

Optical Wireless Communication Theory and Technology

Xizheng Ke
Ke Dong

Optical Wireless Communication



Science Press
Beijing



Springer

Optical Wireless Communication Theory and Technology

Series Editor

Xizheng Ke, School of Automation and Infor. Eng., Xi'an University
of Technology, Xi'an, Shaanxi, China

The book series Optical Wireless Communication Theory and Technology aims to introduce the key technologies and applications adopted in optical wireless communication to researchers of communication engineering, optical engineering and other related majors. The individual book volumes in the series are thematic. The goal of each volume is to give readers a comprehensive overview of how the theory and technology in a certain optical wireless communication area can be known. As a collection, the series provides valuable resources to a wide audience in academia, the communication engineering research community and anyone else who are looking to expand their knowledge of optical communication.

More information about this series at <https://link.springer.com/bookseries/16890>

Xizheng Ke · Ke Dong

Optical Wireless Communication

 Science Press
Beijing

 Springer

Xizheng Ke
School of Automation and Information
Engineering
Xi'an University of Technology
Xi'an, Shaanxi, China

Ke Dong
School of Automation and Information
Engineering
Xi'an University of Technology
Xi'an, Shaanxi, China

ISSN 2731-5967 ISSN 2731-5975 (electronic)
Optical Wireless Communication Theory and Technology
ISBN 978-981-19-0381-6 ISBN 978-981-19-0382-3 (eBook)
<https://doi.org/10.1007/978-981-19-0382-3>

Jointly published with Science Press, Beijing, China
The print edition is not for sale in China (Mainland). Customers from China (Mainland) please order the print book from: Science Press.

© Science Press 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publishers, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publishers nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publishers remain neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Contents

1	Optical Wireless Communication System	1
1.1	System Model of Optical Wireless Communication	1
1.1.1	Transmitter	1
1.1.2	Receiver	2
1.1.3	Channel	3
1.2	Laser Light Source	4
1.2.1	Principles of a Laser Diode	4
1.2.2	Characteristics of a Laser Diode	6
1.2.3	Nonlinearity Correction	6
1.3	Device Response Characteristics	9
1.3.1	Response Characteristics of a Semiconductor Laser	9
1.3.2	Response Characteristics of a PIN Photodetector	14
1.4	Surface Plasmon Polarization	19
1.4.1	Effect of Different Incident Light Directions on the Light Absorption Performance of Silicon Substrates	20
1.4.2	Electric Field Modulus Distribution on the x - z Cross Section of the Photodetector	21
1.5	Signal Detection	22
1.5.1	Direct Detection	23
1.5.2	Direct Detection Limit	24
1.6	Optical Amplifier	25
1.6.1	Classification of Optical Amplifiers	25
1.6.2	Erbium-Doped Fiber Amplifier	26
1.6.3	Semiconductor Optical Amplifier	27
1.7	Spatial Light to Fiber Coupling Technology	28
1.7.1	Single Lens Coupling	28
1.7.2	Array Coupling	30
1.7.3	Special Fiber Coupling	30

1.8	Optical Antenna and Telescope	31
1.8.1	Refractor Telescope	31
1.8.2	Reflecting Telescope	32
1.8.3	Catadioptric Telescope	33
1.8.4	Integrated Transceiver Optical Antenna	34
1.9	Summary and Prospects	35
1.10	Questions	35
1.11	Exercises	36
	References	42
2	Coherent Optical Communication	45
2.1	Basic Principles of Coherent Optical Communication	45
2.1.1	Fundamentals	45
2.1.2	Homodyne Detection	48
2.1.3	Heterodyne Detection	49
2.1.4	Detection of an Amplitude Modulated Signal	49
2.1.5	Dual-Channel Balanced Detection	51
2.2	Coherent Modulation and Demodulation	52
2.2.1	Optical Modulation	52
2.2.2	Coherent Demodulation	53
2.2.3	System Performance	58
2.3	Factors Affecting Detection Sensitivity	60
2.3.1	Phase Noise	60
2.3.2	Intensity Noise	60
2.3.3	Polarization Noise	61
2.3.4	Key Technologies of Coherent Optical Communication Systems	61
2.4	Spatial Phase Conditions for Optical Heterodyne Detection	62
2.4.1	Spatial Phase Difference Conditions	62
2.4.2	Frequency Conditions	65
2.4.3	Polarization Conditions	66
2.5	Adaptive Optical Wavefront Correction	67
2.5.1	Wavefront Distortion Correction System	68
2.5.2	Wavefront Measurement and Correction	69
2.5.3	Wavefront-Free Measurement System	72
2.6	Summary and Prospects	73
2.7	Questions	74
2.8	Exercises	74
	References	78
3	Modulation, Demodulation, and Coding	81
3.1	Modulation	81
3.1.1	Basic Concepts	81
3.1.2	Analog and Digital Modulation	82

3.1.3	Direct and Indirect Modulation	83
3.1.4	Internal and External Modulation	83
3.2	External Modulation	84
3.2.1	Electro-Optic Modulation	84
3.2.2	Acousto-Optic Modulation	90
3.2.3	Magneto-Optic Modulation	93
3.3	Reverse Modulation	95
3.3.1	Cat's Eye Effect	95
3.3.2	Principle of Reverse Modulation	97
3.3.3	Cat's Eye Reverse Modulation System	100
3.4	Pulse-Like Position Modulation	100
3.4.1	Pulse-Like Position Modulation	100
3.4.2	Synchronization Technology	106
3.5	Direct Drive of Light Source	108
3.5.1	Single-Ended to Differential Converter	109
3.5.2	Level Adjustment	110
3.5.3	Laser Driver	110
3.5.4	Principle of Optical Feedback	114
3.6	Subcarrier Intensity Modulation	115
3.6.1	Subcarrier Intensity Modulation	116
3.6.2	BPSK Subcarrier Modulation	117
3.6.3	FSK Subcarrier Modulation	119
3.6.4	Intermodulation Distortion and Carrier-to-Noise Ratio	121
3.7	Orthogonal Frequency-Division Multiplexing	124
3.7.1	Basic Principles	124
3.7.2	Implementation of Discrete Fourier Transform in OFDM	127
3.7.3	Protection Interval and Cyclic Prefix	130
3.7.4	Peak-to-Average Power Ratio and Its Reduction Method	131
3.8	Space-Time Coding	132
3.8.1	Evolution of Space-Time Coding	132
3.8.2	Space-Time Coding in Optical Wireless Communication	135
3.8.3	Space-Time Decoding in Optical Wireless Communication	136
3.9	Channel Coding	137
3.9.1	Channel Coding	137
3.9.2	Linear Error Correction Code	138
3.9.3	Convolutional Code	139
3.10	Summary and Prospects	142
3.11	Questions	142
3.12	Exercises	143
	References	146

