keys with generation rates over short distances from kbit/s up to the recently achieved record values of Mbit/s, but dropping with distance to only a few bit/s for a 100 km QKD link distance. QKD is the slow-but-ITS analogue to the Diffie–Hellman protocol.

The keys distributed by QKD can now be used together with well-established methods from conventional cryptography. The highest security (albeit with low rates) is achieved by using QKD keys in combination with secret key one-time-pad encryption, which is itself ITS. However, an economical way of achieving the usual high data rates while still increasing practical security beyond the currently deployed crypto systems, is to provide frequently changing seed keys for non-ITS secret-key encryption (e.g. AES) using QKD key generation. By doing this, the duration of the seed key's validity, which are typically maintained for months, years, or forever according to current practice, can

now be decreased below a minute. Here, the increased security is based on two facts [22]: that it would be harder to break one AES-session, because a collision of cyphertexts (via the birthday effect) could be excluded for a short ciphertext, and that in the (unlikely) event of a broken key, the leakage of data is limited only to this short period of time. It is worth mentioning that knowledge of one compromised key does not provide the adversary with any assistance in breaking future AES sessions based on subsequent QKD keys, because each key is generated independently. In all cases, the QKD-based key generation and the transmission of the encrypted message are fully independent, so any available public channel can be used to send the encrypted information.

At present, the only suitable quantum systems for long-distance quantum communication are photons in the spectral range of the visible or near infrared. Other carriers of quantum information such as atoms or ions, while well suited for research in quantum memories and quantum computation, do not seem to provide feasible solutions for quantum communication applications, at least in the near future. The quantum information should be encoded on a carrier consisting of only one single photon. A second photon which is accidentally produced alongside the first would normally carry the same information and could be used by the adversary to get increased information [1, 3]. Because generating a single photon on demand is technically extremely difficult, however, other sources approximating a single photon source are used in practice. The two most well-known schemes for QKD are based on weak coherent laser pulses (WCP) and entanglement, with the possible bit values being encoded in single-photon degrees of freedom [1]. In the following, we restrict ourselves to considering the polarization degree of freedom, because it is perhaps the best-studied and best-controlled option available, but in principle, time bins, phase and geometrical modes could also be used.

The next section provides a summary of the most important QKD protocols followed by a short description of how these are commonly implemented in an experiment. Subsequently, we will describe recent results from ESA-funded studies within this field. Finally, we will present a preliminary setup design for some proposed future experiments, including block diagrams for a HAP/satellite and the ground terminals and discuss specific problems for a quantum communication link (e.g. the automatic polarization tracking and the required synchronization of local time bases).

12.3 QKD protocols

12.3.1 BB84 protocol

The concept of the BB84 protocol was published by Bennett and Brassard in 1984 [23]. Typically, the photons are prepared in one of two orthogonal polarization states, associated with the bit values “0” and “1”, in one of two complementary (nonorthogonal) bases (e.g., the $0^\circ/90^\circ$ basis or $45^\circ/135^\circ$ basis), resulting in four possible linear polarization states 0° , 90° , 45° , and 135° . As shown schematically in Figure 12.1, Alice sends single photons randomly in one of these four states to Bob, who receives and analyzes them with a two-channel analyzer in one of the two complementary polarization bases

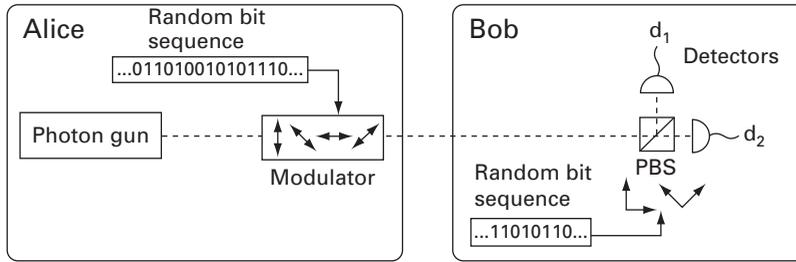


Figure 12.1 An illustration of the BB84 protocol. Alice prepares the light pulses randomly in one of four linear polarization states and sends them to Bob who measures, again randomly, in one of two complementary polarization bases in Detectors d behind a polarizing beam splitter (PBS). Whenever the preparation and the measurement occurred in the same basis, the measurement results (either “0” or “1”) are correlated and enter the cryptographic key while all the other measurement events are disregarded.

(i.e., the $0^\circ/90^\circ$ basis or the $45^\circ/135^\circ$ basis), randomly switching between the two. He records his measurement results together with the basis he used.

After measurement, quantum correlations turn into correlated measurement results stored in tables on computers and subsequent classical processing steps can then use these results to generate a secure key. First, during a *sifting step*, Bob communicates publicly with Alice and tells her which photons actually arrived at his measurement apparatus and the corresponding bases he used to analyze them. In return, Alice tells Bob if she used the same bases to prepare these photons, because only in these cases does Bob obtain the correct result (i.e., he only gets the “right” answer if he analyzes the photon in the same basis that Alice used to prepare the state). Assigning the binary value “0” to the linear polarization states 0° and 45° and “1” to the states 90° (135°) should in principle leave Alice and Bob with an identical set of “0”s and “1”s, although, as in any practical system, there will of course be some errors. This resulting key is called the sifted key.

There are various quantum attacks that one can imagine the adversary might use, but all are in common that an unavoidable trace is left in the measurement results [1]. The most intuitive strategy is to intercept the communication, measure the photons, and re-send them in the observed state. However, due to the superposition principle, some of the re-sent photons will be observed wrongly by Bob and will introduce errors in the QKD protocol which can be characterized by the quantum bit error ratio (QBER). As far as Alice and Bob are concerned, these errors are indistinguishable from errors which may result from their own experimental imperfections or in the transmission fidelity of the quantum channel. All errors therefore need to be corrected during the second important classical processing step of the QKD protocol, known as the *error-correction* step, to yield symmetric bit strings called the error-corrected key. The key point for QKD security is that the amount of error in the generated key (QBER) provides an upper bound on the information an eavesdropper might have gained assuming the best theoretically feasible attack. In practical systems, although experimental imperfections and

detector noise may also contribute to the QBER, it is important to assume that all errors are caused by the adversary to maintain the strict security guarantees provided by the laws of physics are not compromised by any additional assumptions about the experiment. The final classical data processing step is called *privacy amplification*, where the secure key is further refined (and reduced in size) to the extent that no information could have leaked away from that shared by the communicating parties. Assuming that the QKD protocol is subjected to the best theoretically possible attacks, no secure key will be obtained from the privacy amplification step if the QBER exceeds a threshold of 11% – in such a situation, any resulting key could potentially be known to the adversary.

12.3.2 Decoy-state BB84 protocol

In practical realizations of the BB84 protocol, the photon source is usually not a (technologically challenging) true single-photon source, e.g., based on quantum dots or vacancies in an optical crystal, but rather just an attenuated pulsed laser as commonly used in telecom and laser industries [24, 25]. The pulses must therefore be attenuated to a mean photon number smaller than one photon per pulse, but even then it is still possible that a pulse contains extra photons (due to the Poissonian statistics of the number of photons in a given pulse) which would also be carrying the same externally encoded quantum information (see Figure 12.1). Such multiphoton pulses enable an eavesdropper to mount a very powerful attack – the so-called photon number splitting attack (PNS attack) – where one of the extra photons in a multiphoton pulse is extracted and measured by the adversary [26].

An effective method to counteract this attack is to use decoy states – states of different mean photon numbers which are randomly sent in place of the actual signal states [27, 28]. The eavesdropper cannot distinguish between signal and decoy pulses and thus cannot act differently on them. After transmission, the measurement results are monitored by Alice and Bob to uncover the action of the adversary. Since the signal states and the decoy states exhibit different photon-number statistics, any photon-number-dependent eavesdropping strategy (e.g., the PNS attack) will have different effects on the signal and decoy states. Alice and Bob can now independently compute the transmission probability for signal and decoy states and detect, with high probability, any photon-number-dependent attack.

12.3.3 Entanglement-based BB84 protocol

The BB84 QKD schemes described above are both typical prepare-and-measure schemes, where Alice prepares one of several possible initial states and Bob selects one of several possible measurement bases and gets a corresponding measurement result. In the entanglement-based version of this protocol ([29] BBM92, illustrated in Figure 12.2), both Alice and Bob receive photons and carry out a similar measurement procedure to Bob's in the BB84-protocol. If the two photons are entangled, quantum measurements in both $0^\circ/90^\circ$ and $45^\circ/135^\circ$ bases will then yield theoretically perfect

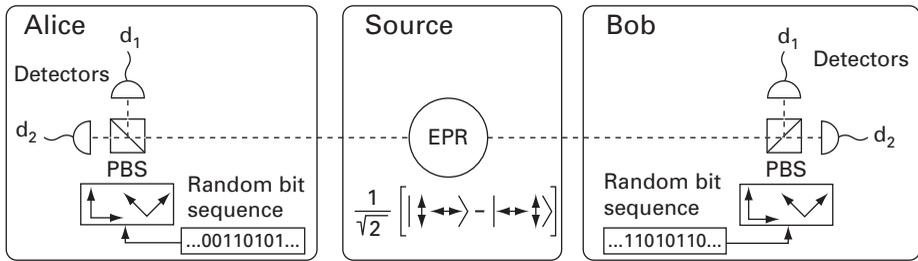


Figure 12.2 A schematic of the entanglement based QKD protocol. A source generates entangled photon pairs, usually in the state $|HH\rangle + |VV\rangle$ (meaning the polarization of the photons emitted in different directions are known to be the same, but the individual state is unknown), which are then sent to Alice and Bob. Both measure the incoming photons randomly in one of two complementary polarization bases in detectors d behind a polarizing beam splitter (PBS). The cryptographic key consists of those results which are obtained from measurements in the same basis. If the EPR source, which was first described by Einstein, Podolski, Rosen, emits a state $|VH\rangle - |HV\rangle$, as indicated in the figure, the ideal measurements are instead fully anticorrelated and one receiver needs to invert the bits.

correlations whenever they choose the same measurement basis [30]. The source of entangled photon pairs is located either as a third station in the middle [30] or perhaps, for practical reasons, at one of the two main stations [31].

Entangled photons have novel features not present in standard classical laser systems. One striking difference is that, individually, they have no well-specified polarization, but as soon as the first photon is measured, the behavior of the second photon depends on (is correlated to) the result of the measurement on the first photon. Clearly, non-entangled pairs of photons could be prepared in the four desired polarization states (e.g., $0^\circ 0^\circ$, $45^\circ 45^\circ$, $90^\circ 90^\circ$ or $135^\circ 135^\circ$), but if either photon is measured in the other basis from which it is not prepared (a choice made independently by Alice and Bob), the measurement results would then be uncorrelated. Moreover, even if the source was itself operated by the adversary, no information of the quantum states, measurement results or key is left at the source in the hands of Eve, because this would be reflected in the quality of the observed quantum correlations.

Afterwards, the subsequent protocol steps are nearly identical with the BB84 protocol, the only difference being that the quantum correlations are between two measurement outcomes rather than between a preparation and a measurement. Specifically, Alice and Bob publicly communicate which photons they detected and the corresponding measurement bases, discarding events where only one photon is detected, as well as those in which the photons were both detected, but using different measurement bases. Because a shared entangled state exhibits perfect polarization correlations, Alice's and Bob's results should be perfectly correlated. Again, this key might contain errors from experimental imperfections or eavesdropping attacks, so an error correction protocol step is used to remove the differences and exactly specify the QBER. This error rate is important for the last processing step of privacy amplification, which cuts the length of the key to the extent that the adversary loses all knowledge about the key.

12.4 Technical implementation of a free-space setup

This section describes the principal design issues for a free-space QKD apparatus (Section 12.4.1) and a specific design for a hybrid QKD link planned to be used in free-space experiments to HAPs or even satellites (Section 12.4.2). The final Section (12.4.3) describes some results of the most recent free-space experiments.

12.4.1 Basic design considerations

There are two main possible ways to send photonic information over long distances – through optical fibers or free space. The different properties of these two media give rise to different design considerations when trying to implement QKD in the real world. Although we will focus primarily on the free-space implementation here, we will briefly touch on some of these differences below. Some key practical issues for a practical QKD system are:

- **Optimal transmission wavelength:** Although photons are largely immune to most forms of decoherence (e.g. the scrambling of polarization), they are quite susceptible to loss, which can take many forms, such as absorption and scattering events. Effective loss can also be induced by effects such as mode distortion. For transmission in optical fibers, the optimal wavelengths for minimizing absorption loss are in the standard telecom band around 1550 nm (loss < 0.2 dB/km). In free space, on the other hand, transmission is more strongly affected by turbulence and scattering. However, free-space communication is obviously vital for any possibility of satellite-based communication.
- **Single-photon detection:** A detector capable of detecting the smallest possible bundle of light – a single photon – is a very sophisticated system made of the detector itself and the electronics required to transfer the arrival of the photon to an electrical pulse (e.g. TTL, NIM) with a steep training edge carrying the timing information. The most advanced single-photon detectors that can be used without specialist cryogenic equipment (an essential requirement for QKD applications) use avalanche photodiodes (APDs) operated at voltages above breakdown and quenching electronics. Of these, the best currently available single-photon detectors are silicon APDs which are mainly sensitive to visible and very near-infrared wavelengths (typical efficiency of 40–55%), with a design-dependent peak sensitivity for light from 550 nm to 680 nm (green to red for the human eye). Silicon APDs are therefore ideal for free-space communication in the red/near-infrared range, but absorption at these wavelengths in fiber is extremely high. For 1550 nm, Indium Gallium Arsenide (InGaAs) APDs provide a feasible option for single-photon counting, but the operational parameters, such as quantum efficiency, dark-count rate and timing jitter are far worse than for Silicon APDs.
- **Signal tracking and beam capture:** A laser pointing and tracking system is used to set up and maintain a stable link between the terminals. To maintain the transmittance of the quantum channel using terminals moving relative to each other (or with

changing atmospheric conditions between them), a closed-loop bi-directional corsage and fine pointing and tracking need to be implemented. To maximize this transmittance, the transmitter terminal should emit a diffraction-limited quantum signal. Likewise, for the receiver terminal, to limit the influence of stray light (causing an increased QBER), the field of view (FOV) should be kept as small as possible. If we consider as a worst case scenario, e.g., the movement in orbit of the ISS (orbit: 333×348 km, inclination 51.6°) relative to an optical ground station, the station will only be visible (above a 10° elevation) for about 5 minutes. This implies a very fast required movement of the telescope of around 1 deg/s both in elevation and azimuth, while the static pointing and dynamic tracking accuracies need to be in the range of up to $\sim \pm 1$ arcsec. Several telescopes are available in Europe and other countries in geodetic stations which have been designed to track GEO and also LEO satellites, with apertures of the order of 50 cm or larger. Besides these fixed optical ground stations, transportable versions are also being developed in Germany (DLR) and Japan (NICT) for ad-hoc installations in the field and at various global locations.

- **Polarization tracking and correction:** In polarization-encoded quantum communication using moving platforms, it is necessary to correct for any relative rotation of the polarization basis frame that the two terminals will experience as they move past each other. One conceivable solution might be to use a polarization-reference beacon. Due to the extremely low power of the quantum signal channel, separate beacon lasers (well isolated from the signal wavelength) are required. A polarization analyzer (or polarimeter) could detect the state of polarization of the beacon laser and this information could then be used in real time to compensate the polarization misalignment of the quantum channel. Alternatively, the polarization reference frame could be adjusted according to a precise computational prediction of the rotation angle from accurate knowledge of the platform's trajectory and rotation.

12.4.2 Overall system design

After the review of critical design parameters mentioned above, we now explain the main quantum subunit of a basic system design – consisting of the sources and receivers required to obtain quantum correlations (see Figure 12.3). The source module comprises either a weak laser source (Section 12.4.2.1) or an entangled photon source (12.4.2.2). The design of the optical antennas, beam pointing, acquisition and tracking does not differ from the design required from conventional optical communication techniques and is not included here and can be found in other relevant chapters of this book. Some propagation effects during transmission are revisited in Section 12.4.2.3 and the receiver is explained in Section 12.4.2.4. The timing synchronization to maintain a common time frame is topic of Section 12.4.2.5.

As mentioned earlier, distributing provably secure keys for communication is possible using various techniques: namely, using weak coherent states (potentially with decoy states) or entanglement sources. In this section we give a brief overview on the possible source designs.

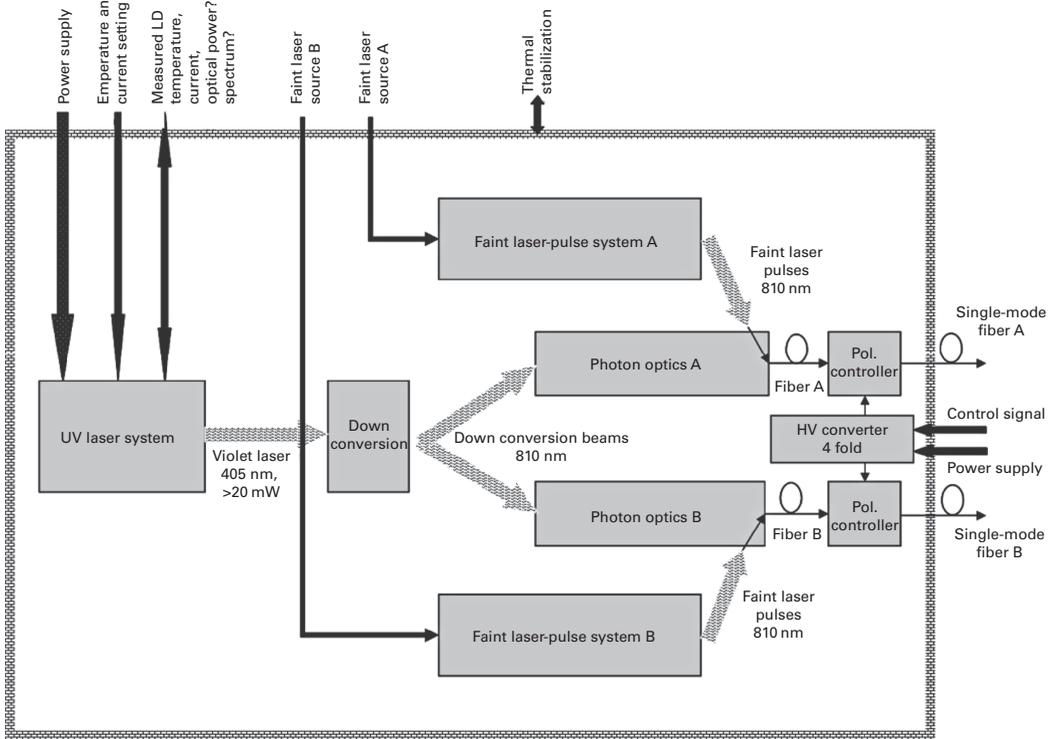


Figure 12.3 Overview of a possible quantum communication source module comprising an entangled photons source (UV laser system, down conversion as well as the required coupling of photon pair A and B into single-mode fibers), a weak laser option (BB84), and fiber polarization compensation for each of the photon channels. The fibers are required also for guiding the photons to the optical antennas. Additionally the required interfaces for the operation of the system is shown here as well as the photons pair output (fiber A and B).

12.4.2.1 Weak-coherent-laser-pulse-based systems

The weak-coherent-pulse scheme described above require the generation of the four different polarization states mentioned above (0° , 45° , 90° , and 135°) and this can be achieved using a range of different methods [28]. One example is to use four separate linearly polarized light beams from different laser diodes that are set to the desired polarization angles. The individual signals can then be combined (e.g., using a conical mirror) into a single spatial mode as defined by a single-mode optical fiber. By randomly driving only one of the laser diodes with a short electrical pulse, a laser pulse at the chosen polarization is generated. The diode must be driven quite strongly, well in the laser regime, in order to avoid unwanted fluctuations in the output power. Subsequently, therefore, the photon number must be decreased to single-photon levels by attenuation. The mean photon number per pulse can be fine-adjusted by varying the strength of the electric pulse and the attenuation of the optical signal. Usually a repetition rate of

singe-photon states of some MHz can be achieved, limited by the driving electronics of the laser diodes and the repetition rate of the detectors.

12.4.2.2 Entanglement-based systems

Entangled photon pairs are usually generated via spontaneous parametric down conversion (SPDC). An optical pump beam is incident on a birefringent crystal where nonlinear effects lead to the conversion of pump photons into pairs of correlated down-converted (i.e., lower-frequency) photons, usually called the signal and idler. Because of energy conservation, the sum of signal and idler energies equals the energy of a single pump photon, so the wavelengths of the down-converted photons are strongly anticorrelated. If the phase-matching conditions (momentum conservation) of the nonlinear crystal is fulfilled for the incident pump radiation and down-converted photons, constructive interference over the full crystal length allows an emission of the down-converted photons. Depending on the type of nonlinear interaction in the crystal, the phase matching can be designed to produce photon pairs with parallel polarization (type-I phase matching) or orthogonal polarization (type-II phase matching), allowing generation of quantum states $|VH\rangle - |HV\rangle$ or $|HH\rangle + |VV\rangle$ for type-II or type-I, respectively. These two systems are explained in more detail in the following.

One of the first schemes developed uses type-II SPDC in a BBO (β -BaB₂O₄: beta barium borate) crystal (a standard nonlinear optical crystal). An ultraviolet laser with a power of several mW is used to generate the signal and idler photons, which are emitted along separate cones that intersect along two straight lines [30, 31]. Photon pairs emitted along these lines are entangled in their polarization since it cannot be known to which cone they belong without measuring their polarization. They can be described by their quantum state $|VH\rangle - |HV\rangle$ as indicated in Figure 12.2. These photon pairs can be collected with a single-mode optical fiber and used for entanglement-based QKD experiments.

Another successful scheme for entangled-photon generation is the so-called polarization Sagnac interferometer [32]. This interferometric scheme uses a periodically poled (pp) type-I nonlinear crystal (e.g., ppKTP: periodically poled potassium titanyl phosphate) and collinear quasi-phase matching to generate photon pairs with the same polarization ($|HH\rangle$). The Sagnac loop is pumped in two directions and the $|HH\rangle$ -polarized photon pairs in one direction are rotated to $|VV\rangle$. Combining both arms then gives the entangled quantum state $|HH\rangle + |VV\rangle$. Because of the high nonlinearity of such periodically poled crystals, the Sagnac source provides photon pair rates approximately ten times higher than what is achievable with standard BBO-based sources. This is critically important when a QKD experiment involves very high channel losses.

The entanglement based source is also included in Figure 12.3. The relative alignment constraints of the UV laser, the nonlinear crystal and the single mode fibers are severe and must be kept at high mechanical stability.

12.4.2.3 Polarization issues in the free-space channel

Since in all of the QKD protocols described above the quantum state is encoded in the polarization of a photon, relative rotations of the polarization reference frame cause bit

errors in the communication protocol and therefore need to be corrected for [33–35]. For example, a misalignment of the two complementary bases settings (i.e., the $0^\circ/90^\circ$ basis and the $45^\circ/135^\circ$ basis) used in the analyzer modules of the communication partners of about 2° will result in a $QBER \approx 0.5\%$.

It turns out that atmospheric effects, like scattering, turbulence, and the Faraday effect due to the Earth's magnetic field, affect the polarization state of light only very slightly: for typical experimental parameters these effects give an overall rotation of the linear polarization plane less than 10^{-3} rad. Hence, the main polarization transformations arise from the optical elements within the transmitter and receiver, and the motion/rotation of the HAP or the satellite terminals relative to the ground station. Due to the extremely low power of the signal in the schemes presented in this text, polarization compensation schemes might be based on either a polarization reference beam (well isolated from the signal wavelength attached to some pointing assembly), or on a computational prediction of the rotation angle from accurate knowledge of the platform's trajectory and rotation.

The feedback frequency required to compensate for the turbulence is determined by the speed of the polarization drift and will probably not exceed the Hz range (except in the case of vibrational effects possibly present on a HAP/Satellite). Note that the compensation has to be performed simultaneously for both basis states used in the respective QKD scheme.

12.4.2.4 Receiver with polarization analysis setup and detector module

A typical single photon detector has no ability to distinguish the incident photons by polarization, but generates an output pulse with a certain probability (quantum efficiency) for every incident single photon. Therefore an external polarization analysis module is needed, where measurement of the polarization corresponds to which output port the photon comes out from. As sketched in Figure 12.4, the incoming single photon enters the module through the input window (e.g. from the telescope). Using a two channel analyzer, each measures the incoming photons randomly (e.g., using a 50/50 beam splitter or an active, but random switch) in either of the two complementary polarization bases (i.e., the $0^\circ/90^\circ$ basis and the $45^\circ/135^\circ$ basis). As the output port, four multimode fibers are used as an interface to the single photons detectors. Such a module can be designed very compactly.

Each of the four optical outputs of the polarization analysis modules are connected to single photon detectors. These avalanche photo diodes (APD) have to be cooled to an operating temperature of about -30°C and must be stabilized to $\pm 1^\circ\text{C}$.

Some electronics is used to passively quench the avalanche triggered by single photons (and the dark counts) as well as to clearly indicate the moment of arrival of the photons by electrical pulses. A convenient way to further process the arrival time is to record the times of detection events in time tags. The detector module can be housed in one box suitable to harsh environment as well.

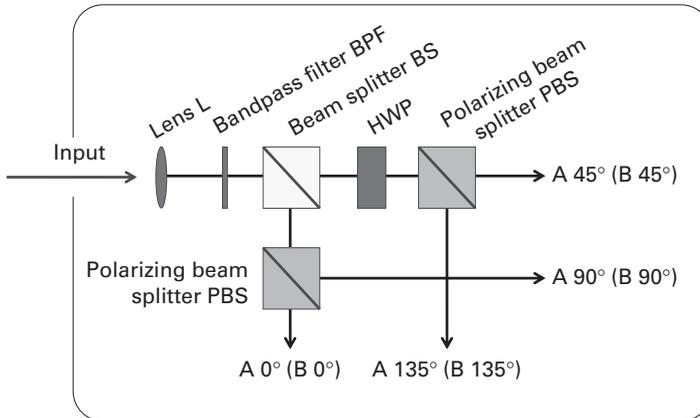


Figure 12.4 Polarization analyzer module. The lens (L) collimates the input beam which is then spectrally cleaned from the background by the narrowband filter BPF. For the random selection of the basis, each incident photon reaches the non-polarizing beam splitter (50:50 BS) and is subsequently steered to one of the two outputs corresponding the $0^\circ/90^\circ$ basis and the $45^\circ/135^\circ$ basis, where for the latter case a half-wave plate (HWP) rotates the measurement basis to set the different analysis orientation. Each analyzer is a polarizing beam splitter (PBS) and the output is coupled to multimode fibers.

12.4.2.5 Temporal synchronization

To perform quantum communication experiments it is necessary to establish time correlations either between two time-of-arrival events at different locations where the emission times are unknown (for entangled-photon based QKD), or between the emission time and detection time in order to attribute the measurement results to certain polarization basis preparations (for weak-pulse QKD, where the emission time is known). Entangled pairs are correlated to within their coherence time to some femtoseconds, the accuracy of synchronization is therefore limited by the timing jitter of the single-photon detectors (i.e. ≈ 1 ns for Si-APDs) and the synchronization of the two local time bases. In [6], the time-base synchronization was achieved using a combination of global positioning system (GPS) and software-driven time correlation, which enabled maintaining a “coincidence” window of < 0.8 ns over hours. The software was designed such that even drifts of the local timescales were compensated by maximizing the coincident events. This requires accurate knowledge of the distance of the respective terminals in order to be able to set the correct offset corresponding to the (in-flight varying) time-of-flight of the quantum signal when performed with moving terminals – this applies both for weak-pulse and entanglement-based schemes.

12.4.3 Recent results of free-space demonstrations

In order to show the feasibility of QKD using HAPs or satellites in a low earth orbit (LEO), several experiments have been carried out over a distance of 144 km between the two Canary Islands, La Palma and Tenerife. There, the atmospheric conditions, channel

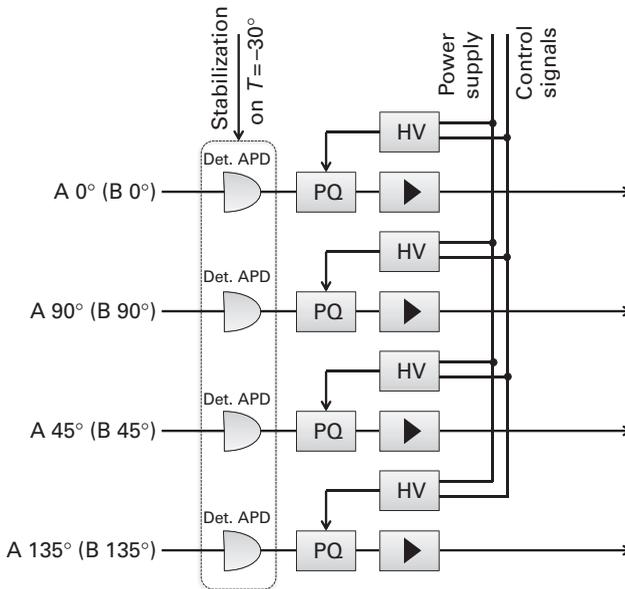


Figure 12.5 The detector module (Det.) consisting of four APDs (avalanche photon diodes), each with a passive quenching circuit (PQ), a DC high-voltage converter (HV) and an amplifier for signal configuration. To reduce dark counts (output signals generated without the presence of a photon), the APDs need to be cooled to temperatures typically $T \approx -30^\circ$.

attenuation and experimental setup contributed about 30 dB of loss, comparable to the loss of an optical link from HAPs or LEO satellites to ground-based receivers.