4	Atmospheric Channel, Channel Estimation, and Channel Equalization	149
4.1	Atmospheric Attenuation	149
4.1.1	Atmospheric Attenuation Coefficient and Transmittance	149
4.1.2	Absorption and Scattering of Atmospheric Molecules	151
4.1.3	Absorption and Scattering of Atmospheric Aerosol Particles	152
4.1.4	Atmospheric Window	153
4.1.5	Estimation of the Attenuation Coefficient	154
4.1.6	Transfer Equation	154
4.2	Atmospheric Turbulence Model	155
4.2.1	Atmospheric Turbulence	156
4.2.2	Atmospheric Turbulence Channel Mode	157
4.2.3	Log-Normal Turbulence Model	157
4.2.4	Gamma-Gamma Turbulence Model	162
4.2.5	Negative Exponential Distributed Turbulence Model	166
4.2.6	Atmospheric Structure Constant	168
4.2.7	Bit Error Rate Caused by Atmospheric Turbulence	169
4.3	Diversity Reception	171
4.3.1	Maximum Ratio Combining	172
4.3.2	Equal Gain Combining	172
4.3.3	Selective Combining	173
4.4	Channel Estimation	173
4.4.1	Concept of Channel Estimation	173
4.4.2	Least Squares Channel Estimation Algorithm	174
4.4.3	MMSE Based Channel Estimation	175
4.5	Channel Equalization	175
4.5.1	ISI and Channel Equalization	176
4.5.2	Time Domain Equalization	176
4.5.3	Linear Equalization	177
4.6	Impacts of Atmospheric Turbulence on BER	180
4.7	Summary and Prospects	182
4.8	Questions	182
4.9	Exercises	183
	References	187
5	White LED Communication	189
5.1	Light-Emitting Principle of LED	189
5.1.1	White LEDs	189
5.1.2	Light-Emitting Principle of LED	190

5.1.3	Light-Emitting Principle of White LED	191
5.1.4	Lighting Model of White LED	192
5.2	Background Noise Model for Internet of Vehicle	193
5.3	Multiplicative Noise Model	197
5.4	Optimal Layout of Light Source	200
5.5	Indoor Visible Light Channel	201
5.6	Receiver and Detection Technology	205
5.6.1	Receiver Front End	205
5.6.2	Receiving Array Design	208
5.7	Uplink of Visible Light Communication	209
5.7.1	Radio Frequency Uplink	209
5.7.2	Infrared Uplink	210
5.7.3	Laser Uplink	210
5.7.4	Visible Light Uplink	210
5.7.5	Isomorphic Uplink	211
5.8	Visible Light Communication Positioning	212
5.8.1	Received Optical Signal Strength Positioning	213
5.8.2	Fingerprint Identification Positioning	213
5.8.3	LED Identity Positioning	214
5.8.4	Visible Light Imaging Positioning	214
5.9	Summary and Prospects	214
5.10	Questions	215
5.11	Exercises	215
	References	218
6	Underwater Laser Communication	221
6.1	Overview of Underwater Laser Communication	221
6.2	Underwater Laser Communication System	223
6.2.1	Principle of Underwater Laser Communication	223
6.2.2	Underwater Channel	223
6.2.3	Characteristics of Underwater Laser Communication	224
6.3	Submarine Laser Communication	225
6.3.1	Forms of Submarine Laser Communication	225
6.3.2	Transmission of Each Dielectric Layer	225
6.3.3	Time Spreading	228
6.3.4	Energy Equation	229
6.3.5	Trends of Submarine Laser Communication	229
6.4	Summary and Prospects	230
6.5	Questions	230
6.6	Exercises	230
	References	231

7	Ultraviolet Communication	233
7.1	UV Light and Its Channel Characteristics	233
7.1.1	UV Light	233
7.1.2	Characteristics of UV Light	234
7.1.3	UV Atmospheric Channel	235
7.1.4	Characteristics of UV Atmospheric Channel	236
7.2	Characteristics of NLOS UV Transmission	239
7.2.1	Ellipsoid Coordinate System	239
7.2.2	UV Scattering Communication	239
7.2.3	NLOS Scattering Characteristics	241
7.3	Solar-Blind UV NLOS Communication Network	242
7.3.1	Wireless Mesh Communication Network	242
7.3.2	Wireless UV Mesh Communication Network	244
7.4	Summary and Prospects	246
7.5	Questions	246
7.6	Exercises	247
	References	247
8	Acquisition, Aiming, and Tracking Technology	249
8.1	Acquisition, Pointing, and Tracking System	249
8.1.1	Concepts	249
8.1.2	Operating Principle	250
8.2	Automatic Acquisition	252
8.2.1	Open-Loop Acquisition Mode	253
8.2.2	Scanning Modes	255
8.2.3	Performance of Acquisition	260
8.3	Automatic Tracking	262
8.3.1	Tracking System	262
8.3.2	Compound-Axis Control System	263
8.3.3	Accuracy of a Coarse Tracking Unit	264
8.3.4	Fine Tracking Unit	265
8.4	Fast Alignment Using Two-Dimensional Mirror	265
8.4.1	Introduction	265
8.4.2	Theoretical Model	267
8.4.3	Experiments	271
8.5	Alignment Error	280
8.5.1	Attenuation Model of Optical Power	280
8.5.2	Geometric Attenuation Model of Gaussian Beam with Alignment Error	282
8.5.3	Average Geometric Attenuation Model with Alignment Error	284
8.6	Summary and Prospects	285
8.7	Questions	285
8.8	Exercises	285
	References	286

9	Partially Coherent Optical Transmission	289
9.1	Basic Parameters of a Light Beam	289
9.1.1	Emission Beam	289
9.1.2	Mutual Interference Function	290
9.1.3	Beam Spreading, Drift, and Intensity Fluctuation	295
9.2	Partially Coherent Light Model	297
9.2.1	Description of Partially Coherent Light	297
9.2.2	Partially Coherent Beam	300
9.3	Beam Propagation in Atmospheric Turbulence	302
9.3.1	Beam Spread and Beam Drift	302
9.3.2	Drift and Spread of a Horizontally Propagating Beam	308
9.3.3	Drift and Spread of a Slant Propagating Beam	313
9.3.4	Fluctuation of Angle of Arrival	316
9.3.5	Influence of Beam Drift and Spread on a Communication System	321
9.4	Summary and Prospects	323
9.5	Questions	323
9.6	Exercises	323
	References	326
10	Optical Communication in the Future	329
10.1	X-ray Space Optical Communication	329
10.1.1	Backgrounds	329
10.1.2	X-ray Communication System	330
10.1.3	Development Directions and Prospects	335
10.2	Orbital Angular Momentum Multiplexing Communication	336
10.2.1	Vortex Beam	336
10.2.2	Generation of a Vortex Beam	337
10.2.3	OAM Multiplexing Communication System	341
10.3	Neutrino Communication	345
10.3.1	Neutrino	345
10.3.2	Neutrino Communication	346
10.3.3	Neutrino Communication System	346
10.3.4	Key Technologies of Neutrino Communication	346
10.3.5	Characteristics of Neutrino Communication	348
10.4	Gravitational Wave Communication	348
10.4.1	Detection of Gravitational Waves	348
10.4.2	Generation of Gravitational Waves	351
10.4.3	Difficulties in Gravitational Wave Detection	352
10.5	Terahertz Wave Communication	352
10.5.1	Terahertz Wave and Its Advantages	352
10.5.2	Terahertz Wave Transmitting Antenna	354
10.5.3	Terahertz Detector	356
10.5.4	Terahertz Wave Modulator	357

10.5.5	Transmission of Terahertz Waves in the Atmosphere	358
10.6	Summary and Prospects	359
10.7	Questions	359
10.8	Exercises	360
	References	360

Chapter 1

Optical Wireless Communication System



Optical wireless communication (OWC) shares the advantage of large communication capacity with microwave wireless and optical fiber communications, without the requirements of optical fiber cable installation or spectrum license. In this chapter, the system model and basic concepts of optical wireless communication are introduced.

1.1 System Model of Optical Wireless Communication

An optical wireless communication terminal comprises an optical antenna (telescope), laser transceiver, signal processing unit, and acquiring, pointing, and tracking (APT) system. Usually, a laser diode (LD) or light emitting diode (LED) is adopted as the light source in the transmitter, while a positive-intrinsic-negative (PIN) diode or avalanche photodiode (APD) is adopted as the detector in the receiver. The model of an optical wireless communication system is shown in Fig. 1.1.

1.1.1 Transmitter

In a transmitter, an optical carrier is modulated by a certain form of information, such as time-varying waveforms and digital symbols, generated by the information source. Then, the carrier (called light beam or light field) is emitted into the atmosphere or free space. The transmitter comprises source coding, channel coding, modulation, optical signal amplifier, and transmitting antenna.

Channel coding is an insertion of some redundant symbols to the source data stream, which facilitates error detection and correction at the receiver end. With a fundamental task to reduce the bit error rate and increase the reliability of the communication, channel coding results in a reduction in the data transmission rate due to the addition of redundancy.

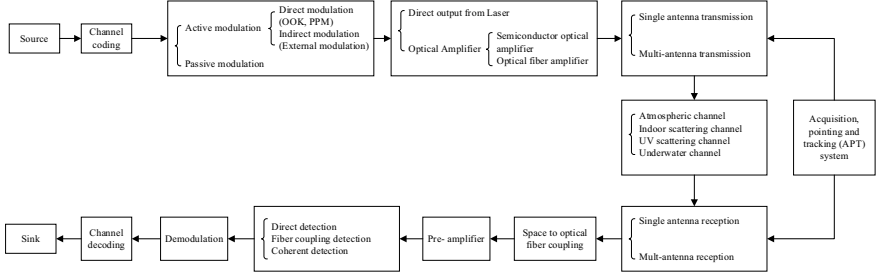


Fig. 1.1 Model of optical wireless communication

Modulation is a process of signal transformation. It changes certain characteristics of the optical signal, such as the amplitude, frequency, and phase, according to the characteristics of the encoded signal, and these changes happen in a regular way determined by the source signal itself. Thus, the relevant information of the source signal is successfully carried by the optical signal.

Modulation can be divided into two categories: active and passive. Modulation is considered active if the light source and the modulation process are co-located at the transmitter end, and modulation is passive if the modulated signal is generated on the other side of the light source; this is also known as reverse modulation. In contrast, modulating the laser power supply refers to direct modulation, whereas modulating the beam emitted by the laser is called indirect, or external, modulation.

If the required communication distance is large and the optical power directly output by the laser is insufficient, an optical amplifier is used to amplify the optical signal. Optical amplifiers include semiconductor optical and optical fiber amplifiers.

The transmitting antenna has several configurations, such as multi-antenna transmit/reception and single antenna transmit/reception. Herein, multi-antenna transmit/reception can effectively suppress the influence of atmospheric turbulence.

1.1.2 Receiver

The receiver comprises receiving antenna for optical signal collection, spatial light-fiber coupling unit, preamplifier, detector, demodulator, etc.

The optical signals sent by the transmitter are first collected by the receiving antenna, and then the light in free space is coupled into the optical fiber by the spatial light-fiber coupling unit. The optical fiber detector performs photoelectric conversion. Energy loss occurs in the process of coupling optical signals into the optical fiber.

Under certain circumstances, the signal coupled into the optical fiber is very weak, and it must be pre-amplified by an optical amplifier, the preamplifier, before the photoelectric conversion process.

Signal detection includes direct detection by a detector, spatial light-fiber coupling detection, distributed detection, and coherent detection. Here, direct detection means that the optical detector directly receives the optical signal collected by the antenna. Spatial light-fiber coupling detection means that the spatial light is coupled into the optical fiber, and the signal in the optical fiber is detected by the photo detector. Owing to the small dimensions of the optical fiber, small photosensitive area of the photoelectric converter, and small optical signal intensity required, the signal can be detected with high rate and sensitivity by coupling the spatial light into the optical fiber.

1.1.3 Channel

Wireless optical channels can be divided into atmospheric, indoor, ultraviolet light scattering, and underwater channels. The atmospheric channel is the most complicated because of the inevitable impacts of atmospheric turbulence and complicated meteorological conditions on the wireless optical channel. The channel transfer function can be expressed as

$$H(f) = H_T(f)H_c(f)H_r(f), \quad (1.1)$$

where $H_T(f)$, $H_c(f)$, and $H_r(f)$ represent the transfer function of the transmitter, channel, and receiver, respectively. The corresponding expression in time domain is

$$h(t) = h_T(t)h_c(t)h_r(t), \quad (1.2)$$

where $h_T(t)$, $h_c(t)$, and $h_r(t)$ represent the unit impulse response of the transmitter, channel, and receiver, respectively. The channel model is shown in Fig. 1.2.

The signal input to the demodulator can be expressed as

$$r(t) = A(t)[s(t) * h(t)] + n(t), \quad (1.3)$$

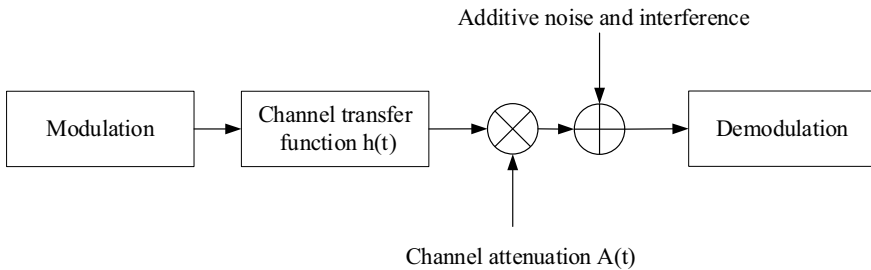


Fig. 1.2 Channel model

where $A(t)$ represents the fading of the channel, $s(t)$ is the output of the modulator, and $*$ denotes the convolution operation. For atmospheric laser communication, $A(t)$ results from atmospheric turbulence. For non-line of sight ultraviolet communication, $A(t)$ is primarily caused by single and multiple scattering of atmospheric molecules to the ultraviolet light. Finally, for indoor visible light communication, $A(t)$ is primarily caused by reflection in the indoor environment.

Without channel fading, the received signal can be expressed as

$$r(t) = s(t) + n(t), \quad (1.4)$$

where $n(t) \sim N(0, \sigma^2)$ is the white noise in additive Gaussian distribution, which generally represents the electronic noise of the receiver detector and its auxiliary circuit.

1.2 Laser Light Source

LD is a solid-state semiconductor device that can directly convert electrical energy into light. Other lasers, such as gas and liquid lasers, can also be used as light sources, but semiconductor lasers are the most common.