One experiment [6] demonstrated entanglement-based QKD. Entangled photon pairs were generated on the island of La Palma, one photon was measured locally whereas its partner was sent over a 144 km free-space link to Tenerife where it was collected by ESA's optical ground station (OGS). During a measurement run of 75 seconds, 789 coincidences could be detected between the two islands. With the obtained $QBER = 4.8\%$ this resulted in a secure key of 178 bits in total, corresponding to a secure key rate of 2 bit/s. With the sources available today, which are emitting almost an order of magnitude more photons per second, this would provide a key generation rate again almost an order of magnitude higher.

Another experiment [5] demonstrated decoy state QKD with weak coherent laser pulses on the very same optical link between La Palma and Tenerife. Due to stray light and the high optical attenuation through the 144 km free-space channel, secure key exchange solely based on the BB84 protocol would normally no longer be secure. By using the decoy-state protocol, however, the secrecy of the resulting quantum key could be ensured, despite the Poissonian photon statistics of the emitted pulses, and an averaged secure key rate of up to 42 bit/s was possible.

Finally, we report an experiment aimed at confirming the feasibility of establishing a link between an orbiting source of single photons and a ground telescope. In this

experiment, we simulated a quantum communication single-photon source on a satellite, and showed how the very dim signals could still be detected. Using a Japanese satellite (Ajisai) covered with retroreflectors at about 1600 km of slant distance, we emulated the space-based single-photon source by attenuating an outgoing pulsed laser such that only one single photon is reflected by the satellite back into the communication channel. This experiment showed that it was possible to distinguish the received single photons from the background (mainly starlight in the FOV of the receiver) with sufficient SNR to be used for quantum communication protocols.

12.5 QKD networks

In Sections 12.1–12.3, we discussed only single point-to-point QKD links, where Alice and Bob are connected by a quantum channel for single-photon signals and a classical channel for conventional communication: i.e., one particular Alice is always connected to the same Bob. This is the concept used for almost all current QKD systems, both commercial, fiber-based systems and the experiments discussed in Section 12.4.3. In the following discussion, the names “Alice” and “Bob” refer to sender/receiver pairs for each QKD link.

In the future, when one user wants to share information securely with a communication partner, they will establish a connection via an intermediate QKD network. Unconditional security over such a network will require special quantum repeater technology (currently under active development) [36], because standard optical amplification would destroy the fragile quantum correlations which provide both the communication mechanism and the security. Such quantum repeaters would enable a “truly quantum” network [37], allowing the distribution of genuine entanglement between far-separated communication partners, which they could then, for example, use to directly establish a secure communication key. Alternatively, even with existing technology, partners will be able to share secret information over a network of individual QKD links combined using “trusted” repeater nodes [38, 39], where, as part of the information sharing protocols (discussed in detail in Section 12.5.2 below), the secret information must be electronically stored and protected inside the node. These nodes must be trusted by all users wishing to utilize the network. In the following we will focus on this scenario of trusted-repeater networks.

The SECOQC QKD network has already demonstrated the possibility of combining independent point-to-point QKD links to form a trusted-repeater QKD network using fiber-based QKD links [4]. Each QKD link has its own “Alice” and “Bob” and these device pairs are never changed. The QKD devices at each node of the QKD network have been combined by node modules in that sense that the generated secure key is stored in key stores. These key stores, which are independent of the underlying realization of the QKD links, are filled by secret keys distributed over each QKD link and used by applications that request these keys.

In order to communicate beyond the limits of a local metropolitan network, however, users will need to establish a connection from the ground to a HAP or a satellite. Since

the partner will presumably not be on location there at the time, the flying object will serve as a trusted repeater. The underlying quantum architecture of HAPs and satellites is different to fiber-based QKD networks, because one Alice on the flying object needs to communicate sequentially with many Bobs. If we assume there is only one Alice on the flying object, she must select a specific Bob with high optical connectivity and high demand for a fresh key by steering the sender telescope to send her photons to the receiver. Alice and Bob behave like a single link, her photons will be collected and the quantum measurements form a key, which will be saved in the corresponding key store. If Alice selects another Bob later on, the same procedure will be repeated. At the end, the various key stores for the selected Bobs will be used as appropriate. Despite this 1: N architecture (one Alice to many Bobs), the principle at the layer of QKD links in a broader network (e.g., with the problems of synchronization to maintain a common time frame, etc.) is no different from the already well-developed fiber-based QKD network which has one central node and N independent links to the users.

In the following section, we consider the special case of entanglement-based QKD links for QKD networks. In Section 12.5.2, different “QKD transport” protocols will be described to form a QKD network and finally we will discuss application scenarios permitting global QKD link connections in Section 12.5.3.

12.5.1 Entanglement based QKD links in QKD networks

In the special case where an entangled source on the flying object sends the photons in each pair simultaneously to different ground locations, the two photons need to be measured on arrival at nearly the same time. Depending on the different distance between the source and the two receiver stations, the time of travel will vary marginally (e.g., a difference in path length of 30 km causes one photon to arrive earlier by 10 μ s). However, as discussed in Section 12.4.2.5, a common time frame with sub-nanosecond precision needs to be defined for Alice and Bob to match their clocks and to maintain correlations.

Although the source is high above Alice and Bob, such a system could be treated as a single QKD link, because the measured quantum correlations and further steps are equivalent to a QKD link based on, e.g., WCP operated over a direct quantum channel. The information exchanged during these subsequent classical protocol steps could be sent over the Internet. This architecture helps to reduce the communication load to the flying object. Moreover, in contrast with the other QKD network architectures discussed below, the flying sender station containing the source which generates the entangled photon pairs does not need to be trusted. The drawback is that the distance is limited to simultaneous line of sight.

For longer distances (e.g., different continents), the same source of entangled photon pairs can instead be operated in the single-link mode, where one photon is measured immediately at the flying object, which acts as Alice for the QKD link, and the second photon is sent to Bob as usual. In this case, in order to span longer distances, the system loses its previous advantage of cutting out the middle station. Ultimately, such entanglement-based QKD links operated along highly asymmetrical path lengths

are equivalent to a single QKD link based on, e.g., WCP. In the following parts of this section, we will consider only this latter case.

12.5.2 QKD transport protocol

Whenever two users want to share a key with each other without line-of-sight connection or a common fiber, they can achieve this by setting up two free-space quantum channels to a common object. This could be a high tower or HAP for metropolitan areas or a satellite for global quantum communication. In all cases the common object acts as node in a QKD network. The node establishes separate keys with each user, fully independently, because unlike an entanglement-based QKD system, simultaneous key generation is not necessary at all. The keys will be saved in key stores and kept secret on both sides electronically in the QKD hardware ready for later use.

The starting point of the QKD transportation protocol is as follows [4]: two users sharing their own secure keys, K_1 and K_2 , with the common node, want to use these keys for a security application. As soon as the users send a request for a secure application to the node, different methods can be used to establish a common bit sequence, or even direct communication over the node.

1. **Direct communication:** The first and most intuitive scenario is a direct communication over this node. This is possible in principle, when the first user encrypts his message M using his key K_1 with a bitwise XOR operation, $(M \oplus K_1)$, to an ITS-secure one-time-pad encrypted cyphertext, which is then sent to the node. Inside the node the received bit string, $(M \oplus K_1)$, is decrypted, $(M \oplus K_1) \oplus K_1 = M$, and the secret message is available in plain text. The message is then encrypted with the second key, $(M \oplus K_2)$, and sent to the second user. He is the only one with the necessary key, K_2 , to obtain again the original message, $(M \oplus K_2) \oplus K_2 = M$. This message was encrypted at all times during transmission, so provided the node is trusted, the message is secure. The first disadvantage of this scheme is that the node must be capable of all higher-layer applications possible with secure keys. It sounds easy for a simple one-time-pad encryption at low data rates, but it can be extremely difficult in the case of an AES-encrypted video conference with ongoing seed-key exchange, when the node must be able to continuously decrypt/encrypt this high amount of data, paired with the difficulty of a satellite simultaneously communicating data with several concurrent users. Furthermore, another broad disadvantage for future customers is that the node would have not only the key, but also the message fully decrypted. This may still be acceptable if secure messages are communicated between branches of one single company or institution, but would not be satisfactory if the service of “secure communication” were being offered by a communication provider, due to the high resulting risk of legal dispute. How could a serious communication provider prove that no secret information present at their nodes was leaked to an outside party either accidentally or deliberately?
2. **Secure bitstring:** Instead of sending the main message, it turns out to be advantageous to send only a secret key K_M . The first user would generate an additional secret

key only locally (e.g., using a quantum random number generator) and encrypt this bit string using the key shared with the node, $(K_M \oplus K_1)$, for transmission to the node. The node would again decrypt it, recovering a meaningless bit string, K_M , and encrypt this bit string with the second secure key, $(K_M \oplus K_2)$, for transmission to the second user. As a result, both users (and the node) share a random bit sequence K_M .

3. **XOR:** This even simpler QKD transportation protocol minimizes the ground-to-satellite communication. After the node receives a key-distribution request, it then encrypts the key from the first user, K_1 , with the one from the second user, K_2 , with a bitwise XOR operation, $(K_1 \oplus K_2)$, and sends it to the second user. Only the second user can extract the secret key of the first user by another bitwise XOR operation with the locally stored key, $(K_1 \oplus K_2) \oplus K_2 = K_1$. Again, both users (and the node) share the same key, which is the QKD key of the first user with the node, K_1 .

All three concepts have the common feature that the node must be trusted to be robust against electronic, physical, or software hacking to extract secret keys or secure information. For security reasons, the communication of all possible applications using the secret key should not be communicated over the same node, because the key from both users, K_1 and K_2 , as well as the additional key, K_M , was stored there and the secret information could be easily extracted with the knowledge of the ciphertext and the key. In the first protocol which uses direct communication, the QKD provider routinely possesses the secret message M , but for the other two alternatives (using a secure bitstring or XOR procedure), a corrupt QKD network provider would need to actively search for the encrypted ciphertext ($(M \oplus K_M)$ or $(M \oplus K_1)$) on the communication channels from both users to get access. Using two alternative QKD network providers with the delivered keys, K_{M1} and K_{M2} , and building up a combined key, $K_M = (K_{M1} \oplus K_{M2})$, to encrypt the secure data, $(M \oplus K_M)$, a single corrupt QKD provider could not reveal the secret message, M . Only two cooperating corrupt QKD network providers would make the message M insecure.

As a result of QKD transportation protocols 2 and 3, both users share a common key without having a private QKD link. The system of trusted repeaters could easily be extended to more nodes in between. It is straightforward to show that all scenarios can be scaled to a linear chain of more nodes in a larger QKD network.

For each secure bit exchanged between two users (as for K_M in the second QKD protocol) they need to invest one bit of information for each link K_1 and K_2 . This lost overhead the same bit length as the secret information is needed, because the additional information is fully in the hands of the adversary. This is particularly evident for the third method of QKD transport, because the result of the XOR combination of both secret keys is sent to one user and can therefore be considered to be public.

12.5.3 QKD networks via HAPs and satellites

Relay stations, similar to radio link systems at microwaves frequencies, clearly accessible at high altitudes would be needed to span longer distances between metropolitan areas by QKD links. Unfortunately, the atmosphere is very dense for optical wavelengths

due to pollution and scattering (this same effect causes a blue sky and beautiful sunsets). Nevertheless, it is possible to reach objects such as HAPs (Section 12.5.3.1) and satellites (Section 12.5.3.2) with losses acceptable for QKD. Clearly, it is advantageous to cross the disturbing layer in the vertical direction to reach a ground station from a flying node (or vice versa) with minimal losses. This could be achieved if each metropolitan area had its own HAP.

In the following we assume that QKD links could be technically realizable between different HAPs as well as between HAPs and satellites. This opens new interesting scenarios and applications for QKD networks with flying objects: the HAPs over metropolitan areas could act as nodes and be connected by QKD links over a network of HAPs (Section 12.5.3.3) for shorter distances, or again over satellites (Section 12.5.3.4) to achieve globally spanning QKD. To increase the secure key rate, more than one satellite could be used (Section 12.5.3.5). If intersatellite QKD connections are also feasible, a highly dynamic QKD network could perform superior interconnection over global distances. These scenarios will be outlined in detail below for the case of two users (for simplicity), although it would of course also be possible to connect far more users with the same infrastructure.

12.5.3.1 QKD links to a single HAP

In this first, most basic scenario one HAP establishes two QKD links with two users at line of sight. The HAP is stationary in the sense that it moves only slightly away from the expected position, but bad weather conditions would decrease the practical availability. The schematic is shown in Figure 12.6(a). For simultaneous link connections, the HAP would need two sending telescopes, but this is actually not necessarily required when using the concept of QKD networks described above. It is also possible for the HAP to have only one telescope pointing alternately to both users, with the minor consequence that the key stores would then be filled up sequentially.

To change the number of users at the ground, there is no need to change the hardware of a HAP simultaneously, but only provide a new target point for the sending telescope. For the other users, the potential time slots available to them during operation would be decreased, giving a correspondingly lower secure key rate. The number of offered time slots and single photons could again be increased by an additional sender telescope, expanding the already installed capacity linearly.

12.5.3.2 QKD links to a single satellite

This simple scenario is sketched in Figure 12.6(b). Whereas HAPs are stationary relative to the ground, satellites are moving rather fast. Thereby high-rate QKD links with short times to establish connections are needed, because the time a satellite could be used for key exchange is less than 10 minutes before the receiver telescope would then have to wait for the next opportunity to receive photons. Weather conditions and radiation from the Sun decreases the time of availability further.

In principle, the satellite serves as a node of a QKD network and stores the established key in the key buffer. For the QKD transport protocol outlined in Section 12.5.2 the first method of direct communication clearly would fail. Also the second, improved version

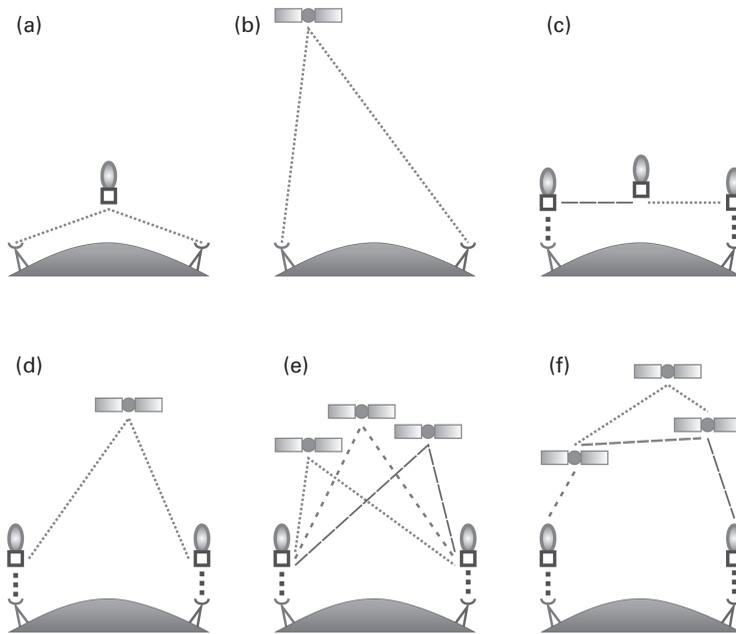


Figure 12.6 Six different use cases of quantum communication between users, HAPs and satellites are illustrated schematically. Two user stations (marked as two ground telescopes) want to establish free-space links over HAPs (indicated as balloons close to ground) and satellites (objects far above ground). Scenario (a), (b), and (d) correspond to QKD networks with one possible path. The path in scenario (e) is selected according to the best-suited satellite. The HAPs shown at scenario (c) indicate a stationary QKD network with different possible paths for non-neighbour communication. For scenario (f), the two HAPs could be connected by many flying satellites building a time-dependant mesh network as a QKD network.

would need additional communication to the satellite. In contrast to that, a flying node performing the third method of QKD transport protocol would only receive commands from the ground and would send the XOR combination of the secret key back. Interestingly, it would not only be the base station of the user that could receive this XOR combination of secrets. Instead, any arbitrary base station globally could receive the bit string and communicate it to the target user over a public network such as the Internet.

To summarize, this scheme has the disadvantage of intermittent short-term optical visibility, which can potentially be further complicated by bad weather conditions. The latter issue can be heavily relaxed in the next scenario for shorter distances.

12.5.3.3 QKD links to HAPs connected by a network of HAPs

The best optical path through the atmosphere is to transmit vertically, to try to travel through the disturbing layers of air as undistorted as possible. If a HAP could be placed directly above the ground station, a large volume of keys could be distributed vertically during good weather conditions and saved in key stores. At such high altitudes the quality of horizontal optical connections is far improved compared to the ground. Therefore, it would be possible to connect the HAPs directly or launch other HAPs working as

relay stations as shown in Figure 12.6(c). This solution would work for a linear chain of HAPs to span the required link distances and, moreover, it is also possible to arrange a 2-dimensional grid to cover the required area of a large country for long-distance quantum communication. Each HAP in the network would serve as a potential access node for users. The routing in a simple 2-dimensional cell as a basic building block was already demonstrated in the first fiber-based integrated QKD network experiment (SEQOQC). Thus, alternative paths can be used to avoid malfunctioning QKD links or damaged or problematic HAPs. Multiple paths can also be used to enhance the accessible secure key rate.

12.5.3.4 QKD links to HAPs connected by a single satellite

To span even longer distances, too many HAP relay stations would be needed, and such a scenario calls again for satellite-based key distribution. This scenario, sketched in Figure 12.6(d), is an extension of the one shown in Figure 12.6(b) using two HAPs as additional relay stations.

The additional effort required to establish QKD link connections to satellites over HAPs would be worthwhile in order to relax constraints due to bad weather conditions. The altitude of a HAP is well above problematic fog and most opaque disturbing layers of clouds and greatly increases the availability of a QKD link connection, so the unreliability of establishing a fresh key with the satellite is heavily reduced. Moreover, it would normally be easy to fill the key stores of the QKD link between the ground and the HAP. If the satellite delivers a new secret key to the HAP when there is no visibility to the ground due to bad weather, transport protocol 3 could be used to send the bit-sequence of the XOR combination of both secret keys to the ground station using radio frequency transmissions and it would still be possible to establish a key successfully. Even during a longer period of limited visibility, such as during a rainy season in equatorial regions, this method is still usable if enough keys had been distributed with the HAP in advance.

12.5.3.5 QKD links to HAPs connected by many single satellites

As outlined in Figure 12.6(e), it is possible to increase traffic over satellites by using more than one of them in parallel. At any given time, the optimal satellite can be chosen to establish a secret key between the users. The more satellites available for this scenario in different orbits, the more functionality of the connections that can be accessed.

12.5.3.6 QKD links to HAPs connected by a network of satellites

Again, as in the last scenario, each HAP is able to establish a key to different satellites, with the relaxed requirement that different HAPs do not need to choose a common satellite. Additional intersatellite QKD connections could then connect satellites either directly, if the distance is not too large, or via other satellites in different orbits. Consequently, the HAPs establish a key with an arbitrary satellite (whichever is the closest) and the network of satellites is responsible for then connecting the HAPs and its users.

From a network perspective, the backbone network in such a situation would not be highly constant and stable as in fiber-based networks, but would change constantly over time. As soon as two satellites would come close, they could establish new QKD links leading to a highly flexible access to the network architecture.

References

- [1] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden. Quantum cryptography. *Rev. Mod. Phys.*, 74:145–195, 2002.
- [2] M. Dušek, N. Lütkenhaus, and M. Hendrych. Chapter 5 quantum cryptography. volume 49 of *Progress in Optics*, pages 381–454. Elsevier, 2006.
- [3] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf *et al.* The security of practical quantum key distribution. *Rev. Mod. Phys.*, 81(3):1301–1350, Sep 2009.
- [4] M. Peev, C. Pacher, R. Alléaume *et al.* The secoqc quantum key distribution network in Vienna. *New J. Phys.*, 11(7):075001, 2009.
- [5] M. Sasaki, M. Fujiwara, H. Ishizuka *et al.* Field test of quantum key distribution in the Tokyo QKD Network. *Opt. Express*, 19:10387–10409, 2011.
- [6] R. J. Hughes, J. E. Nordholt, D. Derkacs, and G. Peterson. Practical free-space quantum key distribution over 10 km in daylight and at night. *New J. Phys.*, 4:43, 2002.
- [7] M. Aspelmeyer, H. R. Böhm, T. Gyatso *et al.* Long-distance free-space distribution of quantum entanglement. *Science*, 301(5633):621–623, 2003.
- [8] C-Z Peng, T Yang, X-H Bao, J Zhang *et al.* Experimental free-space distribution of entangled photon pairs over a noisy ground atmosphere of 13km. *Phys. Rev. Lett.*, 94:150501, 2005.
- [9] R. Ursin, F. Tiefenbacher, T. Schmitt-Manderbach *et al.* Entanglement-based quantum communication over 144 km. *Nature Phys.*, 3:481–486, 2007.
- [10] T. Schmitt-Manderbach, H. Weier, M. Fürst *et al.* Experimental demonstration of free-space decoy-state quantum key distribution over 144 km. *Phys. Rev. Lett.*, 98(1):010504, Jan 2007.
- [11] C. Erven, C. Couteau, R. Laflamme, and G. Weihs. Entangled quantum key distribution over two free-space optical links. *Opt. Express*, 16(21):16840–16853, Oct 2008.
- [12] T. Scheidl, R. Ursin, A. Fedrizzi *et al.* Feasibility of 300 km quantum key distribution with entangled states. *New J. Phys.*, 11(8):085002, 2009.
- [13] J. G. Rarity, P. R. Tapster, P. M. Gorman, and P. Knight. Ground to satellite secure key exchange using quantum cryptography. *New J. Phys.*, 4:82, 2002.
- [14] R. Kaltenbaek, M. Aspelmeyer, M. Pfennigbauer *et al.* Proof-of-concept experiments for quantum physics in space. In *Proceedings of SPIE*, 5161, pages 252–268, 2003.
- [15] M. Aspelmeyer, T. Jennewein, M. Pfennigbauer, W. R. Leeb, and A. Zeilinger. Long-distance quantum communication with entangled photons using satellites. *IEEE J. Sel. Top. Quantum Electronics*, 9(6):1541–1551, 2003.
- [16] M. Pfennigbauer, M. Aspelmeyer, W. Leeb *et al.* Satellite-based quantum communication terminal employing state-of-the-art technology. *J. Opt. Netw.*, 4:549–560, 2005.
- [17] P. Villoresi, T. Jennewein, F. Tamburini *et al.* Experimental verification of the feasibility of a quantum channel between space and earth. *New J. Phys.*, 10(3):033038, 2008.
- [18] R. Ursin, T. Jennewein, J. Kofler *et al.* Experiments with quantum entanglement in space. *59th International Astronautical Congress (IAC) Proceedings A2.1.3*, 2008.
- [19] M. Toyoshima, T. Takahashi, K. Suzuki *et al.* Results from phase-1, phase-2 and phase-3 kirari optical communication demonstration experiments with the nict optical ground station (koden). In *24th International Communications Satellite Systems Conference of AIAA, AIAA-2007-3228, Korea*, 2007.
- [20] J. Perdigues, B. Furch, C. de Matos *et al.* Quantum communication at ESA: Towards a space experiment on the ISS. In *Conference Proceedings IAC2007 Hyderabad, India*, 2007. Accepted for publication in *Acta Astronautica*.
- [21] N. Perlot, M. Knapek, D. Giggenbach *et al.* Results of the optical downlink experiment KIODO from OICETS satellite to optical ground station Oberpfaffenhofen (OGS-OP).

- In *Conference on Laser Communication and Propagation, Proceedings of SPIE*, 6457A, 2007.
- [22] R. Alléaume, J. Bouda, C. Branciard *et al.* Secoqc white paper on quantum key distribution and cryptography. arXiv:quant-ph/0701168v1, 2007. Quantum Physics (quant-ph); Cryptography and Security (cs.CR); Information Theory (cs.IT).
- [23] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. *Proceedings IEEE Int. Conf. on Computers, Systems and Signal Processing, Bangalore, India, IEEE, New York*, pages 175–179, 1984.
- [24] W.-Y. Hwang. Quantum key distribution with high loss: Toward global secure communication. *Phys. Rev. Lett.*, 91(5):057901, Aug 2003.
- [25] X.-B. Wang. Beating the photon-number-splitting attack in practical quantum cryptography. *Phys. Rev. Lett.*, 94(23):230503, Jun 2005.
- [26] G. Brassard, N. Lütkenhaus, T. Mor, and B. C. Sanders. Limitations on practical quantum cryptography. *Phys. Rev. Lett.*, 85(6):1330–1333, Aug 2000.
- [27] H.-K. Lo, X. Ma, and K. Chen. Decoy state quantum key distribution. *Phys. Rev. Lett.*, 94(23):230504, Jun 2005.
- [28] C.-Z. Peng, J. Zhang, D. Yang *et al.* Experimental long-distance decoy-state quantum key distribution based on polarization encoding. *Phys. Rev. Lett.*, 98(1):010505, Jan 2007.
- [29] C. H. Bennett, G. Brassard, and N. D. Mermin. Quantum cryptography without Bell's theorem. *Phys. Rev. Lett.*, 68(5):557–559, Feb 1992.
- [30] T. Jennewein, C. Simon, G. Weihs, H. Weinfurter, and A. Zeilinger. Quantum cryptography with entangled photons. *Phys. Rev. Lett.*, 84(20):4729–4732, May 2000.
- [31] A. Poppe, A. Fedrizzi, R. Ursin *et al.* Practical quantum key distribution with polarization entangled photons. *Opt. Express*, 12(16):3865–3871, Aug 2004.
- [32] A. Fedrizzi, T. Herbst, A. Poppe, T. Jennewein, and A. Zeilinger. A wavelength-tunable fiber-coupled source of narrowband entangled photons. *Opt. Express*, 15(23):15377–15386, Nov 2007.
- [33] C. Bonato, M. Aspelmeyer, T. Jennewein *et al.* Influence of satellite motion on polarization qubits in a space-earth quantum communication link. *Opt. Express*, 14(21):10050–10059, Oct 2006.
- [34] C. Bonato, C. Pernechele, and P. Villoresi. Influence of all-reflective optical systems in the transmission of polarization-encoded qubits. *J. Optics A: Pure and Applied Optics*, 9(10):899, 2007.
- [35] G. Anzolin, A. Gardelein, M. Jofre, G. Molina-Terriza, and M. W. Mitchell. Polarization change induced by a galvanometric optical scanner. *J. Opt. Soc. Am. A*, 27(9):1946–1952, Sep 2010.
- [36] N. Sangouard, C. Simon, H. de Riedmatten, and N. Gisin. Quantum repeaters based on atomic ensembles and linear optics. arXiv:0906.2699v2, 2007. 52 pages, 34 figures, to appear in RMP.
- [37] A. Acin, J.I. Cirac, and M. Lewenstein. Entanglement percolation in quantum networks. *Nature Phys.*, 3:256–259, 2007.
- [38] M. Dianati, R. Alléaume, and M. Gagnaire and X. Shen. Architecture and protocols of the future European quantum key distribution network. *Security Comm. Netw.*, 1:57–74, 2008.
- [39] R. Alléaume, F. Roueff, E. Diamanti, and N. Lütkenhaus. Topological optimization of quantum key distribution networks. *New J. Phys.*, 11(7):075002, 2009.

13 Optical modulating retro-reflectors

William Rabinovich

13.1 Introduction

Free-space optical communication (FSO) offers many potential advantages. These include: high bandwidth, security, non-interference, and a large amount of available spectrum. Many of these advantages are related to optics' short wavelength relative to radio-frequency (RF) or microwave systems. At optical wavelengths even a small transmit aperture can produce highly directional beams, and optical receivers can focus light down to very small spots. To make use of this high optical antenna gain very good pointing is required. Even a small telescope, with a 1 cm diameter, has a diffraction-limited beam divergence of less than 250 micro-radians in the near-infrared. Pointing with this accuracy is challenging, particularly if one or both ends of the link are in motion. Active FSO systems with high pointing accuracy are often large, heavy, and have high power consumption [1]. Such systems are also complex, which can lead to reliability issues.

Despite these challenges, direct FSO links – those with active terminals on both ends – have many good applications. There are, however, other applications in which the two ends of the link have different payload and power capabilities. Some examples include: unattended sensors, small unmanned aerial vehicles (UAVs), and small, tele-operated robots. For these applications a modulating retro-reflector (MRR) may be an appropriate solution.

Optical MRRs couple passive retro-reflectors with electro-optic modulators to allow long-range, free-space optical communication with a laser, and pointing/acquisition/tracking system, required on only one end of the link. As shown in Figure 13.1, a conventional free-space optical communications terminal, the interrogator, is used on one end of the link to illuminate the MRR on the other end of the link with a continuous-wave beam. The MRR imposes a modulation on the interrogating beam and passively retro-reflects it back to the interrogator. The passive retro-reflector will generally have a large field of view over which incident light will be reflected back to its source, thus eliminating, or greatly reducing, pointing requirements on this end of the link. Despite this, the retro-reflected beam divergence can be very small, preserving the desirable features of direct FSO such as security and non-interference.