1.2.1 Principles of a Laser Diode

LD is a light emitting device that uses semiconductor material as its working material. The working materials commonly used include cadmium sulfide (CdS), gallium arsenide (GaAs), indium phosphide (InP), and zinc sulfide (ZnS). The specific process of generating the laser will be different depending on the structure of the working material. There are three types of excitation methods: electron beam excitation, electrical injection, and optical pumping. Semiconductor lasers can be divided into single heterojunction, double heterojunction, and homojunction lasers. Homojunction and single heterojunction lasers are mostly pulsed devices at room temperature, whereas double heterojunction lasers can achieve continuous operation at room temperature. The basic structure of a semiconductor laser comprises a double heterojunction planar strip structure, as shown in Fig. 1.3. The so-called heterojunction refers to a PN junction (may also be a P-P or N-N junction) composed of two semiconductor materials with different band gap widths. An ordinary PN junction is also called a homojunction.

Population inversion is a premise of laser production. The probability of stimulated radiation between two energy levels is related to the difference in the number of particles between the two energy levels. Usually, the number of atoms at the low energy level (E_1) is greater than the number of atoms at the high energy level (E_2).

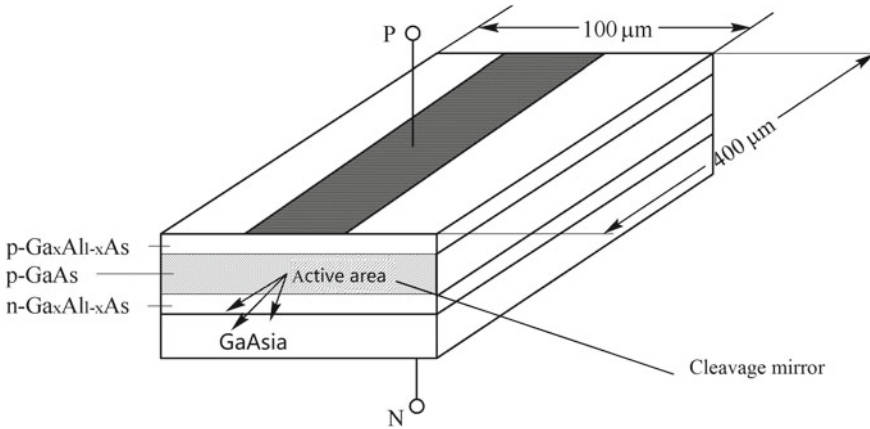


Fig. 1.3 Basic structure of a laser diode bar with double heterojunction plane

In this case, no laser light is generated. To generate laser, the number of atoms at the high energy level (E2) must be greater than that at the low energy level (E1) because stimulated radiation, which leads to light enhancement (known as optical amplification), occurs when a large number of atoms are at the high energy level (E2). To achieve this goal, a large number of atoms in the ground state must be excited to the metastable state (E2), so that the number of atoms in the high energy level (E2) greatly exceeds the number of atoms in the low energy level (E1). Thus, population inversion is achieved between the two energy levels.

Under certain conditions, the gain medium, resonant cavity, and pump source are essential to generate laser. Homojunction LD requires a heavily doped semiconductor material with a “direct band gap”. As the electron transitions from the conduction band to the valence band, the dominant stimulated radiation makes the emitted light laser. Due to the heavy doping, the active region of the degenerate semiconductor has a weak ability to bind electrons and holes, and a large injection current density is required to achieve population inversion. Therefore, it is difficult to achieve continuous operation at room temperature than at low temperatures. To lower the current density threshold, single and double heterojunction semiconductor lasers have been studied. The band gap difference of different semiconductor materials causes the refractive index of the active region to be higher than that of the adjacent medium; therefore, photons are also confined in the active region. The confinement of carriers and photons causes the threshold current density of the laser to drop significantly, thus achieving continuous operation at room temperature.

1.2.2 Characteristics of a Laser Diode

LD is a device in which a certain semiconductor material is used as working material to produce stimulated emission. The working principle is to realize population inversion for non-equilibrium carriers between the energy bands of semiconductor material (conduction and valence bands) or between the energy bands of semiconductor material and the energy levels of impurity (acceptor or donor) through certain excitation methods. When a large number of electrons and holes in a population inversion state recombine, stimulated radiation occurs.

In general, there are three excitation methods for semiconductor lasers: electrical injection, optical pumping, and high-energy electron beam excitation. Electrical injection semiconductor lasers are generally semiconductor junction diodes made of gallium arsenide (GaAs), cadmium sulfide (CdS), indium phosphide (InP), zinc sulfide (ZnS), etc., and are biased in the forward direction. Current is injected to excite and generate stimulated emission in the junction plane area. Optical pumped semiconductor lasers generally adopt N- or P-type semiconductor single crystals (e.g., GaAs, InAs, and InSb) as working materials and use laser from other lasers as the light pump excitation source. High-energy electron beam-excited semiconductor lasers adopt N- or P-type semiconductor single crystals (e.g., PbS, CdS, and ZnO) as working materials and are excited by external injection of high-energy electron beams. The LD is a threshold device: when the injection current is less than the threshold, the gain of the resonant cavity is not high enough to overcome the loss, and the population inversion cannot be realized in the active area. Spontaneous emission is dominant, and ordinary fluorescence is emitted, similar to LED. As the injected current increases and reaches the threshold, population inversion is realized in the active area. Stimulated radiation is dominant, and a sharp-spectrum laser with a clear pattern is thus emitted. Semiconductor lasers are very sensitive to temperature, and their output power varies considerably with temperature. The main reason is that the external differential quantum efficiency and threshold current of semiconductor lasers vary with temperature.

LD is a type of laser device that uses semiconductor material as the working material. In addition to the common characteristics of lasers, it has the following features: small size, low weight, low driving power and current, high efficiency, long operating life, direct electrical modulation, and easy integration with various optoelectronic devices.

1.2.3 Nonlinearity Correction

The laser is a nonlinear device with threshold characteristics. The nonlinearity will produce harmonic distortion under the excitation of the modulation signal.

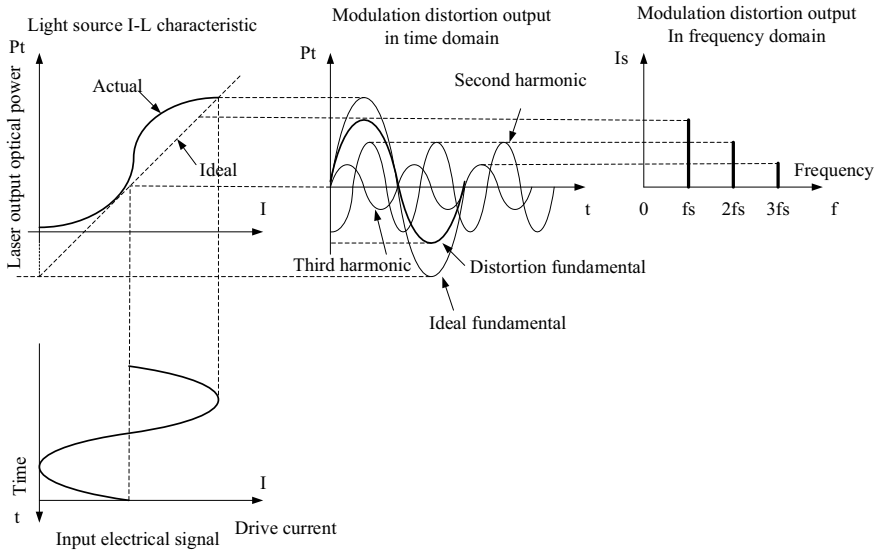


Fig. 1.4 Influence of laser nonlinearity on subcarrier modulation

1.2.3.1 Effect of Static Nonlinearity on Subcarrier Modulation

The I-L characteristic curve using the relationship between the input current and the output optical power can be used to characterize the nonlinear distortion of the laser. As shown in Fig. 1.4, nonlinear distortion can be simply summarized as a distortion-free signal; the output response signal produces waveform distortion in the time domain, and new harmonic components appear in the frequency domain.