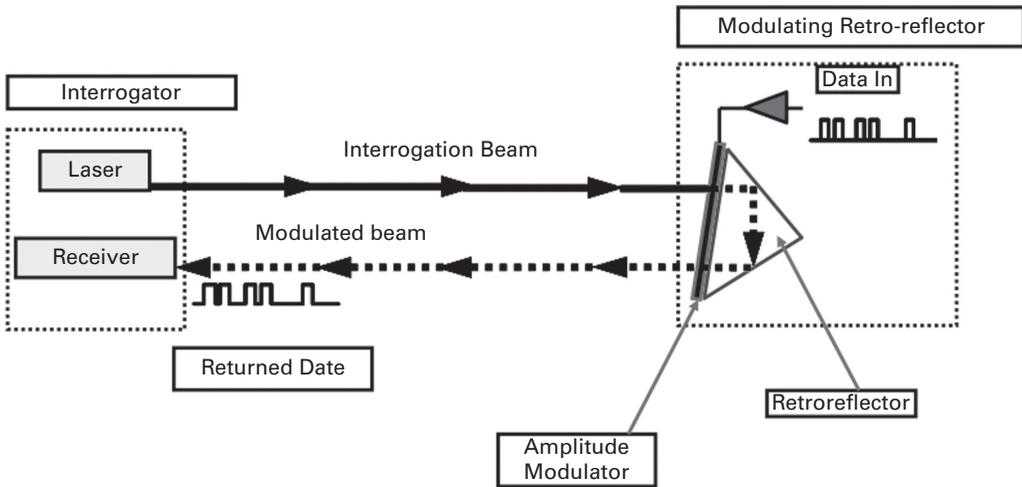


Figure 13.1 Schematic of a modulating retro-reflector communication link.

The concept of a modulating retro-reflector was first proposed in a 1904 patent by Sir Howard Grubb. He conceived of a communications link using a retro-reflector with a shutter, and a searchlight as the interrogating source [2]. In 1948, the idea was re-introduced by Stockman in a seminal paper that also envisioned RFID tags [3]. Optical MRRs can be made using a variety of passive retro-reflectors, including corner cube prisms and cat's eye lenses [4]. The electro-optic modulator may be placed as a window in front of the optic, or may be incorporated into the optic. A wide variety of electro-optic modulators can, and have, been used, ranging from vibrating membranes [5], liquid crystals [6–8], micro-electromechanical systems (MEMS) [9–13] and semiconductor multiple quantum wells (MQW) [14–17].

MRRs are possible for radio-frequency communication systems as well, but are especially attractive at optical frequencies. A reasonable figure of merit for a retro-reflector is the maximum number of distinct optical spots over its field of view. This is proportional to the square of the FOV divided by the beam divergence of the retro-reflected beam. For a diffraction-limited retro-reflector this is,

$$\left(\frac{\theta_{FOV} D_{retro}}{2.44\lambda} \right)^2 \quad (13.1)$$

where θ_{FOV} and D_{retro} are, respectively, the field of view and diameter of the retro-reflector, and λ is the operating wavelength.

For compact systems this figure of merit is much higher for optical retro-reflectors than radio-frequency. This is, once again, due to optics' short wavelengths. Most retro-reflectors have fields of view measured in tens of degrees, regardless of the region of the electromagnetic spectrum that they operate in. The difference lies in their beam divergence. The diffraction-limited beam divergence for a 1 cm retro-reflector in the near-infrared is about 250 micro-radians, yielding a figure of merit of about 4×10^6 .

However, even for high-frequency microwave link at 35 GHz, a retro-reflector diameter of about 80 meters would be needed to produce the same figure of merit.

The advantages of MRRs come with some costs. As will be described below, MRR links fall off as range to the fourth power, and any atmospheric losses are suffered in double-pass. This limits terrestrial MRR links to maximum ranges of tens of kilometers, and often less [18]. In addition the configuration of the modulator for an MRR both restricts the choice of mechanism, and often requires a large modulator area. The net result is that MRR links are generally limited in their bandwidth to less than 100 Mbps, and are often much slower. Despite these limitations there are many applications in which MRRs offer significant advantages.

13.2 Modulating retro-reflector link budgets

Modulating retro-reflector links have similarities and differences from direct free-space optical links. Often, an MRR link is made bidirectional by pairing a wide field of view photodetector with the MRR. In this case the down-link to the MRR can be described by the same equations as a direct FSO link. The return link from the MRR however takes a different form. As with direct links, MRR links depend on laser power, beam divergence, pointing accuracy, and receiver diameter, but for an MRR link these parameters are all determined by the interrogator. The MRR parameters that affect the link are the MRR's optical antenna gain, its modulation efficiency, and its modulation bandwidth.

There are several ways to represent FSO link budgets, but for MRR links it is clearest to adopt the formalism used for RF links [19]. We start with a link source and then express the rest of the link in terms of losses and gains ending with a signal strength that can be compared to detector sensitivity.

The downlink from the interrogator to a detector at the MRR follows the standard FSO link equation

$$P_{sig} = P_{las} G_{tx} L_{tx} L_R L_{atm} G_{rx-MRR} L_{rx-MRR} \quad (13.2)$$

where the terms follow standard definitions described in Table 13.1.

The link budget for the retro-reflector return link is more complex, since the light must propagate out, be reflected, and then propagate back. It also includes two new terms, G_{MRR} , the retro-reflector antenna gain and M , the modulation efficiency [20, 21]. M represents modulator-specific characteristics and will be defined in Section 13.4.

The MRR acts as a receiver, intercepting the light of the interrogator, and a transmitter, re-emitting the light as its retro-reflects it. Thus its optical antenna gain is the product of the classical formulas for receiver gain and transmitter gain. The retro-reflector antenna gain is,

$$G_{MRR} = \left[\frac{\pi D_{retro}}{\lambda} \right]^4 S \quad (13.3)$$

where D_{retro} is the optical aperture of the retro-reflector, λ is the wavelength of light and S is the Strehl ratio of the optic. As can be seen the gain has a very strong dependence on

Table 13.1

| Term | Symbol | Definition |
|--|----------------|--|
| Received power | P_{sig} | Power returned from link |
| Laser power | P_{las} | Output power of laser |
| Transmitter antenna gain | G_{tx} | $\left[\frac{32}{\theta_{div}}\right]^2$ |
| Transmitter loss | L_{tx} | Power loss through optics |
| Geometric propagation loss | L_R | $\left[\frac{\lambda}{4\pi R}\right]^2$ |
| Atmospheric loss | L_{atm} | Loss due to scattering |
| Receiver antenna gain (MRR end) | G_{rx-MRR} | $\left[\frac{\pi D_{rx-MRR}}{\lambda}\right]^2$ |
| Receiver loss (MRR end) | L_{rx-MRR} | Power loss through optics |
| Retro-reflector antenna gain | G_{MRR} | $\left[\frac{\pi D_{retro}}{\lambda}\right]^4 S$ |
| Modulation efficiency | M | Modulator parameter |
| Receiver antenna gain (interrogator end) | G_{rx-MRR} | $\left[\frac{\pi D_{rx}}{\lambda}\right]^2$ |
| Receiver loss (interrogator end) | L_{rx} | Power loss through optics |
| Transmitter divergence | θ_{div} | Full angle to $1/e^2$ |
| Link range | R | |
| Laser wavelength | λ | |
| Receiver diameter (retro end) | D_{rx-MRR} | |
| Retro-reflector diameter | D_{retro} | |
| Retro-reflector aberration | S | Strehl ratio |
| Receiver diameter (interrogator) | D_{rx} | |

retro-reflector aperture. In fact, since both the antenna gain and the range dependence scale as fourth powers, doubling the aperture of an MRR doubles its range. It is also important to maintain near diffraction-limited performance from the optic.

The return link equation for an MRR can be written as [20],

$$P_{sig} = P_{las} G_{tx} L_{tx} L_R L_{atm} G_{MRR} L_{MRR} M L_R L_{atm} G_{rx} L_{rx} \quad (13.4)$$

where terms are defined in Table 13.1.

When the interrogating laser is bistatic, additional losses can occur. This is because the retro-reflector returns light to the transmit aperture, not the receive aperture. In addition, for terrestrial links, atmospheric aberration may limit how small the divergence from a large retro-reflector may get. This can reduce the retro-reflector antenna gain for large apertures. In most situations the geometric range losses of a retro-reflector link scale as $1/R^4$, where R is the one-way range between the interrogator and retro-reflector. This range dependence arises because the retro-reflector captures a fraction of the transmitted beam and re-emits it with a divergence determined by the aperture of the retro-reflector (for good optical quality devices). Exceptions to the range to the fourth-power dependence can occur when either end is in the near-field of the other, or when either the retro-reflector or interrogator apertures are under-filled by the optical beam.

Most MRR links have used simple on-off keying. However some coherent MRR links have been demonstrated [5, 12, 22]. This can be especially attractive since the

laser source and receiver are collocated, providing a phase coherent reference. It is also possible to consider MRR links that use higher-order signalling to increase bandwidth [23]. The strong range dependence of MRR links means that when ranges are less than the maximum for the link, margin is often very high. Higher-order signalling is one way to take advantage of this.

13.3 The optical retro-reflector

There are two broad classes of passive optical retro-reflectors used in MRR links: corner cube prisms and cat's eye optics. Both can be made to exhibit diffraction-limited beam divergence, and both can operate over wide fields of view.

Corner cube retro-reflectors are the most common configuration. Corner cubes can be hollow or solid. For solid retro-reflectors, the field of view over which retro-reflection can occur is determined by the index of refraction of the prism. A solid glass corner cube has a full width half-maximum for retro-reflection of 26 degrees, whereas a silicon corner cube has a field of view of 60 degrees. Corner cube MRRs can be arranged as arrays of prisms canted in different directions so as to increase the field of view. Generally this is done using individual, high-quality, retro-reflectors mounted in the array. As an alternative, embossed arrays of plastic or metal corner cubes are available in large sheets, but have poor optical quality. These arrays are very thin, which may be advantageous in some applications.

Corner cube MRRs can have an optical modulator placed in front of the prism, as shown in Figure 13.1, or used as one or more of the faces of the prism. In general, it is easier to make high-quality MRRs using the modulator as a window, but for modulator technologies that are reflective, rather than transmissive, integration into a hollow corner cube can be effective.

Corner cube MRRs are generally rugged and inexpensive, making them very attractive for many applications. However, they require the modulator to be the same size as the optical aperture of the retro-reflector. This makes high-bandwidth, long-range, retro-reflector links hard to achieve. As either the range or bandwidth increases, more optical return is needed. Increasing the interrogator's laser power or receive aperture, or decreasing the interrogator's beam divergence, can increase the amount of light returned. However, eventually practical limits are reached. The options for increasing return from the retro-reflector end consist primarily of increasing the retro-reflector antenna gain. For a diffraction-limited retro-reflector, as shown in Equation (13.3), this can only be achieved by increasing the retro-reflector optical aperture or working at shorter wavelengths. Increasing the optical aperture can result in large increases in link margin, since the MRR's antenna gain scales as the fourth power of its diameter. However, for corner cube MRRs, larger optical apertures mean larger modulators. As will be discussed in Section 13.4, larger area modulators are often slower, and always consume more power.

The limitations of corner cube MRRs arise because the optical aperture and modulator size are coupled. Decoupling these parameters would enable higher performance devices. One idea that suggests itself is using a lens to increase the optical aperture. It

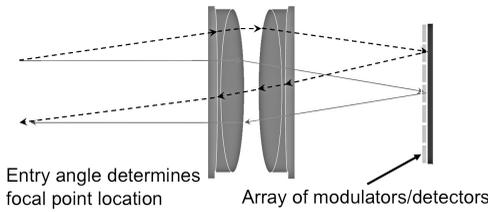


Figure 13.2 Schematic of a cat's eye modulating retro-reflector.

should then be possible to place the modulator in the focus of the lens and maintain a larger optical aperture and a small modulator aperture simultaneously. This can be achieved using a cat's eye MRR [4, 21, 24]. Figure 13.2 shows one form of a cats' eye. Light enters the optical aperture and is focused down onto a reflective modulator. The light is modulated and reflected back to the lens, which recollimates it.

The optic for a cat's eye MRRs must have several characteristics, three of which are:

1. It must preserve the retro-reflective properties of the system.
2. It should have as high an optical antenna gain as possible.
3. It needs a wide field of view to be of application interest.

There is no one form of cat's eye retro-reflector, but all contain some sort of focusing optics and a reflector. If the cat's eye MRR is to operate over a wide field of view, the optics f -number is important. A higher f -number optic implies a larger modulator in the focal plane. This is because the focal spot will move as the angle of incidence changes. The range of motion of the spot determines the modulator size,

$$D_{mod} = f\# D_{retro} \theta_{retro} \quad (13.5)$$

where D_{mod} is the modulator diameter, $f\#$ is the f -number of the cat's eye and θ_{retro} is the FOV that the cat's eye must work over.

Since we'd like to keep the modulator as small as possible this leads to a fourth desirable characteristic for a cat's eye optic:

4. It should have as low an f -number as possible.

Even a sophisticated diffraction-limited cat's eye optic will have an f -number of about 2. If the optic is to cover the same field of view as a corner cube (about 0.5 radians) then $D_{mod} = D_{retro}$, the same situation as with a corner cube. However, a cat's eye MRR can offer two advantages: If the required FOV is not large, a cat's eye MRR can have a small modulator, whereas the modulator size for a corner cube MRR is independent of the FOV. Second, while the focal spot does wander over a large area for a wide FOV, it only covers a small part of the focal plane at any one time. Thus if the angle of arrival can be determined, and if the modulator is divided into sub-pixels, only a small part of the modulator needs to be driven at any one time, greatly reducing the power draw.

Any cat's eye optic will involve some compromises of the four desirable characteristics listed above versus cost, size, complexity, and weight. In addition, cat's eye optics can be divided into classes in which the focal plane is flat or curved. The simplest kind of

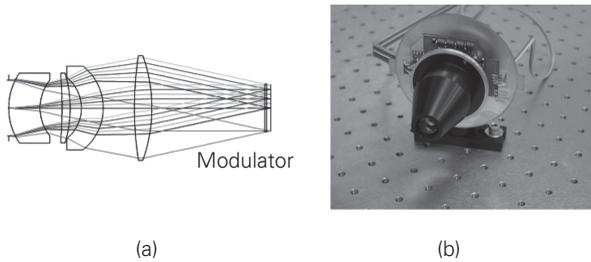


Figure 13.3 (a) Optical design of a cat's eye retro-reflector and (b) a fabricated cat's eye modulating retro-reflector.

curved focal plane cat's eye consists of a hemispherical lens coupled to a hemispherical reflector. These optics have a wide field of view of up to 180 degrees, but have very low Strehl ratios unless they are used at high f -number. The simplest kind of flat focal plane cat's eye consists of a telecentric lens coupled to a flat mirror in its focal plane. These optics exhibit less aberration than spherical cat's eyes but still have Strehl ratios of about 0.05 at $f/4$ and also have fields of view of 30 degrees or less [24].

For near diffraction-limited performance, it is necessary to use multi-element optics specifically designed as cat's eye MRRs. Most of these use a telecentric design as a starting point and then specialize for the aperture and desired field of view [4]. Figure 13.3(a) shows the optical design of diffraction-limited, flat focal plane, cat's eye optic with a 1.6 cm aperture and a 20-degree field of view [21]. Figure 13.3(b) shows a photograph of the cat's eye MRR, with a modulator array in the focal plane, based on this design. The modulator must have a diameter of 1.4 cm to cover the full 20 degree field of view. This is almost the same size as the optical aperture, so for this design to have an advantage, the modulator must be divided into sub-pixels and used with an angle of arrival sensor. However, this same optic could be used with a smaller modulator and a restricted field of view. Operated with a 5 degree field of view the modulator would only need to be 0.35 cm in diameter, but the optical aperture would remain at 1.4 cm. Compared to a corner cube MRR with the same size modulator, the cat's eye MRR offers 24 dB higher antenna gain.

Not all cat's eye MRRs require complex bulk objects. Very small devices have been made using high-index spheres half-silvered to make the back surface reflecting [12]. These types of MRRs will have low antenna gain, but can be very small and will have very large fields of view. Diffractive optics too can be used to make cat's eye retro-reflectors and may have significant weight advantages.

13.4 The optical modulator

In principle, the same kinds of optical modulators that are used for fiber optic communication systems can be used for modulating retro-reflectors. In practice, however, the requirements imposed by optical retro-reflectors restrict the geometry of the modulator.

As a result, optical modulators for MRRs generally have much lower bandwidth and/or extinction ratios than those for fiber optics. An optical modulator for an MRR must work for light entering a large optical aperture, and over a wide range of angles. It must also preserve the retro-reflecting characteristics of the optic. This eliminates single-mode waveguide modulators, such as Lithium Niobate, which are too small to cover a corner cube or the focal plane of a cat's eye MRR. Arrays of single-mode waveguides might work in a cat's eye but the high fill factor required would be hard to achieve.

Most of the MRR systems demonstrated to date use large-area surface-normal modulators with thin active regions. These devices often have lower extinction ratios than waveguide devices. This is less of a problem for most MRR links than it would be for fiber telecommunications links. The lower bandwidth, larger area, detectors used in free-space optical links generally operate far from the shot-noise limit. As a result, the noise in the optical receiver is generally dominated by that of the electronic pre-amplifier circuit. Unlike quantum limited systems, in which the noise level increases as the optical contrast ratio decreases, the noise level in this case is independent of the optical contrast ratio. The optical signal-to-noise ratio (OSNR) can then be defined as [20],

$$OSNR = \frac{P_{On} - P_{Off}}{P_{noise}} = P_{ret} \frac{e^{-\alpha_{On}} - e^{-\alpha_{Off}}}{P_{noise}} \quad (13.6)$$

where P_{On} is the optical power returned by the MRR when it is in its on-state, P_{Off} is the power returned in the off-state, P_{noise} is the noise equivalent power of the detector, P_{ret} is the optical power returned by the MRR excluding losses in the modulator, α_{On} is the double-pass absorption-length product of the modulator in its on-state, and α_{Off} in its off-state. From Equation (13.6) it can be seen that maximizing the OSNR depends on both the optical contrast ratio and the optical transmission of the MQW. This can be seen more clearly by defining a figure of merit for the MQW, its modulation efficiency,

$$M = e^{-\alpha_{On}} - e^{-\alpha_{Off}} = e^{-\alpha_{Off}} \cdot [C_{MQW} - 1] \quad (13.7)$$

where M is the modulation efficiency and C_{MQW} is the optical contrast ratio of the modulator. The OSNR of an MRR link is then simply $P_{Ret}M$.

While MRRs have been made with vibrating membranes [5] and gas-based Stark modulators [25], the most common types of MRR modulators have been: liquid crystal devices used in amplitude modulation mode, micro-electromechanical modulators used either as phase modulators or amplitude modulators, and multiple quantum well electro-absorption modulators.

All these systems are essentially capacitive in nature. While an MRR does not need to consume power for transmission, it does consume power to change its modulation state. This capacitive power draw scales as

$$P = CV^2f \quad (13.8)$$

where P is the power, C is the device capacitance, V is the drive voltage, and f is the modulation frequency. Thus, across all these families of modulators, it is desirable to have as low a capacitance and drive voltage as possible.

Liquid crystal modulators have been used very successfully for long-range MRR links [6]. It is relatively easy to make large-area, high optical quality, liquid crystal films. Devices with apertures of 2.5 cm have been demonstrated. This allows for corner cube retro-reflectors with large antenna gain. For shorter-range links, liquid crystal on silicon modulators with 1 mm diameters have been used to close a 30 meter link at 100 bps [7]. The principal drawback of liquid crystal based MRRs is their limited bandwidth. The extinction ratio of these devices declines rapidly as data rates exceed about 1 KHz. Liquid crystals also have limited temperature ranges of operation.

MEMS modulators have been used in a variety of ways with MRRs. While the details differ, there have been four distinct approaches using MEMS. One, or more, of the surfaces of a corner cube can be spoiled as a reflector using MEMS actuation either to form a diffraction grating or distort the surface so that it is no longer flat [11, 13]. It is also possible to use MEMS to tilt one of the surfaces of a corner cube so that it is no longer a retro-reflector [9]. A MEMS-based Fabry–Perot cavity can be used as a reflective modulator [10] built into a corner cube, or used in the focal plane of a cat’s eye MRR. MEMS Fabry–Perot cavities can also be used transmissively in front of a corner cube retro-reflector. MEMS devices can also be used as phase modulators for coherent MRR links. These devices have been demonstrated by moving a small cat’s eye retro-reflector to change the phase of the reflected light [12].

MEMS approaches offer several advantages. The technology is most often based on silicon fabrication, and thus has the possibility of being inexpensive in large numbers. MEMS MRRs can generally work over a large temperature range. MEMS MRRs also can work over a very wide wavelength range, potentially from the visible to the infrared, depending on the choice of materials. MEMS approaches using Fabry–Perot cavities do have a wavelength dependence within one free spectral range, but this does not limit their overall spectral range. MEMS devices are also much faster than liquid crystal modulators. Maximum bandwidths depend on implementation details, but most devices operate well up to around 100 KHz. Beyond that the extinction ratio often drops and ringing on the edges of the rise and fall pulse can become a problem. Still, with careful design, bandwidths up to 1 MHz are possible. Figure 13.4 shows a picture of a MEMS-based MRR developed by Boston University [13]. The device uses a MEMS actuated grating as one or more of the faces of the corner cube. When the MEMS structure is activated light diffracts from the face and is not retro-reflected. This MEMS MRR has a diameter of about 1 cm and a bandwidth greater than 100 KHz.

MEMS approaches have some limitations as well. It is unlikely that practical MEMS MRRs with bandwidths beyond about 1 MHz will be possible. Some MEMS designs have limited field of view. Also MEMS modulators must generally be pixellated to offer high bandwidth and large area. This can lead to diffractive losses and limited extinction ratios.

The third major modulator type used with MRRs are surface normal multiple quantum well electroabsorption modulators. The majority of MRR links demonstrated in the field have been with this type of device [18, 20, 21, 26]. MQW modulators are semiconductor devices consisting of multiple layers of alternating semiconductor materials embedded in a PN junction. Quantum confinement, due to the varying band-gaps of the layers,



Figure 13.4 A micro-electro-mechanical corner cube modulating retro-reflector.

alters the absorption spectrum of the well material. The excitonic absorption feature at the band-edge is enhanced in strength, and can be changed by the application of a field. By applying a reverse bias across the device its transmission for wavelengths near the band-edge can be changed. Figure 13.5 shows the band-edge absorption versus wavelength of an InGaAs/InAlAs coupled quantum well modulator for 0 and 6 V bias [27]. The regions in which these two curves show significant difference determine the operating wavelengths of the modulator. A variety of semiconductor material systems can be used to choose the wavelength of operation, although most MQW modulators have been designed for the near-infrared. Devices matched to 850 nm laser diodes [15], 1000 nm Neodymium and Ytterbium lasers [14], and 1550 nm telecom lasers [20] have been demonstrated. Device capacitance is typically about 5 nF/cm^2 . As a result devices of about 1 cm diameter consume about 100 mW/MHz of power. MQW modulators used in cat's eye MRRs may be pixellated and thus use less power at a given frequency.

MQW modulators can be used in either transmissive or reflective mode. Most often they have been used transmissively with corner cube MRRs. In this case, a large-area MQW modulator is placed in front of a corner cube retro-reflector. Cat's eye MRRs can use MQW modulators in reflective mode when the cat's eye optic has a flat focal plane. The MQW wafer can have metal deposited on its back surface to act as a reflector, or it is also possible to grow a Bragg mirror into the MQW structure itself. For cat's eye MRRs with curved focal planes there are two ways to use MQW modulators. A transmissive modulator can be placed between the last lens element and a curved reflector [21] or a conformal MQW can be deposited on a curved reflector [28].

The principal advantage of MQW modulators is their bandwidth. Unlike MEMS or liquid crystals the intrinsic switching speed of electroabsorption is in the tens of

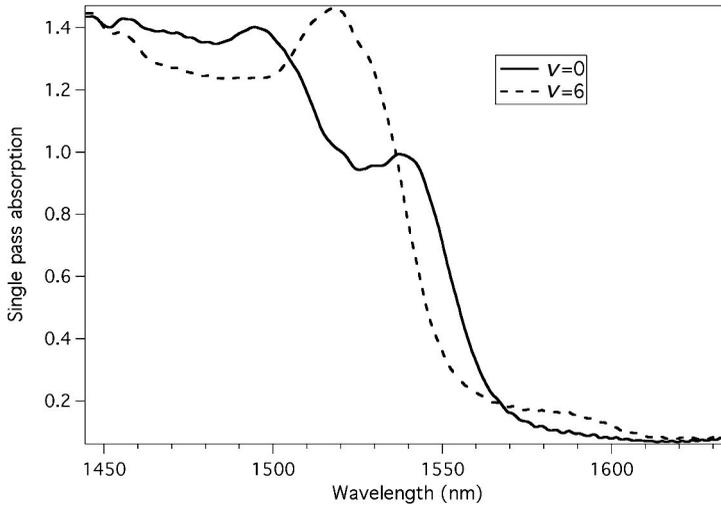


Figure 13.5 The absorption versus wavelength of a InGaAs/InAlAs multiple quantum well at two different applied voltages.

gigahertz. The practical switching rate of MQW modulators for MRRs is determined by their RC time. Because MQW modulators for MRRs are large, their capacitance is high which limits their bandwidth. Two strategies can be employed to get around RC time limits. First, the sheet resistance of the MQW electrode can be reduced using electrode design. Because the mobility of the p-type contact layers is much lower than the n-type layers they dominate the resistivity. One approach to overcome this problem is to use a metal grid on the top p-contact to distribute the electric field [29]. A grid structure is shown in Figure 13.6. Effective resistances below 10 ohms are possible with a grid. Used with a single element device this approach yields about a 10 MHz bandwidth with devices that are about 1 cm in diameter.

The other route to reducing the RC time is to lower the device capacitance. In most cases the only way to drop the capacitance is to reduce the area of the modulator that is being driven. This can be done by dividing the MQW into sub-pixels. In principle, one can use this approach to drive a corner cube MQW MRR to high bandwidths by modulating each sub-pixel simultaneously with its own driver. While this does allow higher bandwidth, power consumption, proportional to the total capacitance, still goes up.

A better approach is to use a cat's eye MRR. By combining the cat's eye MRR with an angle of arrival sensor, only one or two sub-pixels need to be driven at any one time, greatly reducing power consumption. The angle of arrival sensor can be a separate device, but it is also possible to use the MQW itself as the photosensor. In this case, a challenge is separating the small DC photocurrent signal from the much larger modulation signal running through the device. Depending on the choice of the modulation waveform it is possible to use electrical filtering to achieve this [30]. Figure 13.7 shows a pixellated MQW focal plane used in a cat's eye MRR. The 8 mm square device is separated into 8 rectangular sub-pixels each with sufficient bandwidth to support a 45 Mbps signal [21].

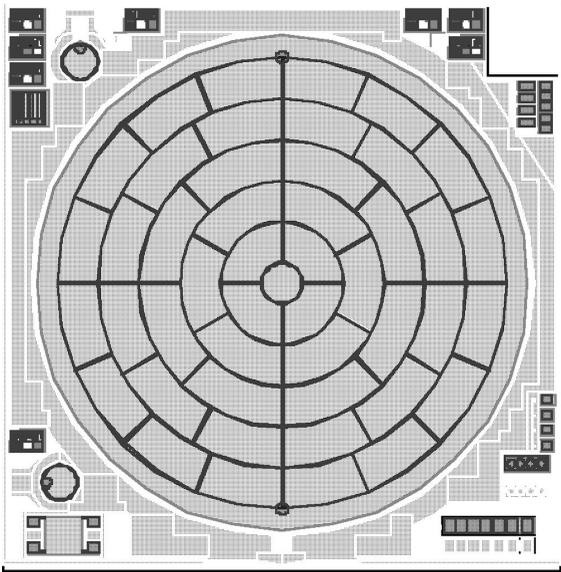


Figure 13.6 A grid electrode pattern for a multiple quantum well modulator.

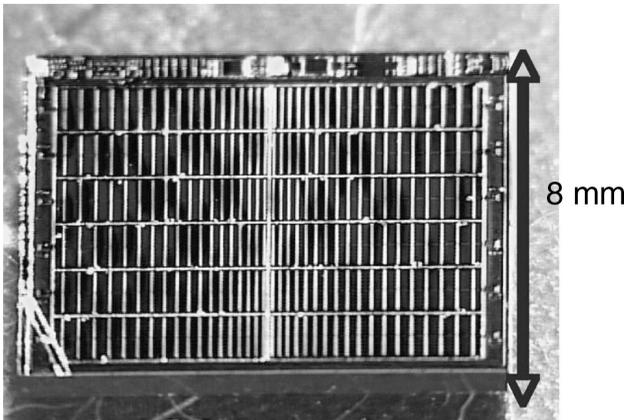


Figure 13.7 A pixellated multiple quantum well modulator designed for a cat's modulating retro-reflector.

The principal drawbacks of MQW modulators are their extinction ratio, device yield, and the temperature dependence of the band of operating wavelengths. Because the interaction length of a surface normal MQW modulator is on the order of one micron, it is difficult to achieve high extinction. Typical extinction ratios are about 2:1. However, as discussed above, limited extinction imposes a smaller power penalty on lower bandwidth optical links. Limited extinction also poses no problems for the highly scintillated waveforms typically seen for links whose range exceeds about 100 meters. While signal fades are much larger than 3 dB, they occur on a much slower time scale than a bit, and so do not distort the waveform.

Higher extinction can be achieved in at least two ways. First, it is possible to grow a thicker active region. This will yield higher extinction, but at the cost of lower throughput due to absorption in the MQW. For most MQW structures the modulation efficiency flattens out beyond an active region thickness of about 1.5 microns. Another way to achieve higher extinction is through an asymmetric Fabry–Perot structure. This type of MQW modulator utilizes a low finesse cavity. The length of the modulator is designed so that front and back reflections are exactly out of phase. When the absorption of the modulator is tuned to balance front and back reflections, complete extinction occurs. Asymmetric Fabry–Perots can produce high extinction with high throughput, but are difficult to use with MRRs. This is because the phase cancellation condition is highly dependent on angle of incidence. Thus for a corner cube MRR this type of modulator would have a limited field of view. A cat’s eye MRR would be restricted to high f -number optics, the opposite of what is normally desired. One way to avoid these problems is to combine MEMS and MQW modulators [31, 32]. If the cavity is tunable via MEMS it becomes possible to use an asymmetric Fabry–Perot with corner cube MRRs. This type of modulator has been demonstrated, but is complex to fabricate. However, for interrogators that allow shot-noise limited detectors these kinds of MRRs may offer significant advantages.

Device yield for large-area MQW modulators can be a problem. Typical fiber telecommunication modulators have areas of 100 square microns or less. An MQW modulator for an MRR has an area that is 10 000 times larger. This can lead to lower yield. Under high reverse bias, defects in the MQW structure can start to conduct, limiting the voltage that can be applied. The larger the device, the more likely this is to happen. Aside from the obvious growth optimizations to produce higher-quality material, there are two approaches to mitigating this problem. First, it is possible to sub-divide a large-area device into isolated pixels. Each pixel can then be characterized. Pixels with high reverse bias current are not connected to the circuit [15, 33]. This approach can be labor-intensive. An alternative is to use more sophisticated quantum well structures that require lower applied field. At lower bias, even material with many defects may operate without problems. One approach that has been demonstrated is to use coupled quantum wells [27]. These structures, which use double wells coupled by a thin barrier, can drop the required voltage by a factor of three or more down to about 3–4 V/ μm . In addition to increasing device yield this drops power consumption, which scales as V^2 , by a factor of ten.

The temperature dependence of MQW modulators arises from the shift of semiconductor band-gaps with temperature. For example, InGaAs, lattice-matched to InP, shifts by 0.7 nm/ $^{\circ}\text{C}$. The operating wavelength of the MQW modulator shifts at the same rate. For most terrestrial temperatures the change is a simple linear shift of the center operating wavelength, with little change in modulation efficiency. The typical full width half-maximum for an InGaAs MQW modulation efficiency is about 15 nm. So temperature shifts of $\pm 10^{\circ}\text{C}$, will have a modest effect on the link. For temperature variations larger than this there are several possible solutions. It is possible to temperature control the MQW with a heater or thermo-electric cooler. The modulator’s small thermal mass makes this relatively simple. This does, however, increase power

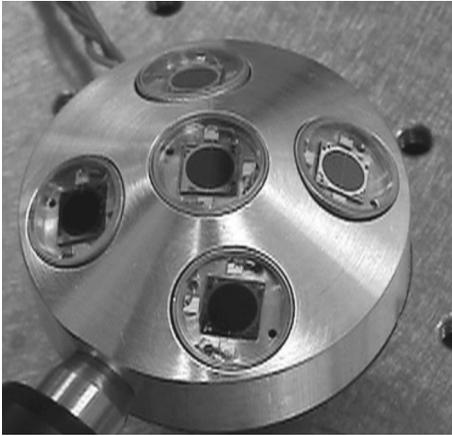


Figure 13.8 A five-element array of corner cube multiple quantum well modulating retro-reflectors.

consumption and complexity on the retro-reflector end of the link. Another option is to tune the interrogating laser to match the center of the MQW's operating wavelength. In the 1550 nm telecommunication band, lasers with more than 30 nm of tuning bandwidth are available yielding an operating temperature range of greater than 50 °C. This is a fairly robust solution that has been used in several MRR field tests [26, 34]. Its drawback is that it increases complexity at the interrogator. A third approach for temperature compensation is to apply a large DC bias field to an MQW to tune the wavelength region over which electroabsorption takes place [35]. An in-situ temperature sensor is used at the retro-reflector to adjust this bias based on ambient temperature. This can yield large temperature ranges of operation. The down-side of this approach is the high field required, often near break-down for the device. This reduces yield for large-area devices, and has an undetermined effect on the lifetime of the modulator.

MQW MRRs have been demonstrated in both corner cube and cat's eye configurations in a variety of sizes and configurations. Figure 13.8 shows an array of five 6.3 mm diameter MQW MRRs using glass corner cubes developed at the US Naval Research Laboratory (NRL) [20]. The array is about three inches in diameter. The devices are canted at different angles so that the entire array covers a $60^\circ \times 60^\circ$ field of view. The MRRs in the array had a bandwidth of about 10 MHz.

13.5 Modulating retro-reflector applications and field demonstrations

Applications for modulating retro-reflector links tend to fall into niches. Their small size, weight and power, as well as loose pointing requirements, make them attractive for small platforms and unattended sensors. However, their limited range and bandwidth rule them out of many of the applications generally considered for free-space optical systems. Most often MRRs compete for applications with RF systems, such as those based on the various IEEE 802.11 standards. MRRs offer more security than RF links, and can

also offer higher bandwidth at lower power at the sensor. Increasingly an advantage of MRRs is their immunity to the problems of spectral congestion that have come to plague sensor systems that use RF communication. As more and more high-bandwidth sensors and platforms are used nearby each other, the available RF bandwidth has become a scarce commodity. Mutual interference is rarely a problem for retro-reflector systems because of the high directionality of the beams, and the large spectrum availability in the infrared.

MRR links face problems similar to direct FSO links, but with some important differences. Atmospheric effects due to scattering from rain, fog, and other obscurants, and optical scintillation due to turbulence, affect both MRR and direct FSO links. In fact, MRR links are more affected by the atmosphere because the link is double pass. In practice, however, this is offset by the fact that most MRR links are shorter range than direct FSO links. Thus the overall system challenge posed by loss of margin due to weather is similar. Link drop-outs due to scintillation may be more severe for MRR links, however.

MRR links also face challenges of pointing and tracking, primarily on the interrogator side of the link. MRR interrogators often use narrow beams to make up for the propagation loss in the link. However, it is easier to acquire a retro-reflector than another FSO system. For acquisition in a direct FSO link, two narrow field of view systems must be pointed at each other. This creates a search problem that scales as the square of the number of addressable points in the field of view of each system. A retro-reflector however is always “pointed” at its interrogator as long as it falls within the very wide retro-reflector field of view. The search problem then scales linearly as the number of points in the interrogators field of view. MRR links do face increased challenge in acquisition and tracking because they work at lower data rates. The typical signal power in an MRR link is one to two orders of magnitude lower than for a high-bandwidth direct FSO link. This means the acquisition sensor has much less power to work with.

FSO interrogators can be either monostatic or bistatic. The transmitter and receiver optical axes are collinear in a monostatic interrogator, sharing an aperture. In a bistatic interrogator, the transmitter and receiver apertures are side-by-side. This design choice has unique ramifications for an MRR system with regard to transmit/receive isolation and minimum range.

All FSO systems must deal with the problem of transmit/receive isolation. In a typical direct FSO link, the ratio of transmit to receive beam power is 30–40 dB. In an MRR link it is often 70–80 dB. Direct FSO links can use different transmit and receive wavelengths, allowing very good isolation. An MRR link does not have this option, as the same laser is used for both transmit and receive. Isolation is most often achieved in MRR links using bistatic systems. This can be effective at long range, but has difficulties at short range. A retro-reflector returns power exactly back to where it originated. In a bistatic system, the power returns to the transmit aperture. If the range is long enough that the retro-reflected beam diameter covers both apertures, there is no problem, but when close most of the power may miss the receiver. This imposes a minimum range of operation for bistatic systems.

Monostatic transceivers pose their own set of challenges. Monostatic systems can be made with beam splitters, using polarization to isolate transmit and receive, or with fiber circulators. These solutions have been demonstrated with some degree of success. A simpler solution is to introduce the transmit beam into the optical path using an occluding mirror contained within a larger receive aperture, or to pass the interrogating beam through a hole in a mirror in the receive path. However, these solutions limit the size of the transmit beam, and hence if the system is to stay eye-safe out of the aperture, the output power must be limited.

While many MRRs have been tested only in the laboratory, there have been some field tests of complete MRR communication links, as well as some theoretical analyses [36]. Surface-to-surface links have been demonstrated with several different types of MRRs. In 1995, Honeywell demonstrated a coherent MRR link at 10.6 microns using a deformable membrane MRR [5]. The link was at a low rate, a few kilohertz, and used to transmit voice data from rooftop to rooftop at ranges up to 24 km.

In 2005, the US Naval Research Laboratory demonstrated a multiple quantum well MRR link from the shore to a boat on the Chesapeake Bay using the MRR array shown in Figure 13.8 [20]. These devices were 6.3 mm diameter InGaAs/InAlAs MQWs designed to work in the telecom band at 1550 nm. Data rates up to 5 Mbps were demonstrated out to a range of 2 km using an interrogator with a four-inch receive aperture and a 1.5 Watt laser. This interrogator had no tracking and was manually aimed at the boat. Scintillation indices for these links varied from about 0.3 to 0.6. These levels are lower than what one would find for a link near the ground because optical turbulence is generally lower over water. A higher data rate version of this link was demonstrated in 2007. Cat's eye MRRs, with a pixellated MQW focal plane were used at a range of 7 km to close a 45 Mbps MRR link from shore to ship [21].

In 2008 a ship-to-ship MRR link was demonstrated by NRL as part of the US Navy's Trident Warrior exercise [26]. A five-element corner cube MQW MRR array similar to Figure 13.8, but this time using 1 cm diameter devices, was combined with a 2 Mbps Ethernet modem to demonstrate video and file transfer at ranges up to 4.5 km between two US Navy vessels underway in the Pacific Ocean. The interrogator, used for this link could auto-track on the retro-reflected return using a quad cell sensor that could drive both internal fast steering mirrors and an external gimbal [37].

Also in 2008, a joint experiment of the University of Maryland's Laboratory for Physical Sciences and NRL, demonstrated an MQW MRR link from one side of the Chesapeake Bay to the other, a distance of 16 km [18]. This link used up to three 1.6 cm diameter cat's eye MRRs and sent data at 2 Mbps using an Ethernet protocol. Both files and video were transmitted. Though this data was taken at the same place as the 2005 Chesapeake Bay experiment, the much longer range resulted in scintillation indices that varied between 0.2 and 2. This experiment also showed that scintillation could be mitigated by aperture averaging. By comparing the return of one, two, and three MRRs in an array, it was shown that the scintillation index was inversely proportional to the number of retro-reflectors. The use of multiple MRRs for aperture averaging must, however, take into account coherent interference effects [38].

Links from the shore to a buoy carrying an array of liquid crystal MRRs were demonstrated by the Swedish Defense Research Agency in 2005 [8]. The MRRs were arranged around the buoy to give 360 degree coverage. Ranges up to 400 meters and data rates of 3 kbps were shown.

While most MRR field demonstrations have been conducted to investigate link properties, there have been some demonstrations of full systems and applications. In 2008, NRL demonstrated an application of MRRs to the problem of communicating with small robots used for explosive ordinance demolition [34]. These robots are teleoperated and require a bidirectional 1.5 Mbps Ethernet link to send video from the robot to the operator and commands from the operator to the robot. Normally this Ethernet link uses RF. However, in military applications of this technology, there is often a great deal of RF interference, which interrupts the link. An array of multiple quantum well MRRs and photodetectors was used to allow a bidirectional 2 Mbps Ethernet link with full azimuthal coverage. The array, mounted on the robot, is shown in Figure 13.9. The MRR link transparently replaced the RF Ethernet link. The robot was controlled using the MRR link out to the maximum line-of-sight range of about 1 km.

A limitation of this link, compared to an RF link, is that it is line of sight. One approach to mitigating this problem is to combine an optical link with an RF link. A pod could be mounted on the robot that contains both an MRR array and an RF transceiver. If both links use Ethernet they can be hooked together. Then a long-range FSO link can run from an interrogator at the base station to the MRRs on the pod. When it is necessary to go non line of sight, the pod can be dismounted from the robot. The link would then run optical from the base station to the dismounted pod, and RF for the shorter, non line-of-sight, link from the pod to the robot.

Not all MRR links have been surface-to-surface. Modulating retro-reflector links to airborne platforms with limited payload capacity are also an attractive application of the technology. In addition, airborne links generally have lower scintillation than links that take place entirely near the surface. These links have some additional challenges as well.

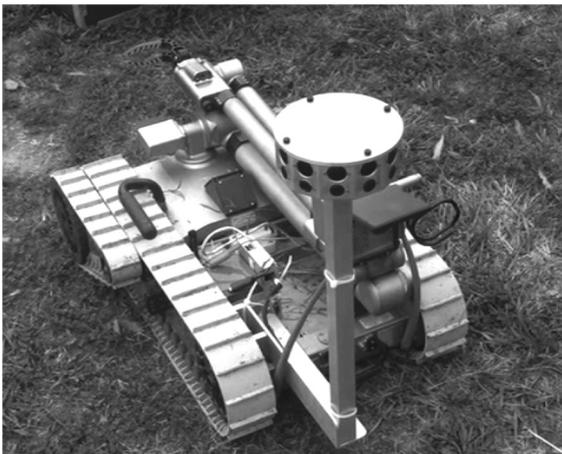


Figure 13.9 A small robot carrying an array of modulating retro-reflectors.

The acquisition and tracking problem for the interrogator is often more difficult than for surface-to-surface links because the air platform moves more rapidly. Likewise, the pointing problem for the retro-reflector is more complicated. An airborne link is more likely to require 360° coverage, at least in azimuth, because of the high maneuverability of the platform. Despite these challenges, a few air-to-ground MRR links have been demonstrated.

In an experiment conducted by Utah State University and the Air Force Research Laboratory, an array of nine 2.5 cm diameter liquid crystal MRRs was mounted on a balloon [6]. The MRRs were canted so that the array as a whole had a 90° full angle of acceptance. A 1.5 meter telescope coupled to an 810 nm, 5 Watt, diode laser source was used as the ground-based interrogator. Data links out to greater than 20 km range were demonstrated at data rates from about 1 kbps to 20 kbps.

In 2000, NRL demonstrated a multiple quantum well modulating retro-reflector link to a small model helicopter [39]. The helicopter carried a InGaAs/AlGaAs MQW MRR array. A 100 mW 980 nm laser diode coupled to a simple image-tracking system and a 2-inch receiver was used as the interrogator. Data at rates up to 910 kbps and live video was transmitted over a very short range, 35 meter, link.

In 2009, NRL demonstrated a much longer range link to an unmanned aerial vehicle (UAV) [40, 41]. In this case a different approach was adopted for covering a wide field of view from the aircraft. A single, 1 cm diameter, InGaAs/InAlAs multiple quantum well MRR and a photoreceiver were each installed in very small and lightweight camera gimbals to allow hemispherical coverage. Since the MRR only needed to be pointed to within its $\pm 15^\circ$ field of view, open loop control of the gimbal could be used.

The MRR and receiver were mounted on a small UAV, which also carried a gimballed camera that served as the data source for the MRR. The MRR transceiver gimbals, stabilized camera, and electronics were installed in the UAV wingpod as shown in Figure 13.10. Hardware providing GPS position, inertial sensing, and heading were included. An on-board processor, using the information from these sensors, and the GPS location of the ground station, maintained the pointing of the micro-gimbals holding the MRR and photoreceiver to the laser ground station.

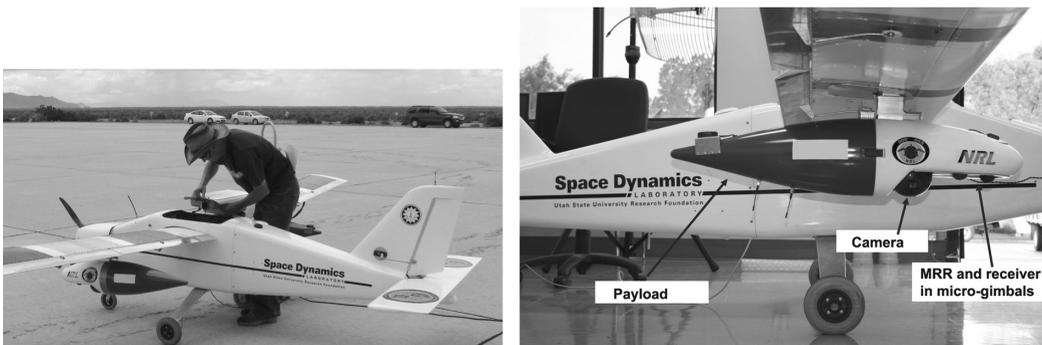


Figure 13.10 A small unmanned aerial vehicle with a wing pod carrying a modulating retro-reflector and receiver mounted in micro-gimbals.

The ground station used a bistatic tracking laser interrogator mounted on a motorized gimbal. The systems had separate 10 cm transmit and receive apertures, and a maximum output power of 2 watts [37]. Using GPS information, sent over a low bandwidth RF link from the UAV, the motorized gimbal provided coarse pointing, putting the aircraft into the field of view of the interrogator's fine steering mirrors, which then optically tracked the modulating retro-reflector. During flight tests a bidirectional 2 Mbps Ethernet link was established and both live video and data files were transferred.

A later test was conducted using another small UAV. During this test, in addition to data and video downlinks, atmospheric scintillation data was measured on the link. At the start of the test a ground-based measurement of scintillation was taken using an interrogator with a 3 cm aperture and a 1 cm retro-reflector. This data was taken at height of about 2 meters above the ground and a range of about 500 meters. The ground-level scintillation index was 0.38 and the frequency knee of the scintillation was about 30 Hz.

Scintillation data was then taken, using the 10 cm aperture tracking terminal, of the optical return of the MRR in flight on the UAV. Figure 13.11 shows the power spectral density plot of the scintillation at a range of 4.3 km. Despite a range that is ten times longer, the scintillation index, σ_1^2 , at 0.366, is about the same as for the 500 m ground based link. The frequency knee of the airborne link however is about 300 Hz, ten times higher than the ground-based link.

This data shows some of the unique challenges faced by airborne links as opposed to ground-based links. In general, though the depth of scintillation fades should be less for airborne links, their frequency should be much faster. This may require different choices of protocols for efficient transfer of data and will certainly impact pointing and tracking systems.

While the majority of modulating retro-reflector systems that have been demonstrated have been for atmospheric links, there have been some analyses and demonstrations of

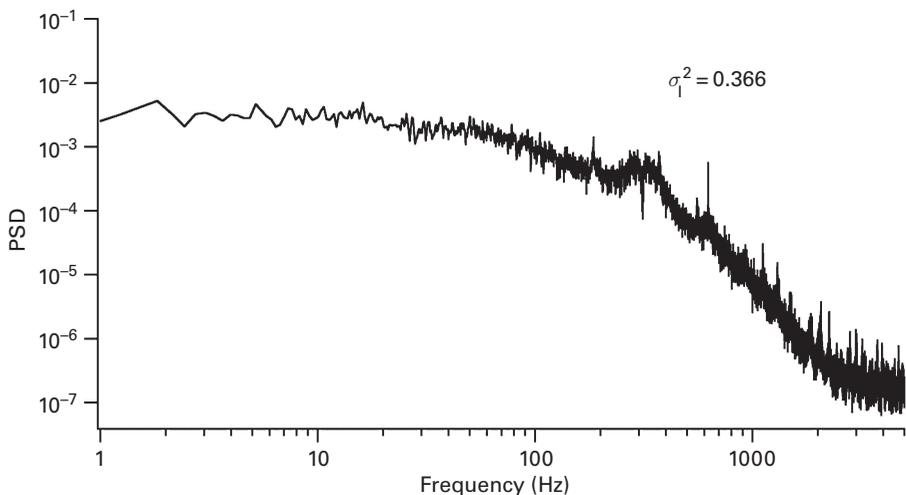


Figure 13.11 The power spectral density of optical scintillation of a ground to air modulating retro-reflector link at a range of 4.3 km.

underwater MRR links [42, 43]. Like atmospheric MRRs, underwater MRRs are applicable to small platforms or unattended sensors. MRR links still provide the advantages of loose pointing tolerances on the retro-reflector end of the link. Despite the fact that underwater propagation produces a much wider beam profile than diffraction-limited atmospheric propagation, underwater links still need to be aimed to within a fraction of a degree to maximize their link range. Propagation losses are very different for underwater links. The $1/R^4$ link dependence of retro-reflector links is dominated by the exponential losses due to scattering and absorption. Underwater MRRs have more limitations on their modulator types. There is a relatively small window in the blue/green portion of the optical spectrum in which light propagates with low loss. While MEMS and liquid crystal devices can be designed to work in the blue/green, the most commonly used semiconductor materials cannot. Materials such as InGaN, which have a wide bandgap, are being investigated and may be used for blue/green MQW modulators in the future. However, these materials are not mature enough for use now. Underwater MRR links must be very short due to the high propagation losses. It is unlikely that link ranges more than 100 meters will be possible, even in the cleanest of water. Within this range however lie several useful applications, such as data transfer at megabit per second rates without the need to make a physical connection.

13.6 Conclusion

Modulating retro-reflectors have a unique niche in the world of free-space optical links. They share characteristics with direct FSO links and RF links. Like direct FSO links they are secure and do not suffer from frequency allocation issues. Like RF links their bandwidths tend to be below 100 MHz and they have loose pointing requirements (on one end at least). MRR links are unique in the low power consumption of the retro-reflecting transceiver, but also in their limited range due to the $1/R^4$ range dependence.

Modulating retro-reflectors, like direct free-space optical links in general, are still in search of the applications that will lead to wide deployment of systems. This is likely to occur as more and more high bandwidth sensors and communication links are used in the field. Increasing frequency congestion and the demand for even higher bandwidth leads naturally to optical systems as an answer. Retro-reflecting links will be part of that answer.

References

- [1] L. Stotts, B. Stadler, G. Lee, "Free space optical communications: coming of age," *Proceedings of the SPIE – The International Society for Optical Engineering* 5550, W1–W15 (2008).
- [2] H. Grubb, "Improvement in and means for signaling and indicating position of objects," UK Patent #21,856, GB190321856, issued October 6, 1904.
- [3] H. Stockman, "Communication by means of reflected power," *Proceedings of the I.R.E.*, 36 (10) pp. 1196–1204 (1948).

- [4] M. L. Biermann, W. S. Rabinovich, R. Mahon, G. C. Gilbreath, "Design and analysis of a diffraction-limited cat's-eye retroreflector," *Opt. Eng.*, 41(7) pp. 1655–1660, (2002).
- [5] G. O. Olsen, H. W. Mocker, N. A. Demma, J. B. Ross, "Coherent CO₂ laser communication system with modulable retroreflectors," *Appl. Optics*, 34 (12) pp. 2033–2044 (1995).
- [6] C. M. Swenson, C. A. Steed, I. A. A. DeLaRue, R. Q. Fugate, "Low power FLC-based retro-modulator communications system," *Proceedings of the SPIE – The International Society for Optical Engineering*, 2990 pp. 296–310 (1997).
- [7] D. C. O'Brien, J. J. Liu, G. E. Faulkner *et al.*, "Design and implementation of optical wireless communications with optically powered smart dust motes," *IEEE J. Sel. Area Comm.*, 27 (9) pp. 1646–1653 (2009).
- [8] F. Kullander, P. Sakari, L. Sjöqvist, J. Ögren, "System trials with modulated retro-reflective optical communication," *Proceedings of the SPIE – The International Society for Optical Engineering*, 5986 pp. 12-1–12-2 (2005).
- [9] D. Pedersen, O. Solgaard, "Free space communication link using a grating light modulator," *Sensor Actuat. A-Phys.*, 83 (1–3) pp. 6–10 (2000).
- [10] C. P. Lao, K. W. Goossen, "Optical microelectromechanical system array for free-space retrocommunication," *IEEE Photonic Tech. Lett.*, 16 (9) pp. 2045–2047 (2005).
- [11] T. K. Chan, J. E. Ford, "Retroreflecting optical modulator using an MEMS deformable micromirror array," *IEEE J. Lightwave Technol.*, 24 (1) pp. 516–525 (2006).
- [12] C. Jenkins, J. G. Brown, L. Li, W. Johnstone, D. Uttamchandani, "MEMS retro-phase-modulator for free-space coherent optical communications," *IEEE J. Sel. Top. Quant. Electron.*, 13 (2) pp. 330–335 (2007)
- [13] L. Ziph-Schatzberg, T. Bifano, S. Cornelissen, "Secure optical communication system utilizing deformable MEMS mirrors," *Proceedings of the SPIE – The International Society for Optical Engineering*, 7209, C1–C15 (2009)
- [14] I. J. Fritz, T. M. Brennan, B. E. Hammons *et al.*, "Low-voltage vertical-cavity transmission modulator for 1.06 μm ," *Appl. Phys. Lett.*, 63 (4) pp. 494–6 (1993)
- [15] G. C. Gilbreath, W. S. Rabinovich, R. Mahon *et al.*, "Large aperture quantum well shutters for fast retroreflected optical data links in free space," *Proceedings of the SPIE – The International Society for Optical Engineering*, 3707 pp. 666–672, (1999)
- [16] W. S. Rabinovich, G. C. Gilbreath, P. G. Goetz *et al.*, "InGaAs multiple quantum well modulating retro-reflector for free space optical communications," *Proceedings of the SPIE – The International Society for Optical Engineering*, 4489 pp. 190–201 (2001).
- [17] S. Junique, D. Agren, Q. Wang *et al.*, "Modulating retro-reflector for free-space optical communication," *IEEE Photonic Tech. Lett.*, 18 pp. 85–87 (2006).
- [18] M. Plett, W. S. Rabinovich, R. Mahon *et al.*, "Free-space optical communication link across 16 km over the Chesapeake Bay to a modulated retro-reflector-array," *Opt. Eng.* 47, 045001 (2008).
- [19] S. G. Lambert and W. Casey, *Laser Communication in Space*, (Boston, Artech House, 1995).
- [20] W. S. Rabinovich, R. Mahon, H. R. Burris *et al.*, "Free-space optical communications link at 1550 nm using multiple quantum well modulating retro-reflectors in a marine environment", *Opt. Eng.*, 44 pp. 056001 1–12 (2005).
- [21] W. S. Rabinovich, P. G. Goetz, R. Mahon *et al.*, "45-Mbit/s cat's eye modulating retroreflectors," *Opt. Eng.*, 46 pp. 104001 1–8, (2007).
- [22] P. G. Goetz, R. Mahon, L. Swingen, J. L. Witkowsky, W. S. Rabinovich, "Coherent modulating retroreflector system," *Proceedings of the SPIE – The International Society for Optical Engineering* 5550 pp. 115–121 (2004).

- [23] P. G. Goetz, E. E. Funk, R. Mahon, W. S. Rabinovich, S. C. Binari, "Quadrature amplitude modulation in modulating retroreflector system," *Electron. Lett.*, 39 pp. 1463–1464 (2003).
- [24] W. S. Rabinovich, R. Mahon, P. G. Goetz, E. Waluschka, "A cat's eye multiple quantum-well modulating retro-reflector", *IEEE Photonic Tech. Lett.*, 15 pp. 461–463 March 2003.
- [25] M. B. Klein, R. H. Sipman, "Large aperture Stark modulated retroreflector at 10.8 μm ," *J. Appl. Phys.*, 51 pp. 6101–6104, (1980).
- [26] W. S. Rabinovich, C. I. Moore, H. R. Burris *et al.*, "Free space optical communications research at the US Naval Research Laboratory," *Proceedings of the SPIE – The International Society for Optical Engineering*, 7585 pp. 02-1–02-15, (2010).
- [27] T. H. Stievater, W. S. Rabinovich, P. G. Goetz, , R. Mahon, S. C. Binari "A surface-normal coupled-quantum-well modulator at 1.55 μm ," *IEEE Photon. Tech. Lett.* 16 pp. 2036–2038 (2004).
- [28] K. Sayyah, A. Narayanan, D. Persechini, P. Brewer, "Conformal pixellated MQW modulator structure for modulating retroreflector applications," *IEEE Photonic Tech. Lett.*, 17 pp. 1854–1856, (2005).
- [29] P. G. Goetz, W. S. Rabinovich, S. C. Binari, J. A. Mittereder, "High-performance chirped electrode design for cat's eye retro-reflector modulators," *IEEE Photonic Tech. Lett.*, 18 pp. 2278–2280 (2006).
- [30] D. Klotzkin, P. G. Goetz, W. S. Rabinovich *et al.*, "Integrated angle-of-arrival sensing and simultaneous bidirectional communication using a cat's eye modulating retroreflector," *Conference on Lasers and Electro-optics*, article#5225567 (2009).
- [31] W. S. Rabinovich, T. H. Stievater, N. A. Papanicolaou, D. S. Katzer, P. G. Goetz, "Demonstration of a microelectromechanical tunable asymmetric Fabry-Perot quantum well modulator", *Appl. Phys. Lett.*, 83 pp. 1923–1925 (2003).
- [32] T. H. Stievater, D. Park, M. W. Pruessner, "A microelectromechanically tunable asymmetric Fabry-Perot quantum well modulator at 1.55 μm ", *Opt. Exp.*, 16 pp. 16766–16773 (2008).
- [33] K. W. Goossen, D. W. Prather, "Fabrication and yield of large-area quantum-well modulators", *IEEE Photonic Tech. Lett.*, 19 pp. 2054–2056 (2007).
- [34] W. S. Rabinovich, J. Murphy, M. Ferraro, R. Mahon, "Free-space optical data link to a small robot using modulating retroreflectors," *Proceedings of the SPIE – The International Society for Optical Engineering*, 7464 pp. 08-1–08-9 (2009).
- [35] D. Varshneya, L. Jeffers, "Temperature compensated dynamic optical tag modulator system and method," US Patent #7603041, October 13, 2009.
- [36] S. Arnon and D. Kedar, "Sensing and communication trade-offs in picosatellite formation flying missions," *JOSA A*, 26 pp. 2128–2133 (2009).
- [37] E. Saint Georges, "Evaluation of cone tracking for free-space optical communication with a retro-modulator," *Proceedings of the SPIE – The International Society for Optical Engineering* 7464 pp. 70903–70903 (2009).
- [38] R. Mahon, W. S. Rabinovich, M. Plett, *et al.*, "Interference effects and aperture averaging in retroreflected light," *Opt. Eng.*, 47 pp. 046002-1–046002-9 (2008).
- [39] G. C. Gilbreath, W. S. Rabinovich, T. J. Meehan *et al.*, "Large-aperture multiple quantum well modulating retroreflector for free-space optical data transfer on unmanned aerial vehicles," *Opt. Eng.*, 40 pp. 1348–1356 (2001).
- [40] P. G. Goetz, W. S. Rabinovich, R. Mahon, "Modulating retro-reflector lasercom systems at the Naval Research Laboratory" (invited), *Proc. MILCOM 2010*, pp. 2302–2307, November 2010.