1.2.3.2 Pre-distortion Compensation of I-L Characteristic

We assume that the relationship between the laser output optical power and the drive current is

$$P = a_0 + a_1 I + a_2 I^2 + a_3 I^3 \quad (1.5)$$

and the pre-distortion signal generated by the pre-distorter is

$$I' = b_2 I^2 + b_3 I^3, \quad (1.6)$$

The basic principle of the pre-distortion can be described as follows. When the signal voltage increases, the pre-distortion model of the laser front end makes a fast increase in the current through the laser. It also cancels out the nonlinear distortion

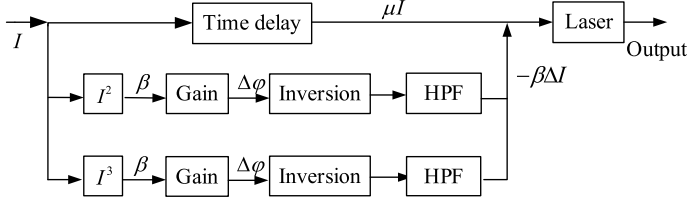


Fig. 1.5 Pre-distortion of I-L characteristic

caused by the smaller increase in optical power compared with the increase in the current. Through the joint action of the pre-distortion model and the laser, the linear relationship between the output optical power and the input current is maintained.

As shown in Fig. 1.5, the original input signal is separated into three channels by the coupler: the main channel (processing the fundamental wave component) and two secondary channels (processing the second and third-order distortion components). Assuming that the gains of the two sub-channels are the same, the pre-distorter can be regarded as the signal generated by the sub-channel. The main channel directly enters the combiner through time delay, and the secondary channel is output by the combiner after passing through the attenuator, inverter, and filter. At this time, the output three-way composite signal is passed through the laser, and the output optical power of the laser changes to

$$P' = \sum_{m=1}^3 a_m (\alpha I - \beta I')^m = \sum_{m=1}^3 a_m \left(\alpha I - \beta \sum_{n=2}^3 b_n I^n \right)^m. \quad (1.7)$$

Equation (1.7) can be expanded as power series

$$P' = \sum_{m=1}^3 D_m I^m, \quad (1.8)$$

where

$$D_1 = \alpha a_1, \quad (1.9)$$

$$D_2 = \alpha^2 a_2 - \beta a_1 b_2, \quad (1.10)$$

$$D_3 = \alpha^3 a_3 - \beta a_1 b_3 - 2\alpha \beta a_2 b_2. \quad (1.11)$$

From Eqs. (1.9)–(1.11), it can be calculated that the condition for eliminating the second-order nonlinear distortion is

$$\beta = \frac{\alpha^2 a_2}{a_1 b_2}, \quad (1.12)$$

and the condition for eliminating the third-order nonlinear distortion is

$$\beta = \frac{\alpha^3 a_3}{a_1 b_3 + 2\alpha a_2 b_2}. \quad (1.13)$$

The condition to simultaneously eliminate the second- and third-order nonlinear distortions is

$$\alpha = \frac{a_1 a_2 b_3}{b_2 (a_1 a_3 - 2a_2^2)}, \quad (1.14)$$

where α is the gain of the main channel, and β is the gain of the secondary channel.

1.3 Device Response Characteristics

1.3.1 Response Characteristics of a Semiconductor Laser

The response characteristics of a semiconductor laser are related to its chirality oscillation frequency, and the frequency of the chirality oscillation f_0 can be obtained from the following rate equations [1]:

$$\frac{dN}{dt} = \frac{I}{qV} - A(N - N_{om})S - \frac{N}{\tau_N}, \quad (1.15)$$

$$\frac{dS}{dt} = \Gamma\beta \frac{N}{\tau_N} - \frac{S}{\tau_p} + \Gamma A(N - N_{om})S, \quad (1.16)$$

where N is the carrier concentration, S is the photon density, I is the current injected into the active zone, N_{om} is the transparent carrier concentration, q is the electron charge, and the volume of the active zone $V = WLD$. Here, W , L , and D are the active zone width, cavity length, and thickness, respectively. Moreover, A is the gain factor, and τ_p and τ_n are the photon and carrier lifetimes, respectively. Finally, Γ is the optical confinement factor, and β is the spontaneous radiation coefficient.

With small signal modulation, the variables in Eqs. (1.15) and (1.16) can be expressed as [2]

$$\begin{cases} I(t) = I_0 + I_1 \exp(i\omega t) \\ N(t) = N_0 + N_1 \exp(i\omega t) \\ S(t) = S_0 + S_1 \exp(i\omega t) \end{cases} \quad (1.17)$$

where I_0 , N_0 , and S_0 are the steady-state direct flows, and $I_1 \exp(i\omega t)$, $N_1 \exp(i\omega t)$, and $S_1 \exp(i\omega t)$ are the transient AC flows respectively.

Substituting Eq. (1.17) into Eqs. (1.15) and (1.16) and neglecting second-order minima yields

$$i\omega N_1 e^{i\omega t} = \frac{I_0}{qV} + \frac{I_1 e^{i\omega t}}{qV} - \frac{N_0}{\tau_n} - \frac{N_1 e^{i\omega t}}{\tau_n} - \left[AS_0(N_0 - N_{om}) + AS_1(N_0 - N_{om})e^{i\omega t} + AN_1 S_0 e^{i\omega t} \right], \quad (1.18)$$

$$i\omega S_1 e^{i\omega t} = \Gamma\beta \frac{N_0}{\tau_n} + \Gamma\beta \frac{N_1 e^{i\omega t}}{\tau_n} - \frac{S_0}{\tau_p} - \frac{S_1 e^{i\omega t}}{\tau_p} + [AS_0(N_0 - N_{om}) + AS_1(N_0 - N_{om})e^{i\omega t} + AN_1 S_0 e^{i\omega t}]. \quad (1.19)$$

When the laser is in steady state, $dS/dt = 0$ and $dN/dt = 0$ [3]. Let $N_{om} \approx 0$, $\Gamma \approx 1$, $\beta \approx 0$, and $A = a_g/(1 + \varepsilon S_0)$, and introduce the gain saturation term [4]. Then Eqs. (1.18) and (1.19) can be simplified to obtain the laser chirality oscillation frequency f_0 as

$$f_0 = \left(\frac{a_g S_0}{\tau_p (1 + \varepsilon S_0)} \right)^{1/2}. \quad (1.20)$$

It can be seen that f_0 is proportional to the differential gain coefficient a_g and the steady-state photon density S_0 and inversely proportional to the photon lifetime τ_p and the gain saturation factor ε . The following simulation analyzes the effect of the above factors on the response characteristics of the laser.