- [41] P. G. Goetz, J. L. Murphy, M. R. Suite, "Lasercom for small unmanned aerial systems using a modulating retro-reflector," *MILCOM 2010 Classified Session*, November 1, 2010.
- [42] L. Mullen, B. Cochenour, W. Rabinovich, R. Mahon, J. Muth, "Backscatter suppression for underwater modulating retroreflector links using polarization discrimination", *Appl. Opt.*, 48 pp. 328–337 (2009).
- [43] S. Arnon, "Underwater optical wireless communication network," *Opt. Eng.*, 49 pp. 015001–015007 (2010).

14 Visible-light communications

Kang Tae-Gyu

Visible-light communications is the name given to a wireless communication system that conveys information by modulating light that is visible to the human eye. Communications may not be the primary purpose of the light; in many applications the light primarily serves as a source of illumination. Interest in VLC has grown rapidly with the growth of visible-light light emitting diodes (LEDs) for illumination. The motivation is clear: When a room is illuminated by LEDs, why not exploit it to provide communications as well as illumination? This sharing of resources can save electric power and raw materials.

14.1 VLC principle

A simple means for achieving visible-light communications is to switch the LED lighting on and off at a speed higher than is perceptible to the human eye. Eyes are organs that can detect changes in light brightness and power when these changes occur over a long time scale, but they cannot perceive light that is switched on and off rapidly, say at 200 Hz or more, depending on the eye. A photodiode, on the other hand, can easily recognize the the rapid on–off modulation. A photodiode is a photodetector that produces an electrical current that is proportional to the optical power that is incident on the photodetector surface. This simple principle makes possible visible-light communication technology that supports both illumination and wireless communication using an LED (see Figure 14.1).

The dual function of illumination and communications means that the user can always recognize a communications area. Unlike RF wireless communication, where specialized tools are needed to find a service area, the presence of a VLC service area will be easily detected.

The visible spectrum is the portion of the electromagnetic spectrum that is visible to the human eye. A typical human eye will respond to wavelengths from about 380 nm to 780 nm. In terms of frequency, this corresponds to a band in the vicinity of 385–789 THz.

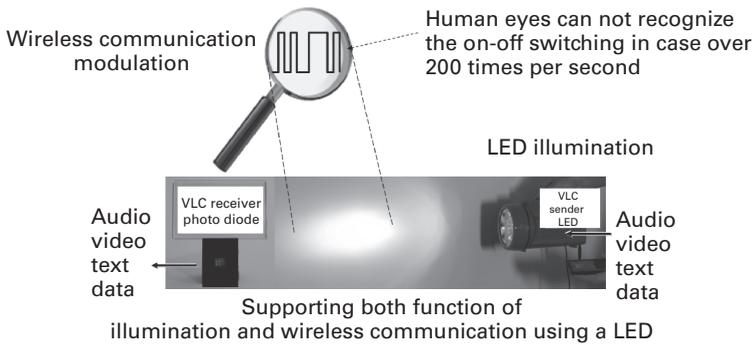


Figure 14.1 Principle of visible light communications.

14.1.1 VLC key components

The components of visible-light communications are the LED (light emitting diode), PD (photodiode), PHY (physical layer), and MAC (media-access control). The LED converts an electrical signal to optical energy that provides illumination as well as communication. Information is line-encoded and modulated by the PHY and then conveyed on the optical signal by modulating the amplitude or some other feature of the LED light. At the receiver, the PD converts the received optical power to an electrical signal, which is then demodulated and decoded by the PHY layer to recover the user message bits. The MAC is the second layer in the Open Systems Interconnection seven-layer model to support a color packet scheme and media-access control, depending on applications (see Figure 14.2).

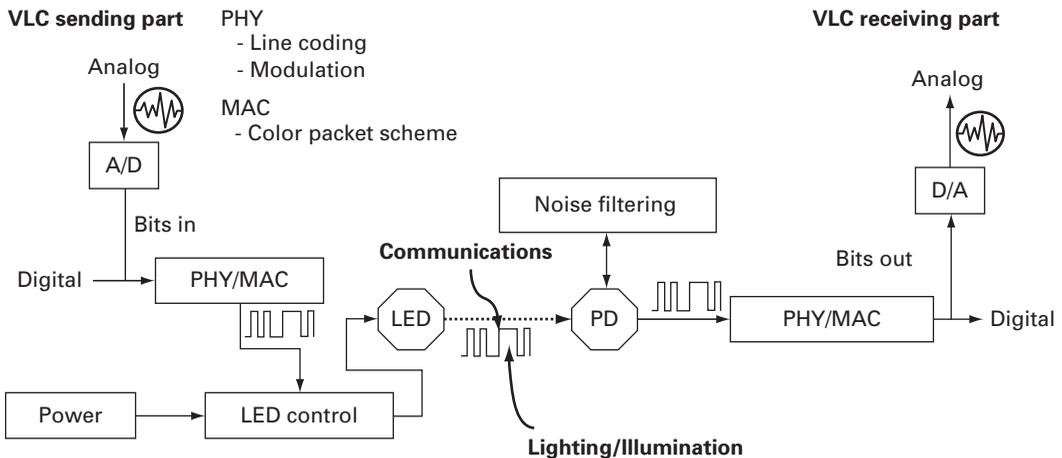


Figure 14.2 VLC PHY Architecture.

Dimming is achieved in a typical LED illumination system by adjusting the duty cycle of the LED drive signal. In other words, PWM (pulse-width modulation) is used for controlling power. To realize communications, we must encode data without disturbing

the duty cycle of the LED driver signal. There is a need for a dual-purpose modulation strategy that supports both functions of illumination and wireless communications.

VLC is advantageous over conventional RF wireless communications because optical signals do not create any ISM (industrial scientific and medical) interference, and because optical signals do not require a license from the FCC or any other regulatory body. VLC also offers some security advantages over radio, since visible light will not pass through solid walls or other opaque boundaries. This same property also prevents interference between separate VLC links in adjacent rooms.

One challenge for VLC is the presence of noise from secondary light sources such as the sun or computer screens. These sources can make it difficult to achieve a high signal-to-noise ratio (SNR). Their impact can be minimized through a careful design of the communication system, including the directionality, optical concentrators and lenses, optical filtering, and electrical signal processing.

14.1.2 VLC approaches

A new market for visible-light technology is driven by both the growing LED lighting market and the growth of green technology. LED is used as indicator lamps in many devices, and is increasingly used for lighting. LED lighting is used in applications as diverse as replacements for automotive lighting such as brake lamps and turn signals as well as traffic signals. Moreover, the LED market will expand to incandescent replacement lamps, sign board, TV BLU (back light unit), and outdoor exterior lighting systems [6].

VLC is a fusion technology between LED illumination and wireless communication. LED technologies are focused on electric power efficiency (lumen per watt) and long life for better lighting technology. Wireless communications technologies are focused on data transmission efficiency for higher data rates, higher reliability, and higher security. The fusion technology of visible-light communication makes new services for an under developed market using the LED light source (Figure 14.3).

Visible-light communication can be applied to machine-to-machine communication display in an LED back-light unit. We can use the wireless communication using visible-light communication even within a restricted wireless communication area, because a visible-light communication system does not cause interference from the restricted specific electric equipment's frequency spectrum.

The lighting in an intelligent transportation system can provide specific information to vehicles or man to guide a designed direction for safe driving. Any application by visible-light communication supports physical security because light cannot pass through opaque barriers.

14.1.3 VLC channel models with lighting reflective index

A VLC system must contend with multipath propagation, since the light from the LED will travel to the receiver along a direct line-of-sight path as well as along multiple reflected paths. A model for the impact of multipath propagation can be defined that

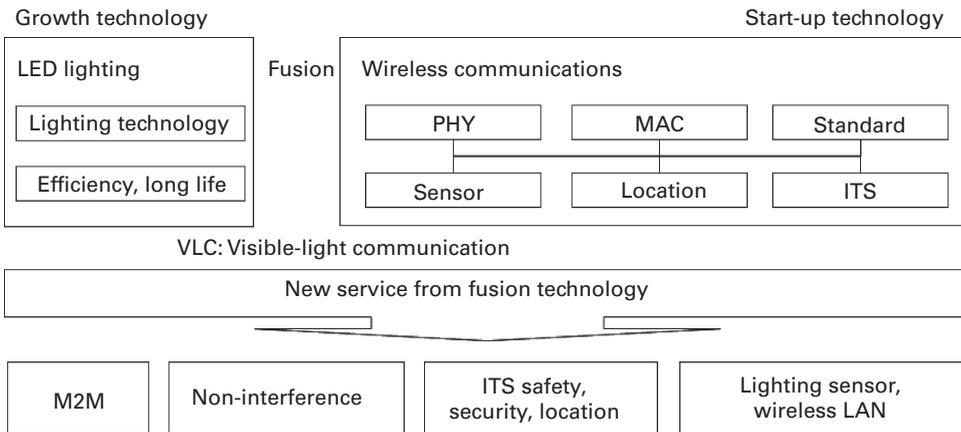


Figure 14.3 Fusion technology with LED illuminations and wireless communications.

accounts for the geometry of the room, the type of obstacles and reflectors, their reflective indices, and lighting sources [1, 8, 9, 11]. The reflective index for a cement wall that is painted an ivory color is measured as 67% in warm white LED and 75% in cool white LED. The reflective index for a vinyl floor with bright-ash color is measured as 61% in warm white LED and 53% in cool white LED. The reflective index for a draped partition wall with white-sky color is measured as 39% for both the warm and cool white LED.

It is assumed that the source of emission and the reflected points on the wall have a Lambertian radiation pattern [12, 13, 14]. The Lambertian emission means that the light intensity emitted from the source has a cosine dependence on the angle of emission with respect to the surface normal.

To survey the illuminance distribution of LED system, we assume two configurations for LED position on the ceiling. In the case of one transmitter, the position is the center of the ceiling, and for four transmitters, the transmitters are uniformly spread across the ceiling as shown in Figure 14.4 adapted from [5].

Figure 14.5 adapted from [5] shows the illuminance with one transmitter, the semi-angle at half power is 30 degrees. The maximum value of luminous flux in the center is 768.10 lx.

Figure 14.6 adapted from [5] is given for comparing our study with previous researches [13, 14]; the numerical calculation in illuminance distribution of our system at the semiangle of 70 degrees is shown in here. All parameters are the same.

14.2 VLC standards

14.2.1 IEEE 802.15.7 VLC

The IEEE Standard for local and metropolitan area networks – Part 15.7: Short-range wireless optical communication using visible light was published September 6, 2011 [16].

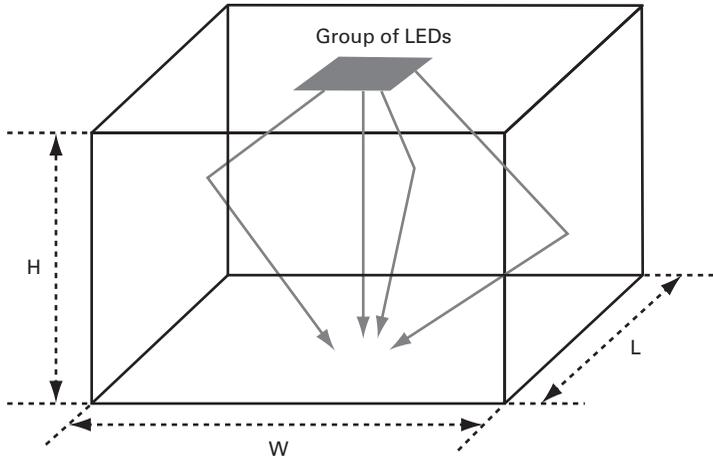


Figure 14.4 An example of indoor visible light communication environment.

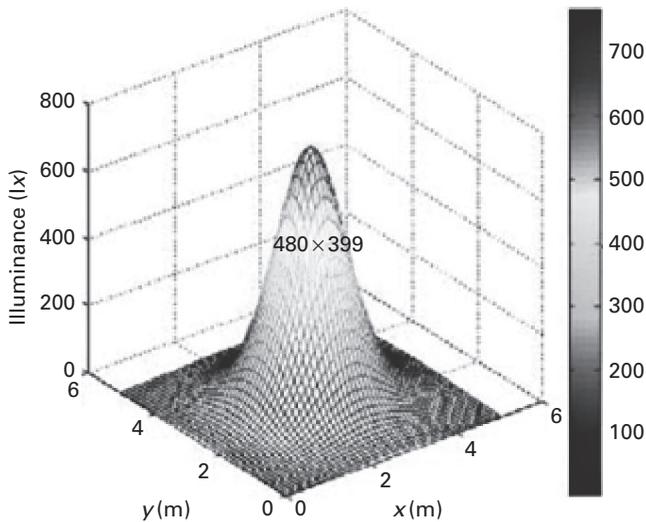


Figure 14.5 Distribution of illuminance in case of one transmitter.

The IEEE 802.15.7 Visible Light Communication Task Group had developed a PHY and MAC Standard for Visible Light Communications (VLC) in 2009; Study Group in 2008, Interest Group 2007 [4]. The standard defines a PHY and MAC layer for short-range optical wireless communications using visible light. The visible-light spectrum extends from 380 nm to 780 nm in wavelength. The standard is capable of delivering data rates sufficient to support audio and video multimedia services and also considers mobility of the visible link; compatibility with visible-light infrastructures; impairments due to noise and interference from, e.g., ambient light; health and other environmental effects; and a MAC layer that accommodates visible links. Potential

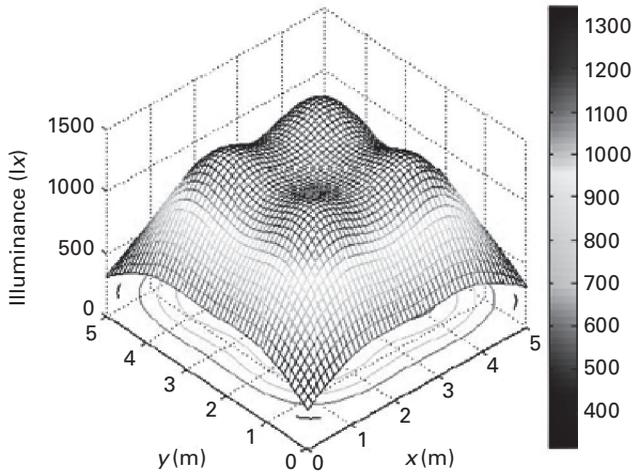


Figure 14.6 Distribution of illuminance with four transmitters with an LED semi angle of 70 degrees. (The maximum is 1342.50 lx, minimum is 315.90 lx, and average is 958.11 lx.)

applications include secure point-to-point communication, indoor Location Based Service (LBS), secure point-to-Multipoint communication (office, hospital, airplane), Intelligent Transportation System (ITS), and information broadcast.

IEEE 802.15.7 VLC has three different PHY that depends on the application: PHY I, PHY II, and PHY III. These three PHY specifications are summarized in Table 14.1.

PHY I is intended for outdoor use with low data rate applications. This mode uses on-off keying (OOK) and variable pulse position modulation (VPPM) with data rates in the tens to hundreds of kb/s.

PHY II is intended for indoor use with moderate data rate applications. This mode uses OOK and VPPM with data rates in the tens of Mb/s.

PHY III is intended for applications using color shift keying (CSK) that have multiple light sources and detectors with data rates in the tens of Mb/s.

The standard provides channel coding support for error correction. PHY I supports concatenated coding with Reed–Solomon (RS) and convolutional coding (CC), since it has been designed for outdoor use with short frames. PHY II and PHY III support only RS coding. PHY I and PHY II also support a run-length limited (RLL) code to provide DC balance, clock recovery, and flicker mitigation. In addition to modulation and coding, multiple optical rates are provided for all PHY types in order to support a broad class of optical transmitters (LEDs) for various applications. The PHY header shall be sent at the lowest data rate for the chosen clock rate. The clock rate does not change through the frame between the preamble, header, and payload.

IEEE 802.15.7 VLC supports LED illumination dimming function and flicker mitigation function.

There are five different dimming mechanisms: Idle pattern and compensation time dimming, visibility pattern dimming, CSK dimming, OOK dimming, and VPPM dimming.

Table 14.1 PHY specifications.

| PHY I | | | | | |
|-------------|-------------|--------------------|-----------------|-----------------|------------|
| Modulation | RLL code | Optical clock rate | FEC | | Data rate |
| | | | Outer code (RS) | Inner code (CC) | |
| OOK | Manchester | 200kHz | (15,7) | 1/4 | 11.67 kb/s |
| | | | (15,7) | 1/3 | 24.44 kb/s |
| | | | (15,7) | 2/3 | 48.89 kb/s |
| | | | (15,7) | none | 73.3 kb/s |
| | | | none | none | 100 kb/s |
| VPPM | 4B6B | 400kHz | (15,7) | none | 35.56 kb/s |
| | | | (15,7) | none | 71.11 kb/s |
| | | | (15,7) | none | 124.4 kb/s |
| | | | none | none | 266.6 kb/s |
| PHY II | | | | | |
| Modulation | RLL code | Optical clock rate | FEC | Data rate | |
| VPPM | 4B6B | 3.75 MHz | RS(64,32) | 1.25 Mb/s | |
| | | | RS(160,128) | 2 Mb/s | |
| | | 7.5 MHz | RS(64,32) | 2.5 Mb/s | |
| OOK | 8B10B | 15 MHz | RS(160,128) | 4 Mb/s | |
| | | | none | 5 Mb/s | |
| | | 30 MHz | RS(64,32) | 6 Mb/s | |
| | | | RS(160,128) | 9.6 Mb/s | |
| | | | RS(64,32) | 12 Mb/s | |
| | | 60 MHz | RS(160,128) | 19.2 Mb/s | |
| | | | RS(64,32) | 24 Mb/s | |
| RS(160,128) | 38.4 Mb/s | | | | |
| 120 MHz | RS(64,32) | 48 Mb/s | | | |
| | RS(160,128) | 76.8 Mb/s | | | |
| PHY III | | | | | |
| 4-CSK | none | 12 MHz | RS(64,32) | 12 Mb/s | |
| 8-CSK | none | | RS(64,32) | 18 Mb/s | |
| 4-CSK | none | 24 MHz | RS(64,32) | 24 Mb/s | |
| 8-CSK | none | | RS(64,32) | 36 Mb/s | |
| 16-CSK | none | | RS(64,32) | 48 Mb/s | |
| 8-CSK | none | | none | 72 Mb/s | |
| 16-CSK | none | | none | 96 Mb/s | |

Idle pattern and compensation time dimming allows an idle pattern to be inserted between the data frames for light dimming. The duty cycle of the idle pattern can be varied to provide brightness variation. Idle pattern and compensation time dimming adapts a compensation time (“ON” and “OFF” time of a light source) to be inserted into either

Table 14.2 Visible light wavelength band plan.

| Wavelength(nm) | Spectral width (nm) |
|----------------|---------------------|
| 380 ~ 450 | 70 |
| 450 ~ 510 | 60 |
| 510 ~ 560 | 50 |
| 560 ~ 600 | 40 |
| 600 ~ 650 | 50 |
| 650 ~ 710 | 60 |
| 710 ~ 780 | 70 |

the idle pattern or into the data frame to reduce or increase the average brightness of a light source.

Visibility pattern dimming adapts in-band idle patterns that are used in the payload of a CVD (color-visibility-dimming) frame. In order to generate high resolution visibility patterns from 0% to 100% in steps of 0.1%, there are certain constraints that need to be used in the design criteria for visibility patterns.

Color shift keying (CSK) supports VLC using multi-color light sources and photo detectors. CSK dimming employs amplitude dimming and controls the brightness by changing the current driving the light source. CSK dimming ensures the average optical power from the light sources is kept constant and keeps the center color of the color constellation with required intensity.

OOK modulation is sent with a symmetric Manchester symbol, compensation time needs to be inserted into the data frame to adjust the average intensity of the perceived source.

VPPM is a modulation scheme adapted for pulse-width based light dimming and offers protection from intra-frame flicker.

A compliant device shall operate with peak radiated energy within the visible light spectrum defined as being from 380 nm to 780 nm. A compliant device shall operate in one or several visible-light frequency bands as summarized in Table 14.2.

The bandplan is non-uniformly distributed across the visible spectrum to account for human eye sensitivity and optical transmitter (LED) manufacturing. LEDs are designed to have narrower bandwidths for middle wavelengths since the human eye is more sensitive to the center wavelengths of the visible-light spectrum. The codes in Table 14.2 are used to indicate the wavelengths containing the spectral peak for the transmitted frame and are indicated in the PHY header. This information may be used by the receiver for optimizing its performance.

14.2.2 VLCC VLC

The Visible Light Communications Consortium (VLCC) was established in November of 2003, with major companies in Japan [10]. The VLCC is aiming to publicize and standardize the visible-light communication technology, which has been discussed and evaluated in various industry fields. This ubiquitous human-interface technology

provides the communication capability through lighting in our offices and homes, LED commercial displays, LED traffic signals, LED small lamps on electronic home appliances, etc, in our daily life. And, this technology can also enable a highly accurate navigation system with the capability to locate with pinpoint accuracy. Therefore, VLC can help people in the elderly society as well as the robot future society. The “visibility” of VLC adds additional value to our safety, security and human comfort. In other words, the VLC gives people “light.” The devices, backed by the development of semiconductor, are originated in Japan.

14.2.3 TTA VLC

Since its foundation in 1988, TTA has been contributing to the advancement in related industries by way of standardization efforts and testing & certification services in the field of information and communication [15].

VLC (Visible Light Communication) Working Group (WG4021) in TTA started in May 2007. VLC WG developed TTA 5 VLC standard specifications in 2008: Basic Configurations of Transmitter PHY for Visible Light Communication, Basic Configurations of Receiver PHY for Visible Light Communication, Basic Configuration of LED Interface for Illumination and Visible Light Communication, Basic Configuration of Light Location Information Service Model using Visible Light Communication, and Basic Configuration of Lighting Identification for Visible Light Communication.

VLC WG developed TTA 16 VLC and LED control related standard drafts in 2010. The 16 specifications are focused on the converged technologies between LED illumination and visible-light communications.

14.3 VLC research and development

14.3.1 VLC modulation with dimming control

VLC needs a different modulation strategy that is compatible with dimming control of illumination. One example of such a strategy is variable pulse-position modulation (VPPM). VPPM is a modulation scheme that is compatible with dimming control that varies the duty cycle or pulse width to achieve dimming, as opposed to amplitude [2].

VPPM combines 2-PPM with PWM for a dimming control. Bits “1” and “0” in VPPM are distinguished by the position of a pulse, whereas the width of the pulse is determined by the dimming ratio. The principles of VPPM are illustrated in Figure 14.7 and Figure 14.8.

14.3.2 VLC line coding

The 4B6B line code expands each block of 4 bits into an encoded block of 6 bits with DC balance. This means there will always be precisely three zeros and three ones in each block of 6 encoded bits. Table 14.3 defines the 4B6B code [3].

Table 14.3 4B 6B Line Coding for non flickering.

| | 4B | 6B |
|---|------|--------|
| 0 | 0000 | 001110 |
| 1 | 0001 | 001101 |
| 2 | 0010 | 010011 |
| 3 | 0011 | 010110 |
| 4 | 0100 | 010101 |
| 5 | 0101 | 100011 |
| 6 | 0110 | 100110 |
| 7 | 0111 | 100101 |
| 8 | 1000 | 011001 |
| 9 | 1001 | 011010 |
| A | 1010 | 011100 |
| B | 1011 | 110001 |
| C | 1100 | 110010 |
| D | 1101 | 101001 |

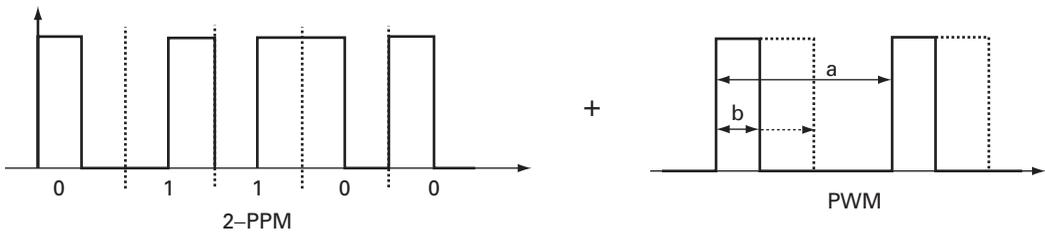


Figure 14.7 The principle of variable pulse position modulation.

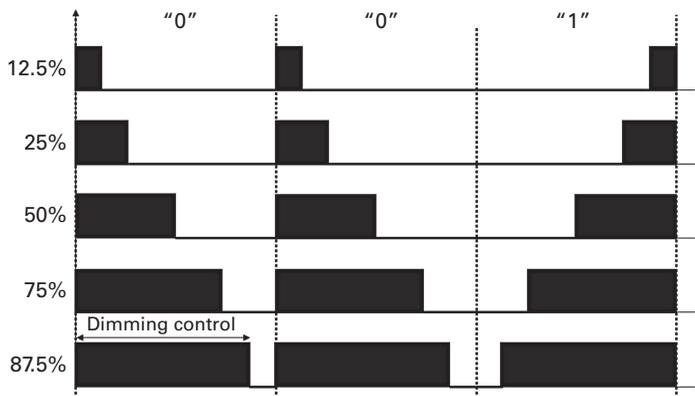


Figure 14.8 An example of variable pulse position modulation with a dimming control (after reference [2]).

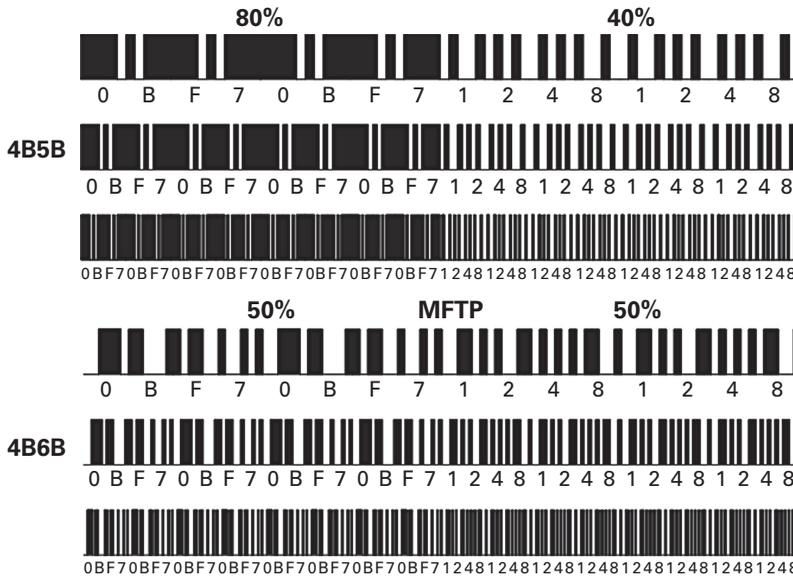


Figure 14.9 An example of comparison MFTP different ratio.

The features of the 4B6B code are the same 50% duty cycle during one encoded symbol for DC balanced run length limiting code and an error detection capability.

The maximum flickering time period (MFTP) is defined as the maximum time period over which the light intensity can be changing but the resulting flicker is not perceivable by the human eye. The 4B6B produce even 0 and 1 codes in a MFTP, regardless of the data, whereas 4B5B produces a ratio of 40% or 80% in MFTP, depending on the data. The even line coding prevents flickering. See Figure 14.9.

14.3.3 VLC video prototyping

There have been several visible-light communications prototyping efforts. One illustration is shown in Figure 14.10, which shows a VLC prototype that can transmit audio, text, numbers, and video. The 4-by-4 red LED array on the left transmits 10Mbps for video and audio. The white lamp with RGB LED on the right transmits text or numbers for an LED illumination identification service.

We can implement with different techniques depending on LED structures; indicator LED, high power LED, RGB white LED, and yellow phosphor white LED. We should also consider illumination structure; connection structure in sign board and LED driving control structure.

14.4 VLC applications

There are already a wide variety of existing wireless communication technologies, such as IEEE 802.11 wireless LAN, IEEE 802.15 wireless PAN, and IEEE 802.16 wireless



Figure 14.10 VLC video demonstration.

MAN. Although they share many similarities, their target applications are fundamentally different. Likewise, visible-light communications is focused on a specific application to support intelligent illumination with lighting as well as wireless communication.

14.4.1 VLC guidance system

In a VLC guidance system, lighting lamps that illuminate a yard, national border, or facility can also be used for guidance as well as for protection from outside attacks. The lamps have an identification number (VLC ID or LED ID) and guidance information.

There are two different illumination strategies for security guiding. One is fixed position with the lamp ID. The other is mobile position with the lamp ID. Visible light communication can provide specific information with illumination ID. We can define specific information such as when to turn the lights on, brightness value, temperature value, facility build date, lamp installed date, and geographic position coordinates. We can make an application for an intelligent security guidance system using the information through light by visible-light communication (Figure 14.11).

Street lamps with visible-light communications provide location or status information of the street. VLC street lamps can enable a convenient information system that can support advertising as well as geographical guidance for pedestrians or vehicles.

14.4.2 VLC color imaginable system

Color often prompts us to action. Color conveys information itself, whether from instinct or education. When we see a left arrow at a traffic signal, for example, we know we can turn left by instinct. The red color at a traffic signal tells us to stop and wait. A number countdown on a signal indicates how many seconds remain before a state change (Figure 14.12).

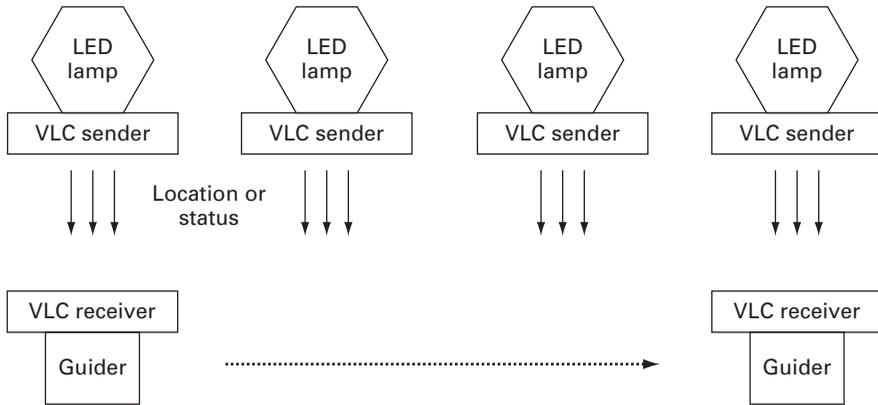


Figure 14.11 Guidance assistance with VLC.

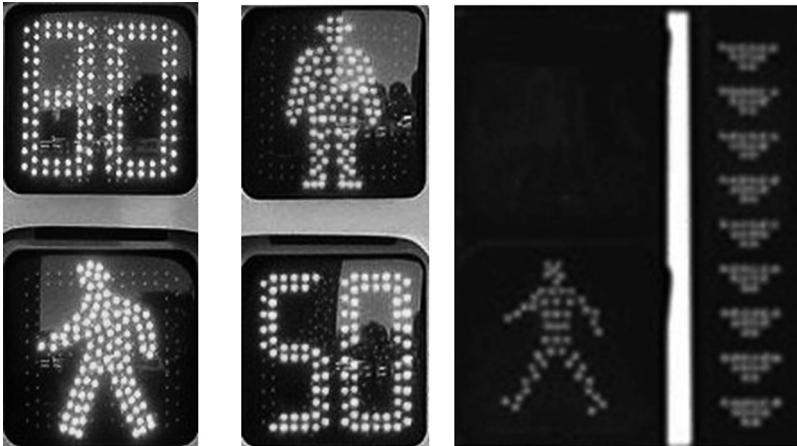


Figure 14.12 Color and information with VLC.

A lamp with visible-light communication provides a user with a receiving device the information or the number. The information or the number utilizes an intelligent traffic system with digitalized data.

The traffic signal can communicate to a vehicle with intelligent ECU (Electronic Control Unit) system. The vehicle can drive by itself according to the traffic signal, without the automobile driver. Such a VLC system offers the promise of being faster, safer, and more convenient.

A VLC traffic signal can also be used to help a person who is visually handicapped.

14.4.3 VLC indoor navigator

A lamp with visible-light communication can provide a function for indoor navigations in areas where GPS are not supported. We can ask someone “Which light are you under?” instead of “Where are you?”

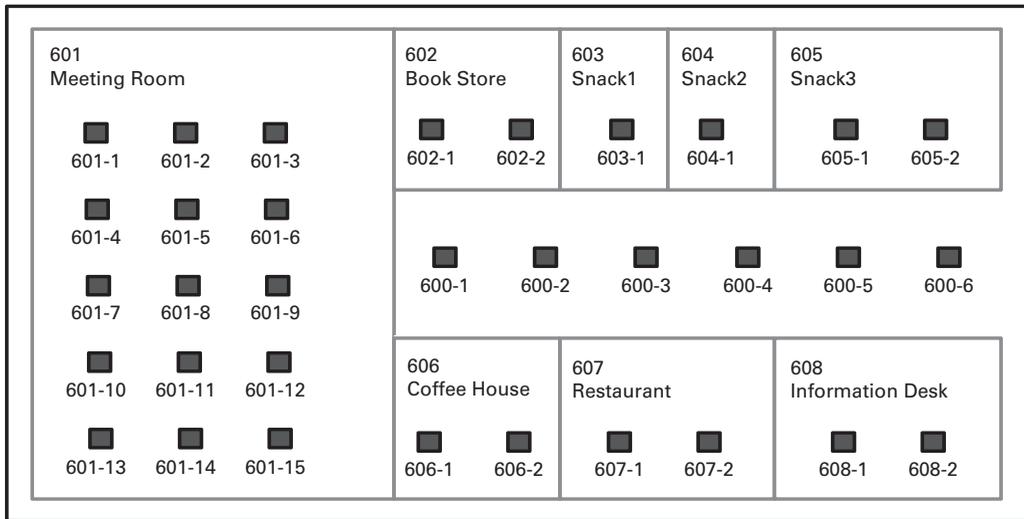


Figure 14.13 Indoor navigator with VLC.

LED lighting is increasingly being deployed in the real world. The convergence of LED lighting and IT technology is increasing the demand for LED. LED lighting is good for the environment, has a long lifetime, and gets excellent power efficiency compared with legacy illumination technologies. In addition, as smartphone usage increases, the market of object identifying information (such as bar codes, QR codes, and RFID) grows rapidly. The convergence of the LED lighting based on VLC and the object identification (ID) technology can open the door to new applications.

Many technologies can be considered for indoor-positioning navigation, such as Wi-Fi, ultrasonic active beacon, RFID tag, and infrared cameras for indoor areas where a GPS signal is not available. Most of all, if the location of VLC-LED lighting is pre-determined, it can be easily recognized where the receiver location is under a lighting fixture. This technology has the LED lighting function and at the same time delivers the positioning-ID information with light. Additionally, the ID information can also provide the status information of an object.

LED lighting with VLC must maintain the high brightness optical power and lack of flicker as required for illumination. At the same time, the VLC-LED lighting can broadcast information messages, for example, or positioning-ID, to receivers. Interestingly, LED illumination can use not only white light but also various colors with digital control technologies. Additionally, many services and applications can emerge when they combine the control networks, the control of LED illumination, and the VLC technology.

Regarding lighting control networks, DMX512 and DALI have been widely used to control dimming. As the use of various LED illuminations (for example: the wall washer, mood lighting, and the media facade) has increased the amount of control messages, the capacity and extension of a legacy for control networks could be limited.

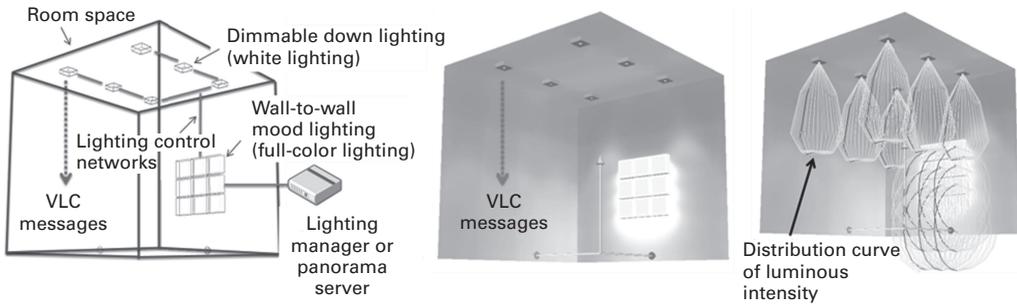


Figure 14.14 The configuration of VLC messages through LED lighting control networks.

In order to solve these limitations, Ethernet and IP-based networks have been widely adapted in lighting control networks as presented in many implementations of DMX over IP. In late 2006, ESTA completed an open, interoperable TCP/UDP/IP based lighting control protocol, known as ACN (Architecture for Control Networks, ANSI E1.17-2006). Soon a dweller can receive VLC messages through lighting control networks [7].

An example of a basic setup with the control network is shown in Figure 14.14. As outlined below, each fixture would be allocated with two kinds of lighting in a room. The first type is the dimmable down lighting of white light and the second type is the mood lighting using the full-color light which is available for the indicator lights, the interior lights, the media facades, kinetic art, etc. on the walls or floors. VLC messages which contain positioning-ID information can be delivered by the down lighting or the mood lighting. These LED lighting fixtures are connected to the control networks and controlled by the central equipment, or the lighting manager.

An emerging new VLC application is the VLC message broadcast system (VMBS), as shown as Figure 14.15. VMBS can transmit not only the VLC positioning-ID but also broadcast messages. A user can easily pair a smartphone with a VLC receiver using Bluetooth. When a VLC-LED lighting fixture broadcasts its positioning-ID, the VLC receiver can get the information under the lighting fixture. The location of each lighting fixture is pre-defined in the floor map. The VLC receiver delivers its own ID and positioning-ID into the smartphone via Bluetooth. The smartphone can report the VLC information to a server system through Wi-Fi or HSDPA. There are a variety of potential applications of VMBS, such as exhibition ID, vehicle parking management systems, and security systems.

14.4.4 VLC automobile driving support system

An automobile has many lamps that facilitate safe driving, including headlights, fog lights, turn signals, and brake lights. Augmenting these lamps with visible-light communication would open up new applications such as car-to-car communications. For example, when a driver sees brake lights on a preceding car, he or she knows that

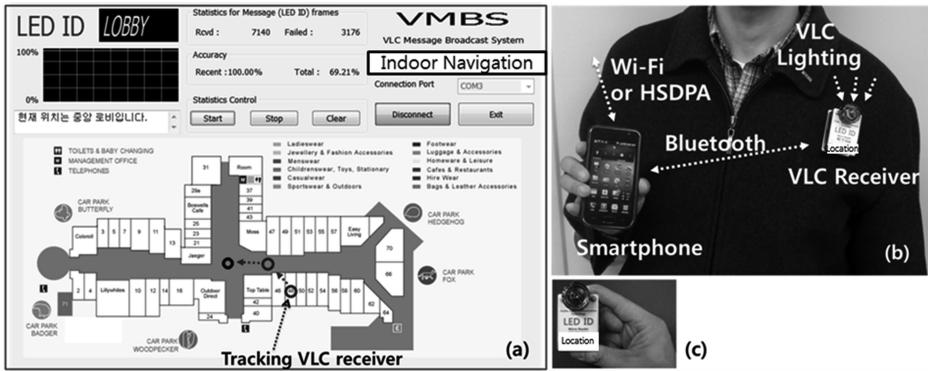


Figure 14.15 (a) Tracking VLC receiver, (b) Bluetooth pairing Smartphone and VLC receiver, (c) VLC receiver for positioning-ID messages. Photographs provided by the LED Communication Research Team of ETRI.

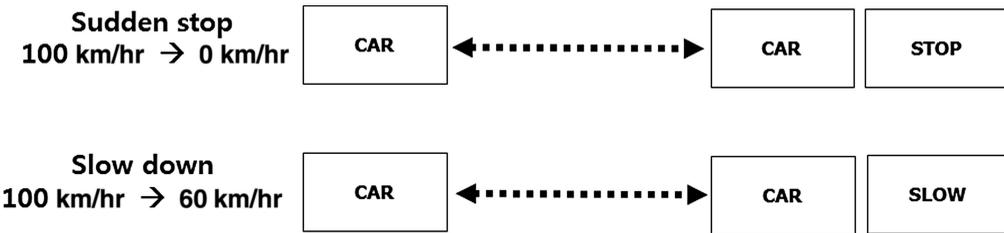


Figure 14.16 An example of car-to-car VLC applications.

the car is slowing down, but has no idea how quickly the car is decelerating. The brake light looks the same, regardless of whether there is a sudden emergency stop from 100 km/hr to 0 km/hr, or whether there is a slight decrease from 100 km/hr to only 60 km/hr, perhaps because the driver is slowing down to enjoy a scenic view (Figure 14.16).

A VLC automobile driving system could convey extra safety information, beyond that conveyed by the traditional lamps. More study of automotive VLC is needed, especially involving issues such as tracking and signal-to-noise ratios. In dense urban environments, a car is often confronted with numerous vehicle brake lamps with high speed in many lanes. Each lamp would have to be distinguished and tracked in order to achieve reliable communications without interruption. The light source would have a strong line-of-sight (LOS) component, which would require careful tracking. The noise produced by sunlight can severely impact the signal-to-noise ratio. During the night there are many man-made sources of noise to contend with, including street lights, vehicle lights, shop illumination, and LED sign boards. More research is needed to achieve an acceptable SNR under these varied conditions.

14.5 Future work

Visible-light communication technology is rapidly evolving. The promise is clear, but there remain significant technical obstacles that require further work before the promise can become a widespread reality. In many ways, VLC technology is still in its infancy. The wireless communications systems based on radio that are so widespread today have benefited from more than one hundred years of research and development. Wireless communications using the visible-light spectrum is poised to similarly expand and improve as research and development continues over the coming years.

One question that remains open is which modulation strategy is best suited to the VLC application. The pulse-width modulation strategy for controlling the brightness of the light (from an illumination perspective) fundamentally conflicts with the data-bearing requirements for VLC. Current strategies involve either pulse-width modulation, OOK, Manchester, or VPPM. More study is needed to find the best modulation strategy that optimally balances the conflicting requirements for illumination and communication.

VLC must operate with any kind of LED lighting system. Unfortunately, there are a wide variety of different functions, architectures, and driving mechanisms for LED light. There is no standard for LED lighting for VLC. VLC would benefit from either a strategy that adapts to different LED systems, or a unique standard for all LED lighting systems.

References

- [1] J. R. Barry, *Wireless Infrared Communications*, Kluwer Academic, 1994.
- [2] S. Lim, "ETRI PHY proposal on VLC band plan and modulation schemes for illumination," IEEE 802.15-09-0674-00-0007, 2009.
- [3] D. Kim, "ETRI PHY Proposal on VLC line code for illumination," IEEE 802.15-09-0675-00-0007, 2009.
- [4] www.ieee802.org/15/pub/TG7.html
- [5] H. Q. Nguyen, C. G. Lee, M. Kang *et al.*, "A MATLAB-based simulation program for indoor visible light communication system, *Communication Systems, Networks and Digital Signal Processing 2010*, 586–590.
- [6] T. G. Kang *et al.*, "White Paper: IEEE 802.15.7 VLC Regulations," July 2010, IEEE802.15 contribution 15-10-0615-00-0007, <https://mentor.ieee.org/802.15/dcn/10/15-10-0615-00-0007-white-paper-ieee-802-15-7-vlc-regulations.pdf>
- [7] Y. Kim, I. Kim, T. Kang, S. Park, "Analysis of IP-based control networks for LED lighting fixture communication," *New Trends in Information Science and Service Science (NISS)*, 4th IEEE Conference, 2010, 307–312.
- [8] D. O'Brien and M. Katz, "Optical wireless communications within fourth-generation wireless systems," *J. Opt. Network* 4, 312–322 (2005).
- [9] T. Komine, S. Haruyama, M. Nakagawa, "A study of shadowing on indoor visible-light wireless communication utilizing plural white LED lightings," *Wireless Pers. Commun.* 34, 211–225 (2005).
- [10] Visible Light Communication Consortium, www.vlcc.net

- [11] D.C. O'Brien *et al.*, Short-range optical wireless communications, Wireless World Research Forum, 2007.
- [12] J.M. Kahn and J.R. Barry, "Wireless infrared communications," *Proc. IEEE* 85 (2), 265–298 (1997).
- [13] T. Komine and M. Nakagawa "Fundamental analysis for visible light communication system using LED light," *IEEE Trans. Consum. Electron.* 50, 100–107 (2004).
- [14] T. Komine and M. Nakagawa, "Performance evaluation of visible-light wireless communication system using white LED lighting," *Proceedings of the Ninth IEEE Symposium on Computers and Communications*, 258–263 (2004).
- [15] www.tta.or.kr/English/
- [16] IEEE Std. 802.15.7-2011, "IEEE Standard for local and metropolitan area networks – part 15.7: Short-range wireless optical communication using visible light," 2011.

15 Optical wireless in sensor networks

Dominic C. O'Brien and Sashigaran Sivathanan

15.1 Introduction

Wireless sensor networks have been an active area of research for the past few decades [1, 2]. The availability of low-power radio communications, microprocessors, and sensors at low cost has led to highly capable sensor “motes” that can be connected in wireless networks (see for instance [3]). Environmental monitoring [2], monitoring of structures [4, 5], industrial process control [6], and mechanical systems [7] have all used wireless sensor networks, and such deployments are likely to become much more widespread. The next generation of appliances and objects will be part of the Internet of Things [8], and these will be connected wirelessly to each other and to an information infrastructure. Reduction in carbon emissions will require a smart electrical grid [9] that allows for local generation of power and control of its use within the home at the appliance level. This will in turn require large numbers of sensors and actuators which will both monitor and control generators and appliances.

In the field of healthcare the ageing population will lead to greater demands on resources, and home monitoring of patients has the potential to prolong life and quality of life whilst reducing demands on resources. Body area networks of sensors are being pursued as means to collect data to allow this [10].

A major challenge for sensor networks is that of prolonging sensor node lifetime, which is usually limited by battery capacity. An ideal sensor node would scavenge all the energy required from its environment, and therefore have an “infinite” operating life, barring failure. There are many schemes to do this, including vibration, air, solar, and radio frequency (RF) energy harvesting [11, 12]. However, most commercial sensor motes require external batteries, which limit the life of the node.

Energy is consumed for sensing, processing and communications, and the balance of the energy demands of each of these is complex to determine. However, communications using RF wireless forms a significant proportion of overall consumption in most sensor nodes. This is due to the almost isotropic “broadcast” nature of the RF channel, which leads to high path loss, and network architectures that require significant energy to be expended on route discovery [13]. In addition the carrier-based RF transmission requires relatively complex and power-hungry transmitter and receiver stages.

Careful management can reduce the need to communicate, by local processing of sensor data and communicating only when action is required, but communications still requires significant energy.

Point-to-point free-space optical (FSO) links have been successfully used between buildings [14], ships [15], aircraft [16], earth and space [17], and from air-to-ground [18]. Free-space optical communications typically uses highly directed channels, and these can have low path loss. In addition simple baseband modulation such as on-off-keying can be used, leading to low-complexity transmitter and receiver circuitry that can have low power consumption. The lenses and mirrors that shape the transmitted beam and capture received radiation in these links can have high gain but be physically small due to the small wavelength of the transmitted radiation.

There are therefore potential advantages in using such links for communications between sensors, but substantial technical challenges to be overcome. Some form of pointing and tracking is required for these links to ensure that the beam of light from a transmitter hits the required receiver, which might be placed anywhere within the coverage area. For most FSO links the terminals are fixed, so manual pointing is used when they are installed. Gimbal-based pointing and tracking are used in typical mobile systems. An alternative is to arrange multiple transmitters and receivers that each have a small field of view in such a way as to cover the much larger required field of view. Such an angle diversity configuration has been demonstrated in [19, 20]. A larger-scale network of atmospheric sensors that use optical communications is proposed in [21, 22]. A Scottish research consortium on “speckled computing” has been investigating various aspects of connection between small computing nodes or specks, including beam-steering [23, 24] and the power consumption of free space communication [25].

There has also been work on underwater sensor networks [26–30] and on topology control of free-space sensor networks [31]. Most of this work considers line-of-sight optical links, but diffuse networks have been proposed [29, 32]. In these ultra violet wavelengths in the solar blind region of the optical spectrum are used, and atmospheric scattering creates a non-line-of-sight (NLOS) channel to create a scattering channel, albeit at the cost of reduced range compared with a line of sight situation.

Retro-reflecting data links [33] (and see Chapter 13) simplify the design of such systems considerably. In this case light from a transmitting terminal propagates to the receiving terminal, which contains a retro-reflector. Light that strikes the retro-reflector is returned to the transmitter terminal by the action of the retro-reflector, creating the bi-directional optical paths required for communication. Information can be transmitted on the downlink path by modulating the source, and on the uplink by modulating the retro-reflector. This removes the need for a source and tracking at the retro-reflector end of the link. In addition the “natural” architecture becomes that of a complex terminal or base station that can direct beams of light to simple sensor nodes equipped with retro-reflectors and communicate with them.

This has the potential advantages of a simplified architecture, in that a central node communicates with all other nodes, and that the sensor nodes can potentially have low power consumption and complexity. However, a line of sight between nodes is required.

Also, the energy consumption is unbalanced in that the tracking node requires a greater energy supply, and this may not always be appropriate.

There have been a number of architectures that use this approach. In [34, 35] the term “smart dust” is first used to describe very small sensor nodes that could be scattered in an environment. In this architecture micro-electro-mechanical (MEMs) based retro-reflectors were proposed for communications [36]. Similar techniques are proposed in [30, 37, 38] and in [39, 40] acquisition techniques are studied.

In this chapter two architectures that use retro-reflectors are described, based on the use of retro-reflecting links. In the first optical links provide all the communications of the system, and in the second this is augmented by the use of short-range RF links. These are given as representative examples only, and are not meant to represent any “best” approach to using optical wireless in sensor networks. However, they do illustrate some of the challenges and potential advantages of using this technique.

15.2 Free-space optical (FSO) sensor network

15.2.1 Overview

Over the past few years a retro-reflecting architecture has been developed at the University of Oxford [41]. Figure 15.1 shows a diagram of the basic architecture. This uses a base station that steers beams of light to retro-reflecting sensor nodes. The nodes use a liquid crystal (LC) shutter-based modulated retro-reflector to modulate the returned beam. In the next sections the components of the architecture and some of the implementation challenges are discussed. The base station (BS) consists of a holographic beamsteering system [43] that directs light to the sensor mote. Light is returned to the

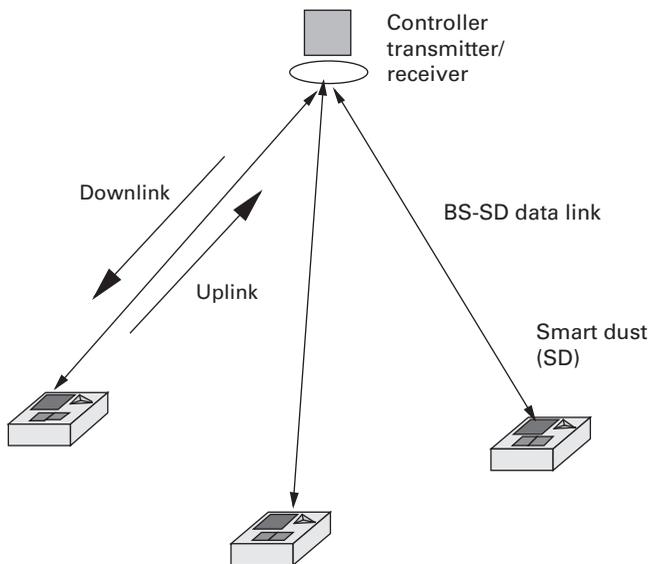


Figure 15.1 Smart dust architecture [42].

base station by a retro-reflector, and this returned power can be modulated by a liquid crystal shutter under control of the mote. This modulation is detected by a complementary metal oxide semiconductor (CMOS) camera in the base station, forming an uplink from mote to BS. A downlink is formed by modulating the BS source and using a receiver on the mote to detect this modulation.

In the implementation shown photodiodes on the mote provide power to operate the communications when illuminated by the beam from the base station. They can also “scavenge” energy from any ambient light that may be present.

15.2.2 Base station

The BS must be able to locate sensor motes and direct a beam of light towards them, in order to provide a communications path and illumination to power them. This might be achieved using a scanning mirror, or by mounting a source on a tracking gimbal [44]. However, these are limited to simple steering a single beam of light, and there is no ability to shape the beam or alter its divergence. Using a holographic beamsteering system allows much greater flexibility, as multiple beams can be independently directed, and defocus and other optical functions can be incorporated into the steering hologram [43]. Figure 15.2 shows the implementation of the BS, with such a programmable hologram implemented using a spatial light modulator (SLM) [43].

The BS uses a CMOS camera as an imaging receiver [45]. Each pixel on the camera receives light from a small range of angles within the coverage field of view, with the retro-reflected beams from the sensor nodes imaged as bright spots against any background radiation. This imaging configuration allows uplink communication with a number of motes simultaneously, as well as allowing their position to be determined. In addition ambient light collected by the imaging system is shared between the pixels of the camera, improving the signal-to-noise ratio at the pixels where a signal from a sensor is received compared with a single large-area detector. Such a detector would have the advantage of simpler readout electronics, but finding the sensor location would be more complex, and parallel communications would not be possible.

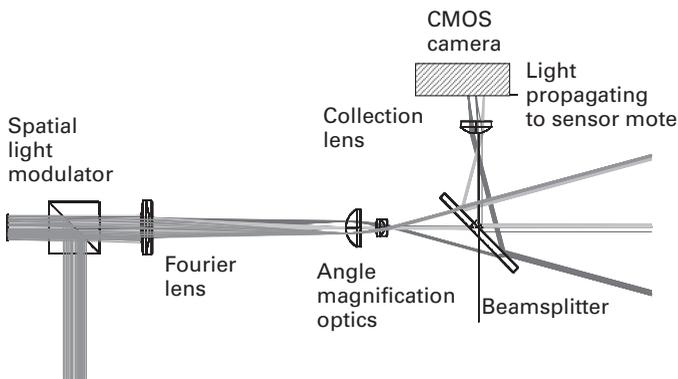


Figure 15.2 Base station implementation [41]. © IEEE2009.

In order to locate the motes within the coverage area the beam from the BS is scanned across the coverage area in a raster scan pattern. When the beam strikes a sensor node retro-reflector the strong return signal is imaged as a bright (relative to background) pixel on the CMOS camera in the BS, and this allows nodes to be located.

In order to speed up the acquisition of the motes a defocusing hologram is also written to the spatial light modulator, which creates a large search beam, thus decreasing the number of scans required to cover the system field of view. This beam does not have sufficient power density to power the motes and maintain the required link budget, and a much narrower divergence beam is used to do this, once the motes' position is known. Correction for different steering angle and barrel distortion caused by angle magnification optics can also be applied by modifying the programmable hologram.

15.2.3 Sensor node

Figure 15.3 shows the sensor node components, based around a CMOS integrated circuit (IC) that is fabricated in a $0.18\ \mu\text{m}$ mixed-mode process. The sensor nodes require a power supply, communications receiver and transmitter, processing and sensing capability. In the architecture shown there is no sensing capability implemented, due to limitations in time and resources available for this project.

Each aspect of the function of the motes is set out below.

15.2.3.1 Power supply

A dual-rail power supply is implemented using p-n junctions in the CMOS process used for the sensor IC. The system is designed to operate at wavelengths between 650–850 nm

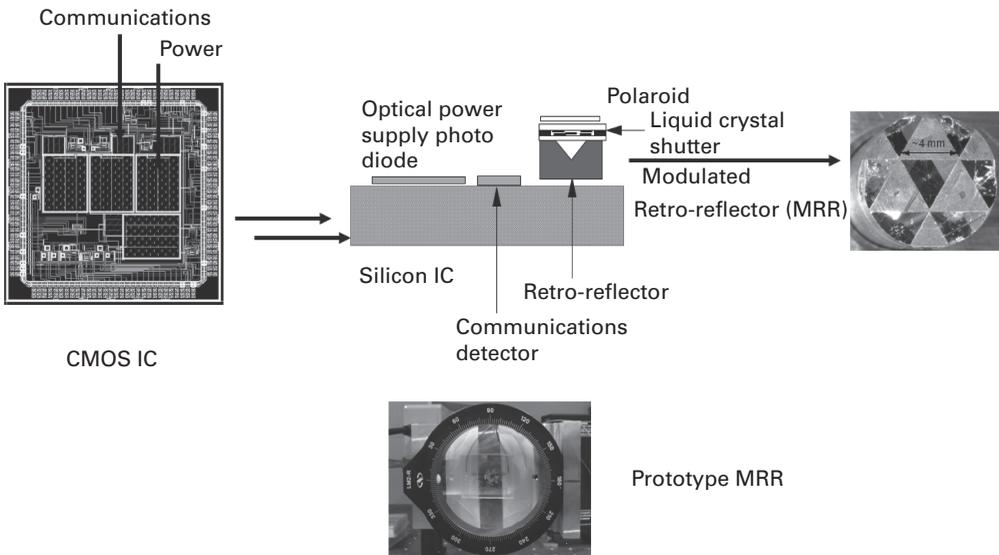


Figure 15.3 Sensor node architecture [42].

and in this region the absorption of radiation in silicon is poor, so deep p-n junctions are chosen to maximize collection of light. A test IC with all possible junctions available with the process was fabricated to assess these. The best junctions were chosen, and a 1.3 mm^2 Pwell/Twell junction was paired with a 0.2 mm^2 Nwell/Psub junction to form a power supply. Under typical illumination the available voltage varied from 0.6–0.9 V supplying 10–1000 nA [41].