Figures 1.6 and 1.7 show the impulse response and frequency response of the laser at different bias currents $I_0(S_0)$. I_0 is set to 40, 50, 60, and 70 mA, respectively, and the initial conditions are $\varepsilon = 1 \times 10^{-25}$, $R = 0.3$, and $a_g = 1.4 \times 10^{-12}$.

As can be seen from Fig. 1.6, when the bias current I_0 is increased, the chirality oscillation frequency increases, the chirality oscillation amplitude decreases, and the pulse waveform distortion is significantly improved. It can be seen from Fig. 1.7 that by increasing the bias current I_0 , the chirality oscillation frequency f_0 can be increased, thus increasing the modulation bandwidth of the laser.

Figures 1.8 and 1.9 show the impulse and frequency response curves of the laser at different differential gain coefficients a_g . I_0 is set to 1.4×10^{-12} , 2.4×10^{-12} , 3.4×10^{-12} , and 4.4×10^{-12} , respectively, and the initial conditions are $I_0 = 40$ mA, $\varepsilon = 1 \times 10^{-25}$, and $R = 0.3$.

Fig. 1.6 Impulse response curves at different bias currents I_0

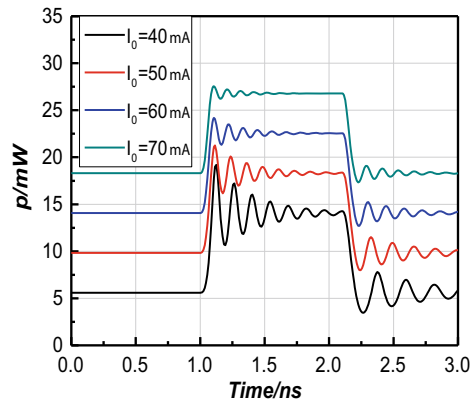


Fig. 1.7 Frequency response curves at different bias currents I_0

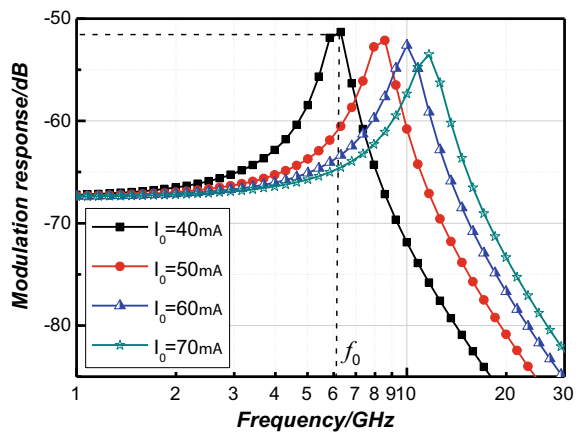


Fig. 1.8 Impulse response curves for different breeze gain coefficients a_g

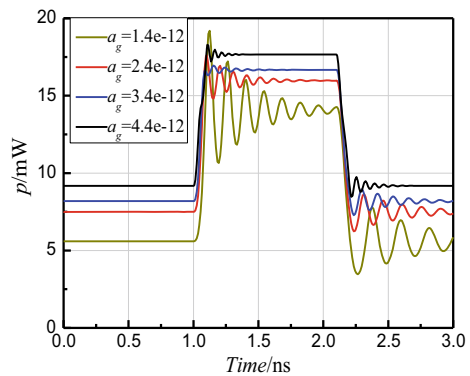
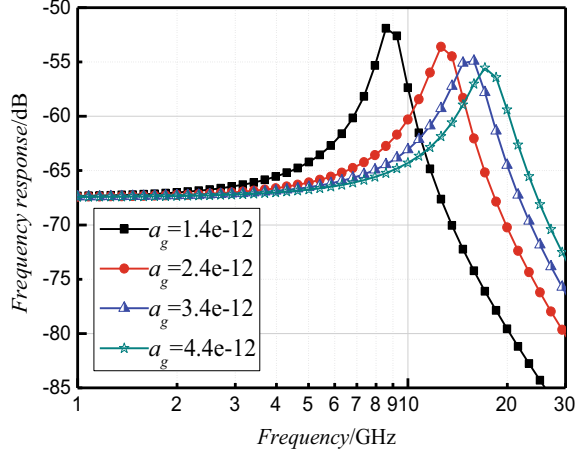


Fig. 1.9 Frequency response curves for different breeze gain coefficients a_g



As can be seen from Fig. 1.8, by increasing the differential gain coefficient a_g , the chirality oscillation process is shortened, and the waveform distortion phenomenon basically disappears. It can be seen from Fig. 1.9 that by increasing a_g , the chirality oscillation frequency f_0 can be increased, which in turn increases the modulation bandwidth of the laser.

Figures 1.10 and 1.11 show the impulse and frequency response curves of the laser at different gain saturation factors ε , which is set as 1×10^{-25} , 25×10^{-25} , 50×10^{-25} , and 100×10^{-25} , respectively. The initial conditions are $I_0 = 40$ mA, $a_g = 1.4 \times 10^{-12}$, and $R = 0.3$.

As can be seen from Fig. 1.10, an increase in the gain saturation factor ε has a damping effect on the chirality oscillation and shortens the chirality oscillation process. Figure 1.11 shows that by increasing ε , the chirality oscillation frequency f_0 decreases and the modulation bandwidth of the laser decreases.

Figures 1.12 and 1.13 show the impulse and frequency response curves of the laser at different reflectance values $R(\tau_p)$, respectively.

Fig. 1.10 Impulse response curves for different gain saturation factors ε

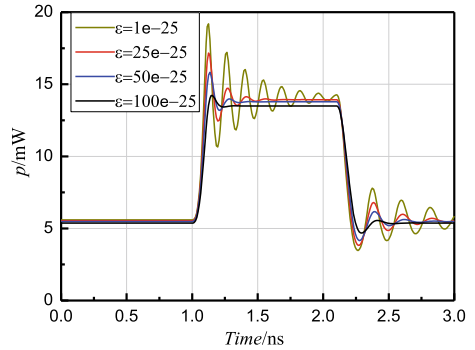


Fig. 1.11 Frequency response curves for different gain saturation factors ε

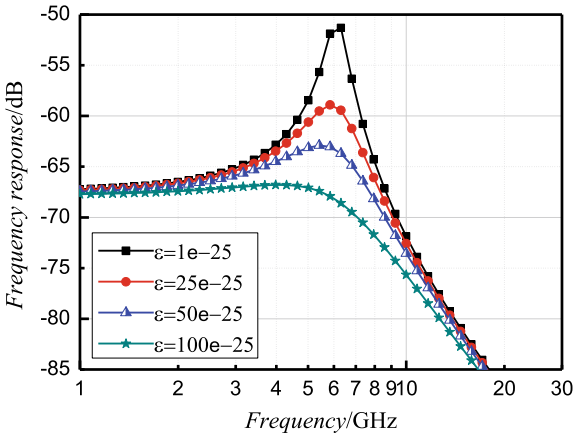


Fig. 1.12 Impulse response curves for different reflectivity values R

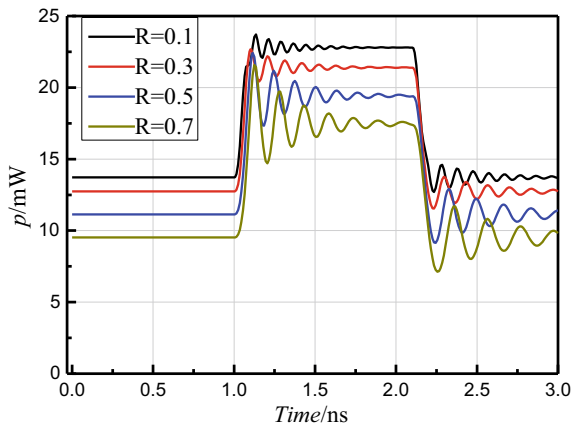
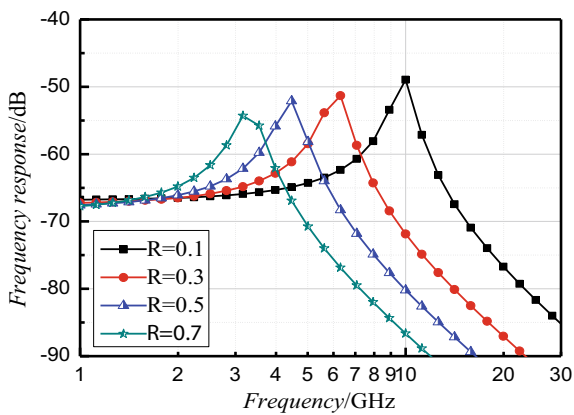


Fig. 1.13 Frequency response curves for different reflectivity values R



The initial conditions are $I_0 = 40 \text{ mA}$, $a_g = 1.4 \times 10^{-12}$, and $\varepsilon = 1 \times 10^{-25}$. As can be seen from Fig. 1.12, by increasing the reflectivity R , the chirality oscillation frequency decreases, the chirality oscillation amplitude increases, and the waveform distortion becomes increasingly severe. It can be seen from Fig. 1.13 that by increasing R , the chirality oscillation frequency f_0 decreases and the modulation bandwidth of the laser decreases.

The above analysis shows that increasing the bias current and differential gain coefficient and reducing the reflectivity and gain saturation factor can suppress the pulse waveform distortion and increase the modulation bandwidth.

1.3.2 Response Characteristics of a PIN Photodetector

As shown in Fig. 1.14, the equivalent circuit model for a PIN photodetector is modelled using Eqs. (1.21), (1.22) and (1.23), assuming that the light is incident from the N region, considering that the I region is electrically neutral [5, 6].

Region N:

$$\frac{dP_n}{dt} = \frac{P_{in}(1-r)}{h\nu} [1 - \exp(-\alpha_n W_n)] - \frac{P_n}{\tau_p} - \frac{I_p}{q}, \quad (1.21)$$

Region P:

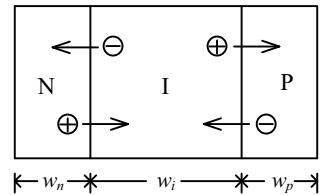
$$\frac{dN_p}{dt} = \frac{P_{in}(1-r)[1 - \exp(-\alpha_p W_p)]}{h\nu \exp(\alpha_n W_n + \alpha_i W_i)} - \frac{N_p}{\tau_n} - \frac{I_n}{q}, \quad (1.22)$$

Region I:

$$\frac{dN_i}{dt} = \frac{P_{in}(1-r)[1 - \exp(-\alpha_i W_i)]}{h\nu \exp(\alpha_n W_n)} - \frac{N_i}{\tau_{nr}} - \frac{N_i}{\tau_{ni}} + \frac{I_n}{q}. \quad (1.23)$$

In Eqs. (1.21), (1.22) and (1.23), P_n and N_p are the total number of excess holes and electrons in N and P region, respectively; τ_n and τ_p are the lifetimes of the electron and hole in P and N region, respectively; I_n and I_p are the electron and hole currents of minority carriers in P and N region, respectively; q is the electron charge,

Fig. 1.14 One-dimensional structure of a PIN photodetector



P_{in} is the input optical power, N_i is the total number of electrons in I region, τ_{nr} is the lifetime of electron's recombination in I region, τ_{nt} is the electron drift velocity in I region, r is the reflectivity; α_n , α_p , and α_i are the light absorption coefficients of N, P and I region, respectively; $h\nu$ is the photon energy, and W_n , W_p and W_i are the widths of N, P and I region, respectively.

To transform the parameters into circuit variables, we introduce a constant C_{nc} and let [7]:

$$P_n = \frac{V_p}{q} C_{nc}, \quad N_p = \frac{V_n}{q} C_{nc}, \quad N_i = \frac{V_i}{q} C_{nc}. \quad (1.24)$$

Then, substituting Eq. (1.24) into Eqs. (1.21), (1.22), and (1.23) and using $I_n = \beta_n P_{in} + V_n/R_{nd}$ and $I_p = \beta_p P_{in} + V_p/R_{pd}$, we obtain [8]

$$\frac{P_{in}}{V_{op}} = C_{nc} \frac{dV_p}{dt} + \frac{V_p}{R_p} + \beta_p P_{in} + \frac{V_p}{R_{pd}}, \quad (1.25)$$

where $V_{op} = \frac{h\nu}{q(1-r)[1-\exp(-\alpha_n W_n)]}$, $R_p = \frac{\tau_p}{C_{nc}}$, $I_{op} = \frac{P_{in}}{V_{op}}$, and $R_{pd} = R_p[\text{ch}(W_n/L_p) - 1]$,

$$\frac{P_{in}}{V_{on}} = C_{nc} \frac{dV_n}{dt} + \frac{V_n}{R_n} + \beta_n P_{in} + \frac{V_n}{R_{nd}}, \quad (1.26)$$

where $V_{on} = \frac{h\nu \exp(\alpha_n W_n + \alpha_i W_i)}{q(1-r)[1-\exp(-\alpha_p W_p)]}$, $R_n = \frac{\tau_n}{C_{nc}}$, $I_{on} = \frac{P_{in}}{V_{on}}$, $R_{nd} = R_n[\text{ch}(W_p/L_n) - 1]$, and

$$\frac{P_{in}}{V_{oi}} + \beta_n P_{in} + \frac{V_n}{R_{nd}} = C_{nc} \frac{dV_i}{dt} + \frac{V_i}{R_{nr}} + \frac{V_i}{R_{nt}}, \quad (1.27)$$

where $V_{oi} = \frac{h\nu \exp(\alpha_n W_n)}{q(1-r)[1-\exp(-\alpha_i W_i)]}$, $R_{nr} = \frac{\tau_{nr}}{C_{nc}}$, $R_{nt} = \frac{\tau_{nt}}{C_{nc}}$, $I_{oi} = \frac{P_{in}}{V_{oi}}$, $I_i = \frac{V_i}{R_{nt}}$, β_n and β_p are the coefficients between the incident optical power and diffusion current, L_n and L_p are the diffusion lengths of the electrons and holes, respectively.

Taking the N-I interface as an example, the current at the output is: $I_j = I_i + I_p$, in addition to the need to consider the chip parasitic parameters (C_c , R_c , R_d) and package parasitic parameters (L_e , R_e , C_e) within the photodetector package parasitic parameters from the carrier and gold wire introduced during the chip packaging process [9]. From Eqs. (1.25), (1.26) and (1.27) and the above analysis, the equivalent circuit model of the PIN photodetector is obtained as Fig. 1.15.