It is possible to create stable voltage references from such supplies but these generally have high power consumption, so in this case circuit designs that operate with a wide range of different voltages were designed. In order to provide a stable time reference the data waveforms transmitted to the IC contain a reference clock.

15.2.3.2 Retro-reflecting modulators.

15.2.3.2.1 *Retro-reflectors*

There are a number of different designs of retro-reflector (see Chapter 13 for a review) but the corner-cube retro-reflector has the advantage that it is available from a number of vendors and a planar LC shutter can easily be integrated with it. In this case a commercial corner cube was machined to the dimensions required and this can be seen in Figure 15.3. The performance of this is likely to be far from the diffraction-limited performance that can be obtained using precision corner-cubes, but is sufficient for the demonstration described here.

15.2.3.2.2 *Liquid crystal shutter*

The liquid crystal shutter consists of an LC cell which is placed in front of the retro-reflector, with a single polarizer placed in the beam path in front of it, as shown in Figure 15.3. The cell consists of a layer of LC aligned between two glass plates coated with transparent electrodes to allow an electric field to be applied. In such devices there is a compromise between LC layer thickness (to achieve good optical path difference and contrast ratio), speed of operation, and voltage to achieve the required electric field. LC cells typically require at least several volts for operation, which is greater than that available here. Low-voltage parallel aligned nematic LC cells with a thickness of $3 \mu\text{m}$ were therefore fabricated. The cells were filled with a LC commonly used in displays and the waveforms to drive the cells are generated in the sensor mote IC. Using this configuration Manchester coded data at 30 bit/s with a voltage of 0.7 V was successfully transmitted on the uplink, indicating sufficient switching at approximately 60 Hz. Cells have a typical capacitance of $\sim 28 \text{ pF/mm}^2$ which leads to switching energies of $\sim 7 \text{ pJ/bit/mm}^2$.

15.2.3.3 Downlink receiver

The receiver needs to operate over a wide dynamic range under ambient light conditions, and was designed to do this without any optical filter to reject broadband radiation. A logarithmic detector [46] can be designed to respond to the contrast ratio of the incoming data stream rather than its absolute value, and a modified camera pixel design [46] that achieves this was developed. The downlink contains an 8 kHz reference clock and a 1 kbit/s Manchester data stream as shown in Figure 15.4. The different modulation depth

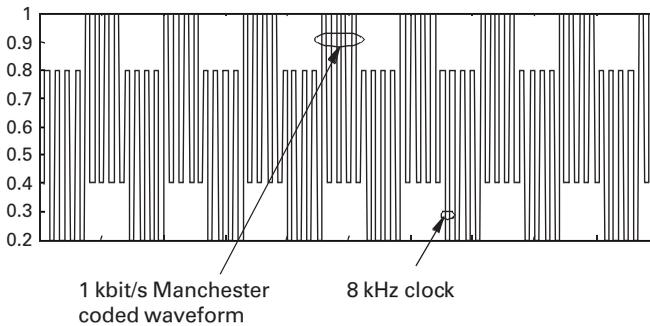


Figure 15.4 Downlink data stream [41] © IEEE2009.

and frequency allows receiver circuitry to separate the two waveforms and the data to be decoded. Measurements indicate the receiver operates successfully for photocurrents in the range of 100 pA to 100 nA.

15.2.3.4 Uplink transmitter

The uplink is controlled by a driver circuit that applies different voltages to the electrodes of the LC cell. These ensure that the net voltage (over time) on the LC cell is zero. A Manchester coding scheme is used to transmit data to make sure the modulation itself is balanced. At the BS a CMOS camera is used to decode the data. The acquisition algorithm described earlier identifies camera pixels that are receiving light from sensor nodes, and a region of interest (ROI) is set around these. Pixel data from this ROI is acquired at a high frame rate so that the low-rate uplink data is sufficiently oversampled, and subsequent data processing is undertaken to recover the received data.

15.2.4 Representative results

15.2.4.1 Link budget.

Successful communications requires that the downlink has sufficient signal to operate the receiver on the sensor mote, and that the BS receiver can decode the modulated uplink correctly. The limit on BS transmitted power is set by eye-safety regulation and in this case class 1 operation [47] was desired, limiting power to < 1 mW. The downlink receiver requires at least 100 pA of photocurrent for correct operation, and the camera has an effective sensitivity of approximately -60 dBm at 100 bit/s. Using this information, together with ray-tracing and analytical modeling predicts an operating range of several hundred meters for the communications links [41].

As designed the sensor nodes use the illuminating beam to provide power, although in practice equal amounts of power are scavenged from the ambient light in the environment. Ray tracing predicts a range of ~ 30 m or so before the beam divergence causes insufficient power to be collected at the sensor node, indicating that the range is limited by the optical power supply rather than the communications link budgets.

15.2.4.2 Loopback operation

At present the architecture has been tested in a loopback configuration over short ranges. To achieve this simple decoding has been built into the smart dust IC. This can decode an ASCII “U” and “D” symbol, corresponding to commands to switch the LC into “up” and “down” states. This modulation can then be detected at the BS using the CMOS camera, creating a loopback test for the link.

Figure 15.5 shows the downlink waveform sending a “U” signal at 1 kbit/s using Manchester coding, together with an 8 kHz clock. The output of the decoder, which is the control line to the LC modulator circuit, can be seen on the lower trace. This can be seen switching to a high level at the end of the symbol as expected.

A message consisting of successive “U” and “D” commands was sent from the BS to the sensor node. Figure 15.6 shows the resulting input to the LC driver, showing up/down transitions as would be expected. These are converted to the drive waveforms sent to the electrodes on the LC cell which can also be seen in the figure. The net voltage on the LC cell is seen on the bottom trace.

The LC cell modulates the optical signal returned to the BS via the retro-reflector, and this modulation is captured on the BS imaging receiver. Figure 15.7 shows the data received at the camera that is used as a detector. It can be seen this is a square wave as would be expected.

15.2.5 Discussion

This architecture shows the potential for very low energy consumption sensor nodes that can communicate over significant distances with a base station. At present the nodes have very limited functionality and only low data rate communications can be achieved. Providing greater functionality is likely to be relatively straightforward, as it can be implemented within the silicon CMOS used for the sensor circuitry.

Increasing the data rate involves more complex design trade-offs. At present the available drive voltage at the sensor node limits the data rate of the LC modulator. However, this is well matched to the frame rate of the CMOS camera, and software

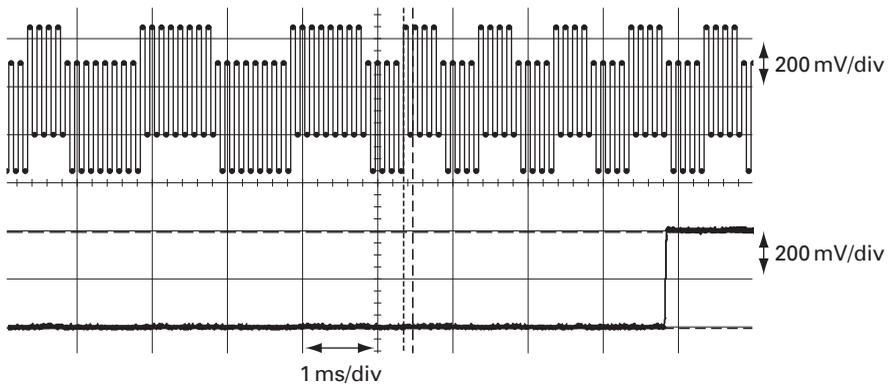


Figure 15.5 Downlink waveform [42].

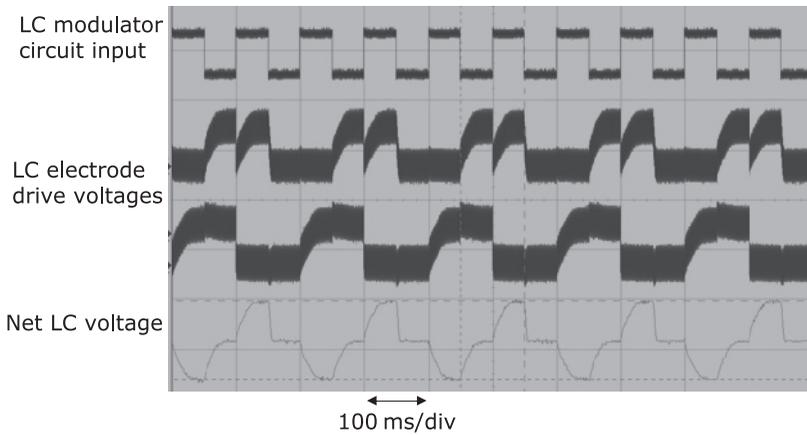


Figure 15.6 Waveforms from LC modulator driver [42].

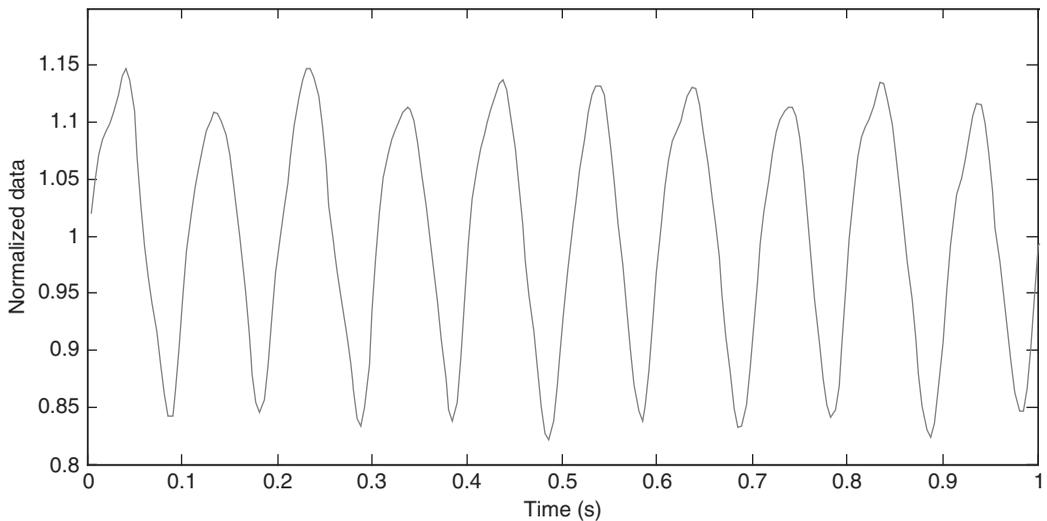


Figure 15.7 Waveform received at base station [42].

based data processing is feasible with a standard personal computer. Such a system has the potential to support communication with a large number of such sensor nodes “in parallel”.

An increased data rate is feasible with ferroelectric LC modulators, where kbit/s rates are possible [48]. This would require an order of magnitude higher drive voltages than the devices used here, together with higher power consumption, and the need for circuitry to boost the available node supply voltages. In addition this is likely to require a receiver array based imaging receiver, rather than the camera based approach used in this system, in order to provide sufficient bandwidth. The limited functionality of the

node makes it unsuitable for a practical application, but it shows that energy scavenging can provide sufficient energy for simple communications tasks.

The major disadvantage of the system is the requirement for a line of sight to a base station node, and in order to overcome this, a hybrid architecture that uses both RF and optical communications has been developed. Such hybrid architectures are attracting increasing research interest, as described in the following sections.

15.3 Radio frequency/Free-space optical (RF/FSO) sensor network system

15.3.1 Hybrid RF/FSO systems

RF/FSO systems have been researched for a number of purposes. Much of the interest in RF/FSO systems is currently devoted to wireless broadband networks [49–51], where the primary means of communications between terminals is a high-speed optical link. However, this requires line of sight (LOS) between the transmitter and receiver, which could be obstructed by clouds, fog, and snow [50]. To complement the optical link, a lower-speed radio link provides backup communications when LOS is obstructed. In the RF/FSO broadband system in [51], the FSO and RF link data rates are 1.25 Gb/s and 100 Mb/s respectively, with an average overall data rate of 183 Mb/s. The RF/FSO broadband system is very robust, and there are almost no obscuration conditions under which both the RF and the FSO links could fail simultaneously [49]. There has also been work on the reconfiguration and topology control of such networks [52, 53], where the application is tactical communications between both airborne and ground-based network nodes.

For wireless sensor networks an RF/FSO system (WSN) has been proposed [41, 54, 55], and is gaining considerable interest amongst researchers [56–58]. For the RF/FSO WSN, the primary means of communications is also a free-space optical link, with RF links serving as backups in the absence of LOS for optical communications. However, unlike the high data rate offered by broadband optical communications, the motivation for using FSO links in WSNs is the low communications energy associated with optical communications. The architecture in the previous section communicates with energies of 10s of pJ/bit, over ranges potentially up to 100s of meters, compared with an order of up to 100 pJ/bit for distances of up to 10 m [59, 60] for networks using RF communication. This is significant as RF sensor nodes expend most energy on communications (compared with sensing and data processing) [2].

In the next section such an architecture is described.

15.3.2 Architecture

Figure 15.8 shows the hierarchical architecture of the RF/FSO WSN, where nodes are divided according to their functions [54]. Nodes 1, 2, and 3, which communicate directly with the base station, are known as gateway nodes. Nodes 2 and 6, which perform data routing, are known as cluster heads. To reduce the data communicated between nodes,

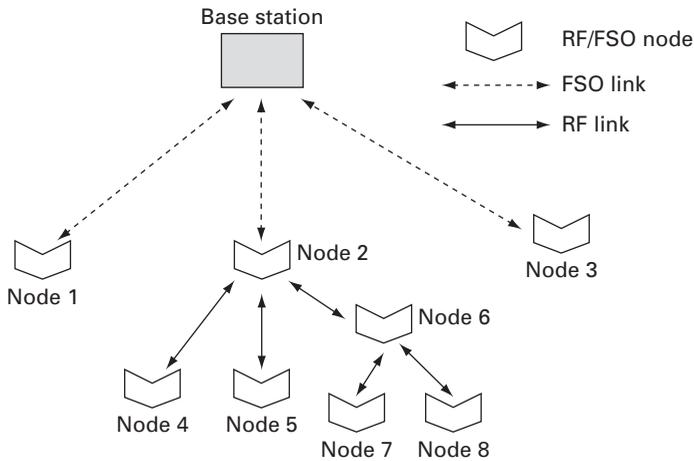


Figure 15.8 The RF/FSO wireless sensor network.

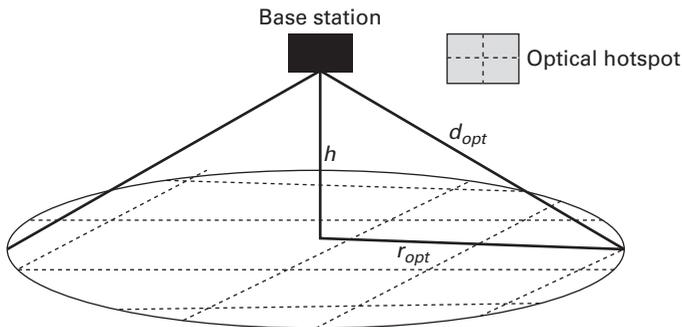


Figure 15.9 Base station's optical hotspot.

cluster heads may perform data aggregation using techniques like suppression, minima, maxima, and averaging [13]. All nodes perform sensing, but nodes 2 and 6 are also responsible for routing the data to and from nodes within their respective clusters.

The optical links are formed using the retro-reflecting technique described in Section 15.2. Nodes which are unable to communicate optically with the base station, due to lack of LOS to the base station, route their data traffic through their closest neighbor using RF multi-hop links. For the RF/FSO WSN in Figure 15.8, RF links serve as back-ups. For RF multi-hop links, a sensor measures the relative radio signal strength of its neighbors before connecting to the node with the strongest signal (which will usually be the closest node to it).

As FSO communications demands less energy than RF communications, the RF/FSO network works best when the nodes are within the base station's optical hotspot, as shown in Figure 15.9. In Figure 15.9, h is the base station's height from the sensor field. For the optical link, d_{opt} is the maximum acceptable communications range to ensure

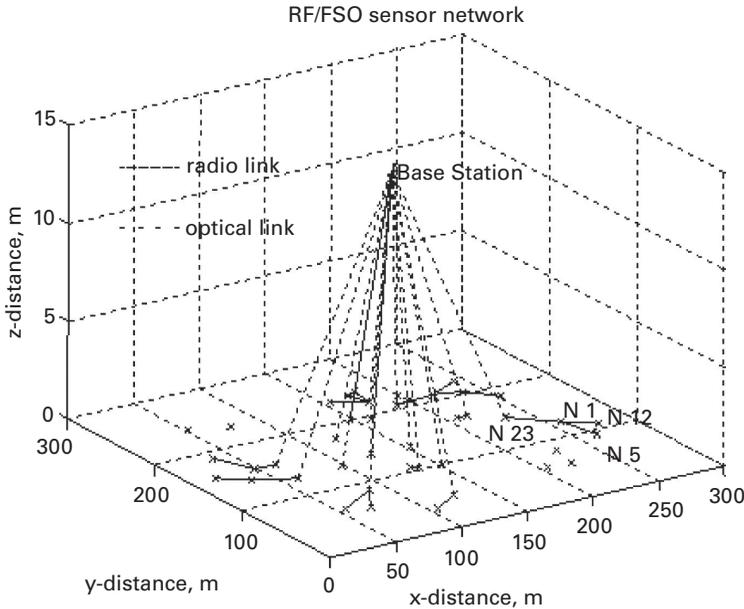


Figure 15.10 RF/FSO network configuration with nodes outside base station's optical hotspot [54].

that the minimum quality of service is met, and r_{opt} is the radius of the circular optical hotspot. Nodes can lie outside the optical hotspot, but they will need to communicate using high-energy RF multi-hop links, as shown in Figure 15.10 and Figure 15.11

15.3.3 Performance

The RF/FSO WSN's lifetime and coverage is compared against an RF-only sensor network [54]. The RF links are omnidirectional whereas the FSO links are directional. A bit rate of 100 bps is assumed for both networks. For the analysis in this section, the maximum RF and optical communication ranges are set to 150 m and 200 m respectively to ensure that quality of service requirements are met. Several parameters are introduced to compare the performance of the RF/FSO WSN against an RF-only WSN. The lifetime ratio, L_r , and coverage ratio, C_r , are defined as:

$$L_r = \left(\frac{\text{RF/FSO WSN lifetime}}{\text{RF - only WSN lifetime}} \right) \quad (15.1)$$

$$C_r = \left(\frac{\text{Average RF/FSO WSN coverage over RF/FSO WSN's lifetime}}{\text{Average RF only WSN coverage over RF only WSN's lifetime}} \right) \quad (15.2)$$

There are a number of definitions for network lifetime. For the work described here, network lifetime is defined as the time from when the network begins sensing and transmitting data, until time when network coverage falls to 0% of total deployed area [54]. In the RF/FSO WSN, the blocking factor, b_f , is the percentage of nodes without

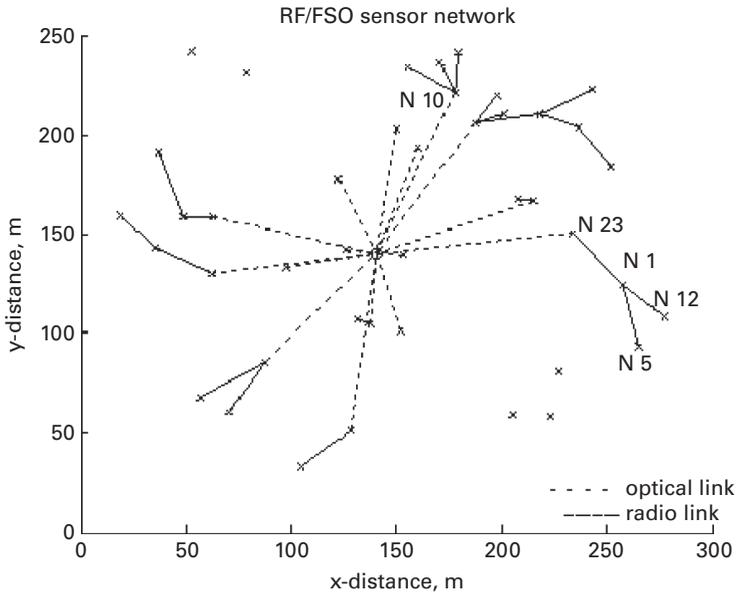


Figure 15.11 RF/FSO network configuration with nodes outside base station's optical hotspot (aerial view) [54].

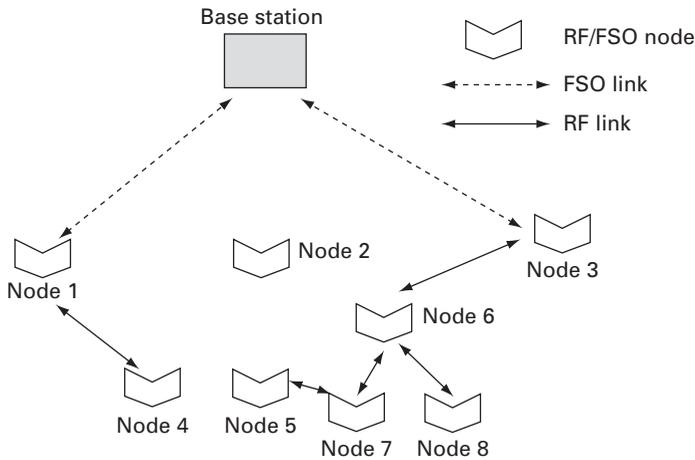


Figure 15.12 RF/FSO network reconfiguration after death of node 2.

LOS to the base station. The node density, μ , is the number of nodes within the radio range of a node and is given as [61]:

$$\mu = \frac{N \cdot d_{radio}^2}{R^2}, \tag{15.3}$$

where N is the number of deployed nodes, R is the radius of the deployment area and d_{radio} is the radio transmission range. Two types of networks are considered. reconfiguring and non-reconfiguring networks. Figure 15.12 shows the reconfigured

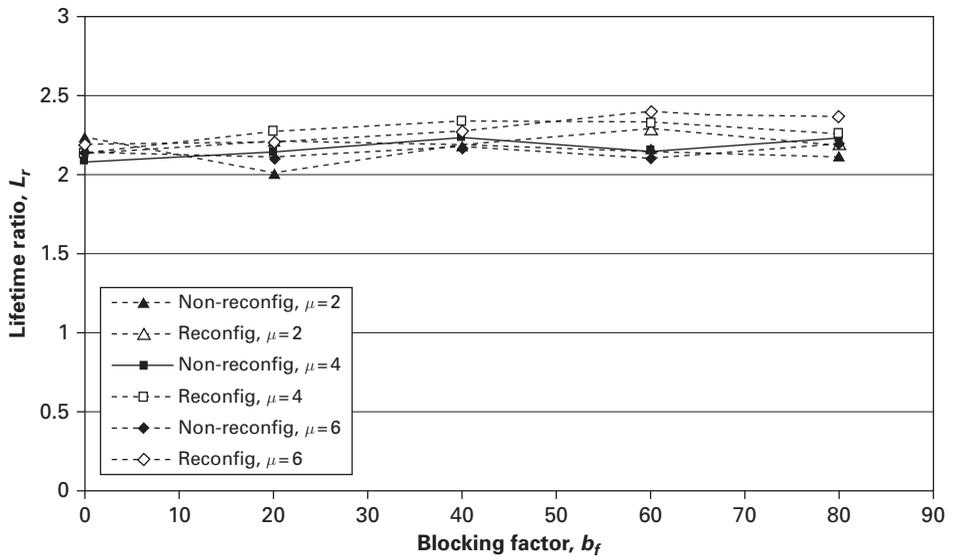


Figure 15.13 Lifetime ratio vs. blocking factor for a range of node densities [55].

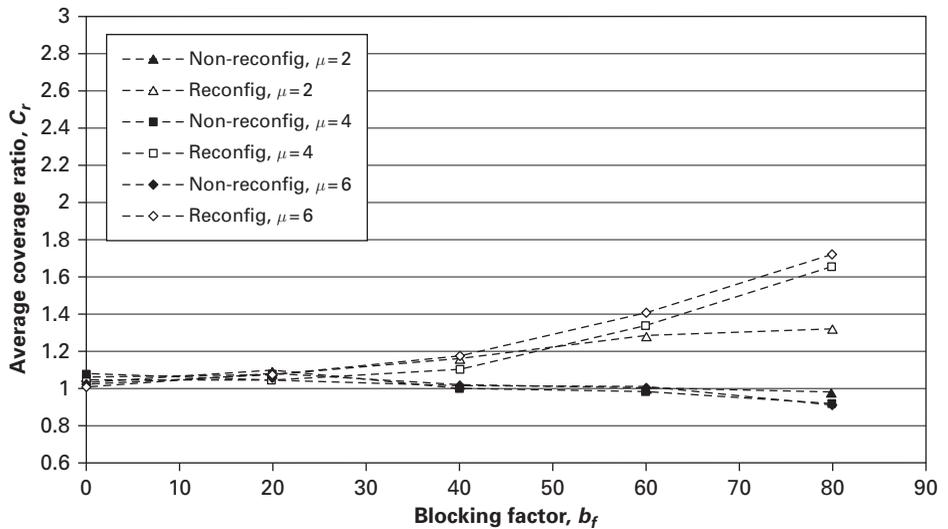


Figure 15.14 Average coverage ratio vs. blocking factor for a range of node densities [55].

RF/FSO network in Figure 15.8 after the death of node 2. In a non-reconfiguring network, all the nodes in node 2's cluster (nodes 4–8) would become isolated after it dies.

Figure 15.13, Figure 15.14, and Figure 15.15 show the performance of the RF/FSO WSN. Figure 15.13 shows that the RF/FSO WSN lasts at least twice as long as its RF-only counterpart for a range of blocking factors and node densities considered.

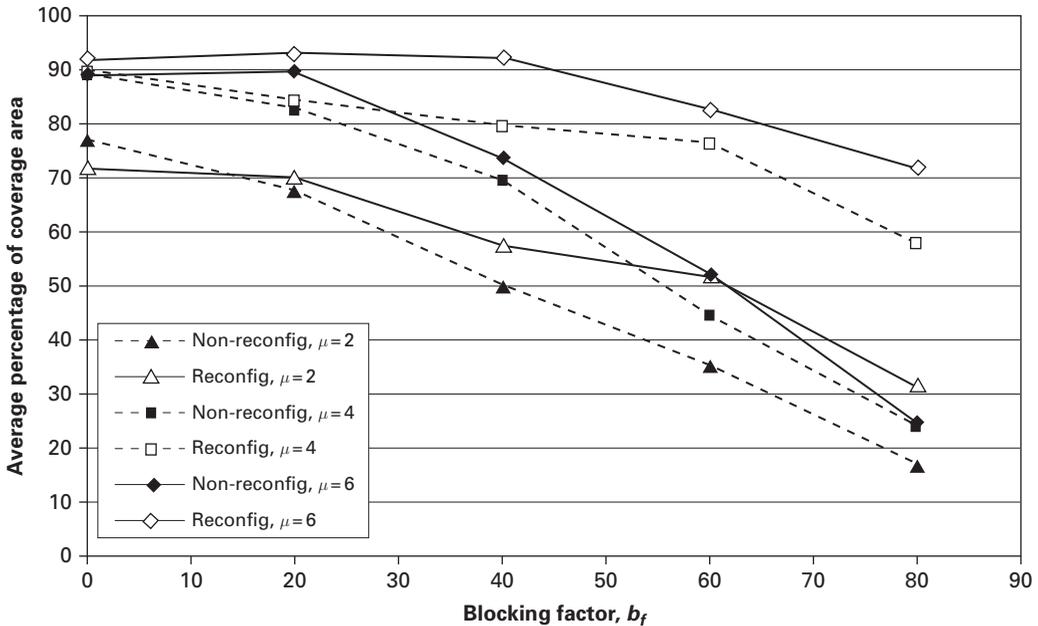


Figure 15.15 Average RF/FSO WSN coverage area vs. blocking factor for a range of node densities [55].

Figure 15.14 shows that the RF/FSO WSN provides almost the same coverage as the RF-only WSNs for non-reconfiguring networks. But, for high blocking factors, the reconfiguring RF/FSO WSNs are able to provide better coverage than their RF-only reconfiguring counterparts. Figure 15.15 shows the RF/FSO WSN coverage for a range of blocking factors and node densities. Reconfiguring RF/FSO networks perform better than non-reconfiguring RF/FSO networks for higher blocking factors.

15.4 Conclusions

The Internet of Things and the need to sense our environment at finer scales will lead to much wider deployment of small sensor and other types of network node, and connecting these in an energy efficient manner is a challenging problem.

In this chapter we have discussed some examples of how free space optics can improve the energy efficiency of sensor networks, and how a combination of RF and optical transmission can either improve reliability, data rate, or overall network performance. Over a range of scales there are situations where the complexity of the additional optical communications channel (over the normally implemented RF systems) can be justified in terms of increased system functionality, lifetime, or both. Currently, the major barrier to the wider deployment of these is seen as the availability of components that can be easily integrated with existing nodes, to demonstrate the benefits of this approach in real-world situations.

15.5 Acknowledgments

The authors would like to thank the current and past members of the optical wireless communications group at Oxford for their contributions to the results presented here.

References

- [1] Akyildiz, I. F., T. Melodia, and K. R. Chowdury, Wireless multimedia sensor networks: A survey. *IEEE Wireless Communications*, 2007. **14**(6): pp. 32–39.
- [2] Yick, J., B. Mukherjee, and D. Ghosal, Wireless sensor network survey. *Computer Networks*, 2008. **52**(12): pp. 2292–2330.
- [3] Crossbow. Crossbow Motes: www.crossbow.com. Accessed 2010.
- [4] Meyer, J. *et al.*, Wireless sensor networks for long-term structural health monitoring. *Smart Structures and Systems*, 2010. **6**(3): pp. 263–275.
- [5] Miyashita, T. and M. Nagai, Vibration-based structural health monitoring for bridges using laser doppler vibrometers and MEMS-based technologies. *International Journal of Steel Structures*, 2008. **8**(4): pp. 325–331.
- [6] Kher, S., J. R. Chen, and A. K. Somani, IEEE 1451 standard and wireless sensor networks: An overview of fault tolerant algorithms. *2006 IEEE International Conference on Electro/Information Technology*, 2006: pp. 227–232
- [7] Pekoslawski, B. *et al.*, Study on application of piezoelectric vibration energy harvesters for powering of wireless sensor nodes in large rotary machine diagnostic systems. *Nanotech Conference & Expo 2009, Vol 1, Technical Proceedings*, 2009: pp. 530–533.
- [8] Toftegaard, T. S., Next generation wireless infrastructure. *Wireless Personal Communications*, 2010. **53**(3): pp. 465–479.
- [9] Toledo, O. M., D. Oliveira, and A. S. A. C. Diniz, Distributed photovoltaic generation and energy storage systems: A review. *Renewable & Sustainable Energy Reviews*, 2010. **14**(1): pp. 506–511.
- [10] Pantelopoulos, A. and N. G. Bourbakis, A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Transactions on Systems Man and Cybernetics Part C—Applications and Reviews*, 2010. **40**(1): pp. 1–12.
- [11] Knight, C., J. Davidson, and S. Behrens, Energy options for wireless sensor nodes. *Sensors*, 2008. **8**(12): pp. 8037–8066.
- [12] Behrens, S. and J. Davidson, Energy harvesting for sensor networks. *17th IEEE International Symposium on the Applications of Ferroelectrics*, 2008: pp. 137–138.
- [13] Al-Karaki, J. N. and A. E. Kamal, Routing techniques in wireless sensor networks: A survey. *IEEE Wireless Communications*, 2004. **11**(6): pp. 6–28.
- [14] Leitgeb, E. *et al.*, High reliable optical wireless links for the last mile access. *Icton 2008: Proceedings of 2008 10th Anniversary International Conference on Transparent Optical Networks*, vol 4, 2008: pp. 178–183.
- [15] Gadwal, V. and S. Hammel, Free space optical communication links in a marine environment. *Free-Space Laser Communications VI*, 2006. **6304**: pp. U166–U176.
- [16] Northcott, M. J. *et al.*, Long distance laser communications demonstration – art. no. 65780S. *Defense Transformation and Net-Centric Systems 2007*, 2007. **6578**: pp. S5780–S5780.
- [17] Toyoshima, M. *et al.*, Ground-to-satellite laser communication experiments. *IEEE Aerospace and Electronic Systems Magazine*, 2008. **23**(8): pp. 10–18.

- [18] Harris, A. and T. A. Gioma, Minimization of acquisition time in a wavelength diversified FSO link between mobile platforms – art. no. 655108. *Atmospheric Propagation IV*, 2007. **6551**: pp. 55108–55108.
- [19] Agrawal, N., S. D. Milner, and C. C. Davis. Free space optical sensor network for short-range applications. *Free-Space Laser Communications X*. 2010. San Diego, California, USA: SPIE.
- [20] Ghosh, A. K. *et al.* Free-space optics based sensor network design using angle-diversity photodiode arrays. *Free-Space Laser Communications X*. 2010. San Diego, California, USA: SPIE.
- [21] Kedar, D. and S. Arnon, Second-generation laser firefly clusters: improved scheme for distributed sensing in the atmosphere. *Applied Optics*, 2005. **44**(6): pp. 984–992.
- [22] Kedar, D. and S. Arnon, Laser “firefly” clustering: a new concept in atmospheric probing. *IEEE Photonics Technology Letters*, 2003. **15**(11): pp. 1672–1674.
- [23] Reardon, C. *et al.*, Optical beam-steering for wireless sensor networks. *2009 IEEE Leos Annual Meeting Conference Proceedings*, vols. 1 and 2, 2009: pp. 583–584.
- [24] Reardon, C. *et al.*, Integrated polymer microprisms for free space optical beam deflecting. *Optics Express*, 2009. **17**(5): pp. 3424–3428.
- [25] Mathews, J., M. Barnes, and D. K. Arvind, Low power free space optical communication in wireless sensor networks. *Proceedings of the 2009 12th Euromicro Conference on Digital System Design, Architectures, Methods and Tools*, 2009: pp. 849–856
- [26] Srinivasan, B. and V. Rodoplu, Energy-efficient joint source-channel coding for optical wireless underwater networks. *2007 Oceans*, vols 1–5, 2007: pp. 725–729
- [27] Liu, L. B., S. L. Zhou, and J. H. Cui, Prospects and problems of wireless communication for underwater sensor networks. *Wireless Communications & Mobile Computing*, 2008. **8**(8): pp. 977–994.
- [28] Anguita, D., D. Brizzolara, and G. Parodi, *Building an underwater wireless sensor network based on optical communication: Research challenges and current results*. *2009 3rd International Conference on Sensor Technologies and Applications (Sensorcomm 2009)*, 2009: pp. 476–479
- [29] Kedar, D. and S. Arnon, Subsea ultraviolet solar-blind broadband free-space optics communication. *Optical Engineering*, 2009. **48**(4).
- [30] Arnon, S., Underwater optical wireless communication network. *Optical Engineering*, 2010. **49**(1).
- [31] Llorca, J. *et al.*, Reconfigurable optical wireless sensor networks. *Optics in Atmospheric Propagation and Adaptive Systems VI*, 2004. **5237**: pp. 136–146.
- [32] Kedar, D. and S. Arnon, An ultraviolet optical wireless sensor network in multi-scattering channels. *Information Optics*, 2006. **860**: pp. 436–445
- [33] Rabinovich, W. S., *et al.*, InGaAs multiple quantum well modulating retro-reflector for free space optical communications. *Free-Space Laser Communication and Laser Imaging*, 2002. **4489**: pp. 190–201
- [34] Kahn, J. M., R. H. Katz, and K. S. Pister. Next century challenges: mobile networking for “Smart Dust”. 1999: Dept. of Electr. Eng. & Comput. Sci., California University Berkeley CA USA.
- [35] Warneke, B., B. Atwood, and K. S. J. Pister. Smart dust mote forerunners. 2001: Dept. of Electr. Eng. & Comput. Sci. California University Berkeley CA USA.
- [36] Lixia, Z., K. S. J. Pister, and J. M. Kahn. Assembled corner-cube retroreflector quadruplet. Technical Digest. 2002. Robotics & Autom: Dept. of Electr. Eng. & Comput. Sci. California University Berkeley CA USA.

- [37] Teramoto, S. and T. Ohtsuki, Optical wireless sensor network system using corner cube retroreflectors (CCRs). *Globecom '04: IEEE Global Telecommunications Conference*, vols 1–6, 2004: pp. 1035–1039.
- [38] Moon, G. B., C. G. Park, and J. G. Lee, Optical Communication Method using CCR. *2008 International Conference on Control, Automation and Systems*, vols 1–4, 2008: pp. 642–646
- [39] Arnon, S., *Swarm of sensor acquisition and identification using optical wireless communication. Free Space Laser Communications IV*, 2004. **5550**: pp. 436–442
- [40] Arnon, S., Network of sensors: acquisition probability. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 2007. **24**(9): pp. 2758–2765.
- [41] O'Brien, D. C. *et al.*, Design and implementation of optical wireless communications with optically powered smart dust motes. *IEEE Journal on Selected Areas in Communications*, 2009. **27**(9): pp. 1646–1653.
- [42] O'Brien, D. C. *et al.* Energy scavenging sensors for ultra-low power sensor networks. *Free-Space Laser Communications X*. 2010. San Diego, California, USA: SPIE.
- [43] Vachiramon P., G. E. Faulkner, and D. C. O'Brien, A DC balancing algorithm for FLCOS binary phase holograms. *Optics Letters*, 2007. **32**(22): pp. 3275–3277.
- [44] Cap, G. A., H. H. Refai, and J. J. Sluss. FSO tracking and auto-alignment transceiver system. *Unmanned/Unattended Sensors and Sensor Networks V*. 2008. Cardiff, Wales, United Kingdom: SPIE.
- [45] Yun, G. and M. Kavehrad. Spot-diffusing and fly-eye receivers for indoor infrared wireless communications. *Conference Proceedings, 1992 IEEE International Conference on Selected Topics in Wireless Communications*. 1992.
- [46] Delbrück, T. and C. A. Mead, *Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor circuits*, in *Vision Chips: Implementing Vision Algorithms with Analog VLSI Circuits*, C. Koch and H. Li, editors. 1995, IEEE Computer Society Press. pp. 139–161.
- [47] IEC 60825-1. Safety of laser products part 1. 2007: British Standards Institution.
- [48] O'Brien, D. C. *et al.* Robust data transmission at high data rates using liquid crystal spatial light modulators. *Spatial Light Modulators and Integrated Optoelectronic Arrays*. 1999: Optical Society of America.
- [49] Kim, I. I. and E. Korevaar, Availability of free space optics (FSO) and hybrid FSO/RF systems. *Optical Wireless Communications IV*, 2001. **4530**: pp. 84–95.
- [50] Desai, A. and S. Milner, Autonomous reconfiguration in free-space optical sensor networks. *IEEE Journal on Selected Areas in Communications*, 2005. **23**(8): pp. 1556–1563.
- [51] Gibson, S. M. *et al.*, Coordinate measurement in 2-D and 3-D geometries using frequency scanning interferometry. *Optics and Lasers in Engineering*, 2005. **43**(7): pp. 815–831.
- [52] Llorca, J., S. D. Milner, and C. C. Davis. Molecular inspired models for prediction and control of directional FSO/RF wireless networks. *Free-Space Laser Communications X*. 2010. San Diego, California, USA: SPIE.
- [53] Milner, S. D., J. Llorca, and C. C. Davis, *Autonomous reconfiguration and control in directional mobile ad hoc networks*. *IEEE Circuits and Systems Magazine*, 2009. **9**(2): pp. 10–26.
- [54] Sivathasan, S., Hybrid radio frequency/free space optical communications for energy-efficient wireless sensor networks, Doctor of Philosophy (DPhil) Thesis, 2008, University of Oxford.

-
- [55] Sivathasan, S. and D. O'Brien, Radio frequency/free space optical and radio frequency-only wireless sensor networks: a comparative study of performance. *IETE Journal of Research*, 2010. **56**(1): pp. 52–61.
- [56] Deng, S.L. *et al.*, Wireless connections of sensor network using RF and free space optical links – art. no. 677307. *Next-Generation Communication and Sensor Networks 2007*, 2007. **6773**: pp. 77307–77307.
- [57] Chonggang, W. *et al.* *ISNOW: Integrated sensor networks with optical and wireless communications*. *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference (OFC/NFOEC) 2010*.
- [58] Nadeem, F. *et al.*, Comparing the life time of terrestrial wireless sensor networks by employing hybrid FSO/RF and only RF access networks. *ICWMC: 2009 Fifth International Conference on Wireless and Mobile Communications*, 2009: pp. 134–139.
- [59] Heinzelman, W. R., A. Chandrakasan, and H. Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference*, 2000.
- [60] Lindsey, S. and C.S. Raghavendra, PEGASIS: Power-efficient GAthering in sensor information systems. *2002 IEEE Aerospace Conference Proceedings*, vols 1–7, 2002. pp. 1125–1130
- [61] Akyildiz, I. F. *et al.*, Wireless sensor networks: a survey. *Computer Networks*, 2002. **38**(4): pp. 393–422.

Index

Figures and tables are indicated in bold typeface.

- absorption, 203, **203**, 204–6, **210**
- adaptive power control, 68–69. *see also* optical orthogonal codes
- albedo, **222**, 226–28, **226**, **227**
- AMC. *see* hybrid RF/FSO wireless networks
- amplifier
 - double heterodyne photoparametric, 259–60, **260**
 - photoparametric, 258–59, **258**
 - trans-impedance, 19, 255–57, **256**, **257**, **258**, **262**
- angle diversity receiver, 118
- AOP. *see* apparent optical properties
- APC. *see* adaptive power control
- applications
 - optical code-division multiple-access wireless network, 81–83, **82**, **83**
 - optical modulating retro-reflectors, 341–47, **344**, **345**, **346**, 353, 361–67, **362**, **363**, **364**, **365**, **366**
- automobile driving application, 365–66, **366**

- barcodes, 129, **130**
- BB84 protocol, 308–10, **309**
- Beer's Law, 203–4, **203**, 212, 214. *see also* channel loss
- BER analysis
 - channel shortening equalizer, **111**, **112**
 - free-space optical wireless networks, 70–71
 - free-space orthogonal frequency division multiplexing wireless networks, **43**
 - in FSO links, **74**
 - generalized hybrid
 - subcarrier/amplitude/phase/polarization modulation, 37–38, **38**
 - heterogeneous optical networks, **47**, **48**
 - infrared optical wireless networks, 251–53, **256**
 - multiple-input multiple-output, 23–25, **24**, **25**, **26**
 - optical code-division multiple-access wireless network, 62–64, **67**, 70–71
 - pulse amplitude modulation, 22–23
 - time equalizer, **109**
- biased-OFDM scheme, 40
- B-OFDM. *see* biased-OFDM scheme

- cat's eye retro-reflector, 333–34, **333**, **334**
- channel inversion, 30–31, **30**
- channel loss, **212**. *see also* Beer's Law
- channel modeling, 12–13
- channel shortening equalizer, 90, 106–12, **111**, **112**
- channel state information, 278–79. *see also* transmitter
- CIDS. *see* cyclic-invariant difference set
- clipped OFDM scheme, 41
- clouds
 - and free-space optical wireless networks, 90–113, **93**, **95**, **97**, **98**
 - RF frequency in, 273
- codes on graphs, 13–19. *see also* modulation
- C-OFDM. *see* clipped OFDM scheme
- communication
 - line-of-sight, 11, 231–32
 - non-line-of-sight, 177–99, **232**, 233, 241–42
- conventional cryptography, 306–8
- corner cube retro-reflectors, 332, **337**, **341**
- cryptography
 - conventional, 306–8
 - modern quantum, 305–6, 312–25
- CSE. *see* channel shortening equalizer
- CSI. *see* channel state information
- cyclic-invariant difference set, 16

- decision-feedback equalization, 103
- decoy-state BB84 protocol, 310
- deep fades, 38–43
- DFE. *see* decision-feedback equalization
- diffuse link, 241–42
- diffuse system design, 119–23, **121**
- discrete-time wireless networks
 - free-space optical intensity, 148, 156–61, **159**, **160**, **161**
 - optical intensity with Gaussian noise, 149, 161–63
 - performance, 149–63, 163–72
 - Poisson, 146–48, 151–55, **155**

- double heterodyne photoparametric amplifier, 259–60, **260**
- eigen analysis, 102–3
- encryption. *see* cryptography
- entanglement-based BB84 protocol, 310–11, **311**
- entanglement-based QKD links, 320–21
- error, pointing, 87–88
- eye safety, 178, 243, **243**
- fountain code, 26
- free-space optical wireless networks
 - atmospheric effects in, 273
 - cryptography, 305–25
 - modulation, 11–49, **39**
 - OCDMA system, 68–75, **69**
 - and optical modulating retro-reflectors, 342
 - system design, **104, 354**
 - turbulence, 11–49, 90–113
- free-space orthogonal frequency division
 - multiplexing wireless networks, 38–43, 39, **43**
- FSO. *see* free-space optical wireless networks
- Gaussian noise, 149, 161–63, 279–80, **282**. *see also*
 - interference
- generalized hybrid
 - subcarrier/amplitude/phase/polarization modulation, 35–38, **36, 37, 38**
- GH-SAPP. *see* generalized hybrid
 - subcarrier/amplitude/phase/polarization modulation
- Gray mapping rule, 21, 24, 25
- HAPP. *see* generalized hybrid
 - subcarrier/amplitude/phase/polarization modulation
- heterogeneous optical networks, 12, 43–48, **44, 45, 47, 48**
- high altitude platforms, 322–25, **324**
- HONs. *see* heterogeneous optical networks
- hybrid RF/FSO sensor network, 378–83, **379, 380, 381, 382, 383**
- hybrid RF/FSO wireless networks
 - modulation, 12, 29–34, **30**
 - performance, 290–95, **291**
 - system design, 275–80, **275, 277**
 - theory of, 281–90, **282, 283, 287, 290**
- imaging receiver, 118–19
- IM/DD. *see* intensity modulation with direct detection
- indoor optical wireless networks
 - fundamentals, 117
 - modulation, 123, 138
 - receiver, 118–19
 - system design, 119–38, **135, 136**
 - transmitter, 117–18, 132–38, 146–48
- infrared optical wireless networks
 - eye safety, 243, **243**
 - interference, 248–51, **249**
 - modulation, 263–66, **264, 265**
 - optical sources in, 242, **242, 267–68, 267**
 - performance, 251–53, **252, 253**
 - receiver, 255–63, **256, 257, 258, 260, 261, 262**
 - system design, 241–42, 244–47
- intensity modulation with direct detection, 11, 91, 247
- interference. *see also* Gaussian noise; scattering; turbulence
 - and free-space optical wireless networks, 11–49
 - infrared optical wireless networks, 248–51, **249**
 - optical code-division multiple-access wireless network, 73–75, **74**
 - underwater free-space optical wireless networks, 202–3, **203, 205, 207, 209, 210, 212**
- laser diode, 242, **242, 243**
- LD. *see* laser diode
- LDPC codes, 14–19, **14, 24, 25, 26, 35**
- LED. *see* light emitting diode
- light emitting diode, 242, **242, 243, 352–58, 354**
- linear minimum mean-squared error equalization, 103
- line-of-sight communication, 11, 231–32
- LOS. *see* line-of-sight communication
- LT code. *see* Luby-transform code
- Luby-transform code, 26–29
- Markov chain model, 90–91, 95–103
- Maximum-likelihood sequential detection
 - equalization, 103
- MCRT. *see* Monte Carlo ray tracing model
- MEMS. *see* micro-electrical mechanical modulator
- micro-electrical mechanical modulator, 336, **337**
- Mie theory, 92–103, **93**
- MIMO. *see* multiple-input multiple-output
- minimum-mean-squared error, **106**
- min-sum-with-correction-term algorithm, 17–19, **17**
- MLMD. *see* multiple lasers - multiple detectors
- MLSD. *see* Maximum-likelihood sequential detection equalization
- modern quantum cryptography
 - and conventional cryptography, 306–8
 - limits, 305–6
 - networking, 319–25
 - system design, 312–19, **314, 317, 318**
- modulation. *see also* codes on graphs;
 - multiple-input multiple-output
 - hybrid RF/FSO wireless networks, 12, 29–34, **30**
 - indoor optical wireless networks, 123, 138
 - infrared optical wireless networks, 263–66
 - intensity, 11, 91, 247

- multidimensional coded, 12, 35–38
 - on-off keying, 75, 77
 - optical modulating retro-reflectors, 334–41, **338**, **339**, 359–61, **360**
 - optical orthogonal codes, 75–82
 - pulse amplitude, 21–25, **21**, **24**, **25**, **76**, 265–66, **266**
 - pulse position, 25–26, 75–78, 230–31, 263–65, **264**
 - underwater free-space optical wireless networks, 230–31
- moment technique model, 99–103, **100**
- Monte Carlo ray tracing model
 - integration, **29**
 - and Markov chain, 90–91, 95–103, **95**
 - performance, 99–103, **100**
 - ultraviolet optical wireless networks, 183–87, **184**, 192–94, **194**
 - underwater, 214–16, **215**
- MQW. *see* surface normal multiple quantum well electroabsorption modulators
- MRR. *see* optical modulating retro-reflectors; retro-reflector communication
- multidimensional coded modulation, 12, 35–38
- multiple lasers - multiple detectors, 19, **26**
- multiple-input multiple-output, 11, 19–26, **20**, 116–38, **128**. *see also* modulation
- multi-scatterings, **97**, **98**
- networking
 - modern quantum cryptography, 319–25
 - underwater free-space optical wireless networks, 234–36
- NLOS. *see* non-line-of-sight communication
- non-line-of-sight communication, 177–99, **232**, 233, 241–42. *see also* optical modulating retro-reflectors
- OCDMA. *see* optical code-division multiple-access wireless network
- OFDM. *see* free-space orthogonal frequency division multiplexing wireless networks
- on-off keying modulation, 75, 77, **264**, **266**, **282**, **283**
- OOC. *see* optical orthogonal codes
- OOK. *see* on-off keying modulation
- OPPM. *see* overlapping pulse position modulation
- optical code-division multiple-access wireless network
 - applications, 81–83, **82**, **83**
 - BER analysis, 62–64, 70–71, **78**
 - interference, 73–75, **74**
 - modulation, 75–77, **76**
 - performance, 77–81
 - receiver, 57–59, 60–62, **61**, 64–66, **64**, **66**, **67**, 71–73
 - system design, 55–57, **56**, **58**, 59–60, **59**, 68–70, **69**
 - transmitter, 60, 66–68
- optical modulating retro-reflectors. *see also* non-line-of-sight communication
 - applications, 341–47, **344**, **345**, **346**, 353, 361–67, **362**, **363**, **364**, **365**, **366**
 - modulation, 334–41, **338**, **339**, 359–61, **360**
 - system design, **232**, 233–34, 329–34, **329**, **333**, **334**, 370
- optical orthogonal codes, 56–57, **66**, 75–82, 79–81, **81**. *see also* adaptive power control
- optical wireless networks
 - optical sources in, 242, **242**
 - optics in, 267–68
- optical wireless sensor networks
 - modulation, 374
 - performance, 380–83, **381**, **382**, **383**
 - receiver, 374–75
 - system design, 371–74, **371**, **372**, **373**, **374**, 375–78, **376**, **377**, **379**, **380**
 - transmitter, 375
- orthogonal frequency division multiplexing, 12
- outage exponent, 285–87, **287**, **290**, **296**
- outage probability, 285
- overlapping pulse position modulation, 76–77, **76**
- PAM. *see* pulse amplitude modulation
- passive optical networks, 43
- performance
 - discrete-time wireless networks, 149–63, 163–72
 - hybrid RF/FSO wireless networks, 290–95, **291**
 - infrared optical wireless networks, 251–53, **252**, **253**
- Monte Carlo ray tracing model, 99–103, **100**
- optical code-division multiple-access wireless network, 77–81
- optical wireless sensor networks, 380–83, **381**, **382**, **383**
- ultraviolet optical wireless networks, 192–99, **192**
- Phase Function. *see* Mie theory
- photodetection model, 279–80
- photoparametric amplifier, 258–59, **258**
- phytoplankton, 206
- point-to-point system design, 131–36, **135**, **136**, 370
- Poisson channel, 146–48, 151–55, **155**, 279–80, **282**, **283**
- PONS. *see* passive optical networks
- PPM. *see* pulse position modulation
- protocol
 - modern quantum cryptography, 308–11
 - QKD transport, 321–22
- prototypes. *see* applications
- public-key cryptography. *see* conventional cryptography

- pulse amplitude modulation, 21–25, **21**, **24**, **25**, **76**, 265–66, **266**. *see also* pulse intensity modulation
- pulse intensity modulation, 146–71. *see also* pulse amplitude modulation
- pulse position modulation, 25–26, 75–78, 230–31, 263–65, **264**
- punctured code, 26
- QAM. *see* Quadrature Amplitude-Modulation
- QC LCPC. *see* quasi-cyclic binary LDPC codes
- QKD. *see* quantum key distribution
- Quadrature Amplitude-Modulation, 22
- quantum key distribution, 305–25
- Radiative Transfer Equation, 214
- radio frequency technology
- bandwidth, 103
 - and fiber-optics, 43
 - incompatibilities, 11
 - and infrared optical wireless networks, 240
 - and optical modulating retro-reflectors, 341–42
 - sensors in, 369
 - underwater, 201
- Raptor code, 12, 26–29, **27**, **29**
- rate-less code, 26
- receiver
- indoor optical wireless networks, 118–19
 - infrared optical wireless networks, 255–63, **256**, **257**, **258**, 260–63, **260**, **261**, **262**
 - optical code-division multiple-access wireless network, 57–59, 60–62, **61**, 64–66, **64**, **66**, **67**, 71–73
 - with polarization analyzer module for cryptography, 316–17, **317**
 - and Viterbi equalizer, 107–12, **108**
- repeaters, 319–20
- retro-reflectors. *see* optical modulating retro-reflectors
- RF. *see* radio frequency technology
- RoF. *see* radio frequency technology
- RTE. *see* Radiative Transfer Equation
- SAA. *see* small-angle approximation
- satellites, 322–25
- scattering. *see also* interference
- albedo, **210**
 - atmospheric, 179
 - in cloud channel, 90–113
- ULTRAVIOLET OPTICAL WIRELESS NETWORKS, 179–89, **180**, **181**, **184**, **188**
- underwater, 203, **203**, 206–13, **207**, **209**, **210**
- scintillation, 12–13, 273, 276–77, 346
- SDMT. *see* spatial discrete multitone modulation transmitter
- Second Largest Eigenvalue Modulus, 102–3, **102**
- secret-key cryptography. *see* conventional cryptography
- security guidance systems, 362
- signal-to-noise
- infrared optical wireless networks, **249**, **252**
 - underwater free-space optical wireless networks, 201–2
- SLEM. *see* Second Largest Eigenvalue Modulus
- small-angle approximation, 216–20, **220**
- smart dust, 371, **371**
- SNR. *see* signal-to-noise
- spatial discrete multitone modulation transmitter, 132–38, **133**, **135**, **137**
- spatial dispersion. *see* water dispersion
- spot-diffusing system design, 123–27, **125**, **126**, 127–38
- surface normal multiple quantum well electroabsorption modulators, 336–41, **338**, **339**, **341**
- system design
- diffuse, 119–23, **121**
 - free-space optical, **104**, **354**
 - hybrid RF/FSO wireless networks, 275–80, **275**, **277**
 - indoor optical wireless networks, 119–38, **135**, **136**
 - infrared optical wireless networks, 241–42
 - modern quantum cryptography, 312–19, **314**, **317**, **318**
 - optical code-division multiple-access wireless network, 55–57, **56**, **58**, 59–60, **59**, 68–70, **69**
 - optical modulating retro-reflectors, 329–34, **329**, **333**, **334**
 - optical wireless sensor networks, 371–74, **371**, **372**, **373**, 375–78, **375**, **376**, **377**, **379**, **380**
 - point-to-point, 131–36, **135**, **136**, 370
 - spot-diffusing, 123–27, **125**, **126**, 127–38
 - ultraviolet optical wireless networks, 189–92, **190**, **191**
 - underwater free-space optical wireless networks, 228–30, **229**, 231–34, **232**
 - visible-light wireless networks, 351–54, **352**, **354**
- TA. *see* trans-impedance amplifier
- Tanner graph, **14**, **27**
- temporal dispersion. *see* water dispersion
- TEQ. *see* time equalizer
- time equalizer, 90, 103–4, **106**, **109**
- TPC. *see* turbo-product codes
- trans-impedance amplifier, 19, 255–57, **256**, **257**, **258**, **262**
- transmitter. *see also* channel state information
- and channel shortening equalizer, 90, 106–12
 - indoor optical wireless networks, 117–18, 132–38, 146–48

- optical code-division multiple-access wireless network, 60, 66–68
- optical wireless sensor networks, 375
- pointing errors, 87–88
- spatial discrete multitone modulation, 132–38, **133, 135, 137**
- and time equalizer, 90, 103–4, **106, 109**
- underwater free-space optical wireless networks, 228–29
- truncated channel inversion, **30, 31**
- turbo product codes, 13
- turbulence. *see also* interference
 - and free-space optical wireless networks, 90–113
 - in FSO wireless networks, 11–49
 - underwater, 210
- uFSO. *see* underwater free-space optical wireless networks
- ultra wideband communications, 43
- ultraviolet optical wireless networks
 - eye safety, 178
 - history of, 177–78
 - interference, 179–89
 - performance, 192–99, **192**
 - system design, 189–92, **190, 191**
- unclipped OFDM scheme, 41
- underwater free-space optical wireless networks
 - history of, 201–2
 - interference, 202–3, **203, 205, 207, 209, 210, 212**
 - modulation, 230–31
 - networking, 234–36
 - optical modulating retro-reflectors, 347
 - system design, 228–30, **229, 232**
 - theory of, 213–28, **215, 219, 220, 222, 225, 227**
- U-OFDM. *see* unclipped OFDM scheme
- UV. *see* ultraviolet optical wireless networks
- UWB. *see* ultra wideband communications
- variable rate variable power adaptation, 31–34, **33, 34, 43–47, 47**
- variable spreading factor, 224–26, **225**
- vector radiative transfer modeling, 217. *see also* water dispersion
- Visible Light Communications Consortium, 358–59
- visible-light wireless networks
 - applications, 361–67, **362, 363, 364, 365, 366**
 - modulation, 359–61, **360**
 - standards, 354–59
 - system design, 351–54, **352, 354**
- Viterbi equalizer, 107–12, **108**
- VLC. *see* visible-light wireless networks
- VSF. *see* variable spreading factor
- water dispersion, 218–28, **219, 220, 222, 225, 226, 227**. *see also* vector radiative transfer modeling
- water optical properties, 211–13