The responsiveness of a PIN photodetector to a high-speed incident light signal is characterized by the response time. The magnitude of the response time is related to the carrier crossing time and the RC time constant in the I region [10], which are analyzed below using the model in Fig. 1.15.

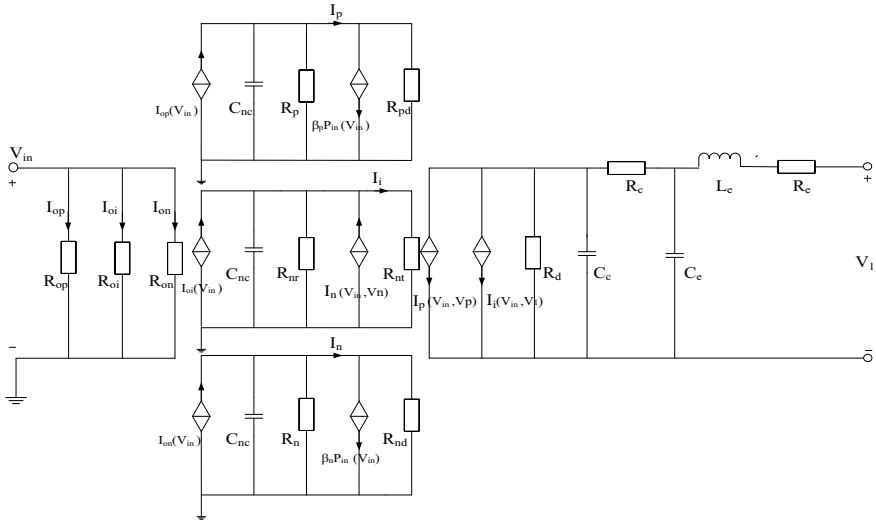


Fig. 1.15 Equivalent circuit model of a PIN photodetector

Figures 1.16 and 1.17, respectively, show the impulse response and frequency response curves of the PIN photodetector for I-zone widths W_i of 3, 10, 50, and 100 μm . Initial conditions are $R_c = 10 \Omega$ and $A = 12 \times 10^{-9}$.

It is observed that, choosing a suitable I-zone width W_i , which minimizes the impulse response time, is an effective way to well suppress the impulse waveform distortion and increase the frequency response bandwidth.

Figures 1.18 and 1.19 show the impulse response and frequency response curve of the PIN photodetector for chip parasitic resistance R_c of 10, 20, 30 and 40 Ω , respectively. The initial conditions are $W_i = 4.2 \mu\text{m}$ and $A = 12 \times 10^{-9}$.

Fig. 1.16 Impulse response for different I-zone widths

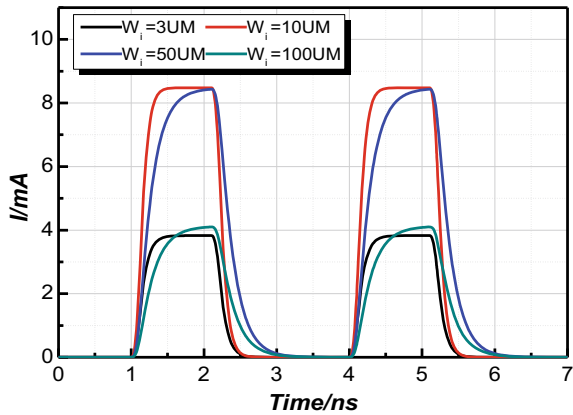


Fig. 1.17 Frequency response for different I-zone widths

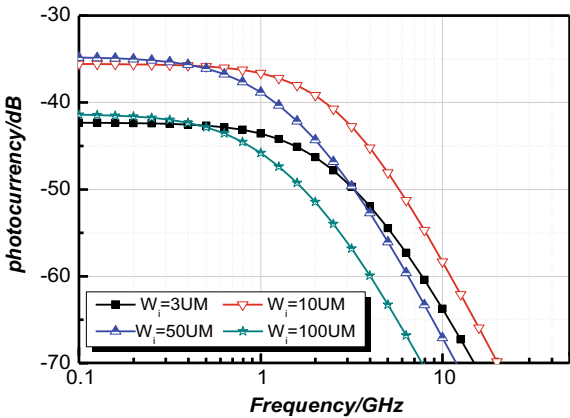


Fig. 1.18 Impulse response for different parasitic resistances

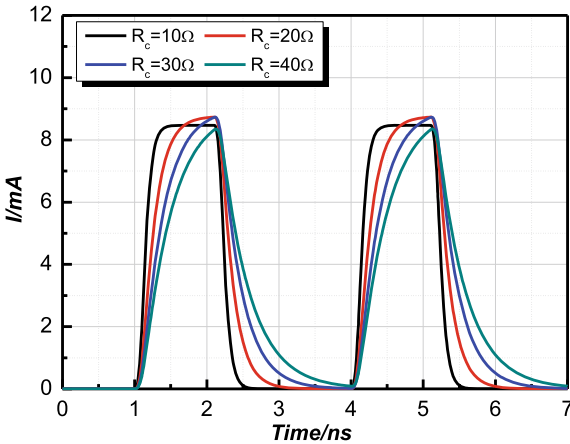
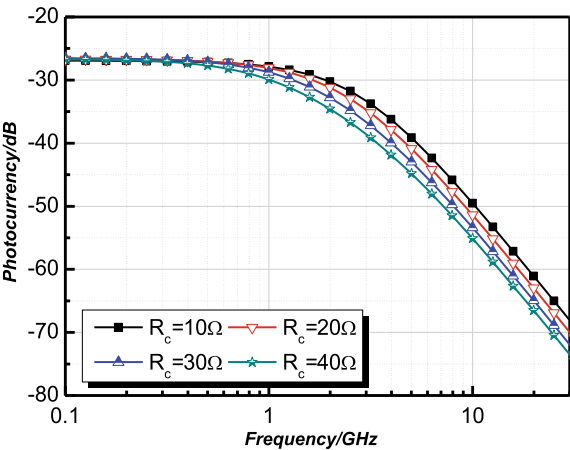


Fig. 1.19 Frequency response for different parasitic resistances



It is observed that, reducing the parasitic resistance R_c and decreasing the pulse response time is an effective way to suppress the pulse waveform distortion and increase the frequency response bandwidth. The simulation is omitted for chip parasitic capacitance C_c since it has a same principle as the resistor.

Figures 1.20 and 1.21 show the impulse response and frequency response curves of a PIN photodetector for different photosensitive surfaces A . The photosensitive surfaces A are taken as: 6×10^{-7} , 12×10^{-8} , 6×10^{-8} , 12×10^{-9} . The initial conditions are $W_i = 4.2 \mu\text{m}$, $R_c = 10 \Omega$.

It is observed that, reducing the photosensitive surface A is beneficial to suppress the pulse waveform distortion and increase the frequency response bandwidth.

The above analysis indicates that by selecting a suitable I-zone width, reducing the chip parasitic resistance, capacitance and photosensitive surface area can suppress pulse waveform distortion and increase the frequency response bandwidth.

Fig. 1.20 Impulse response at different photosensitive surfaces

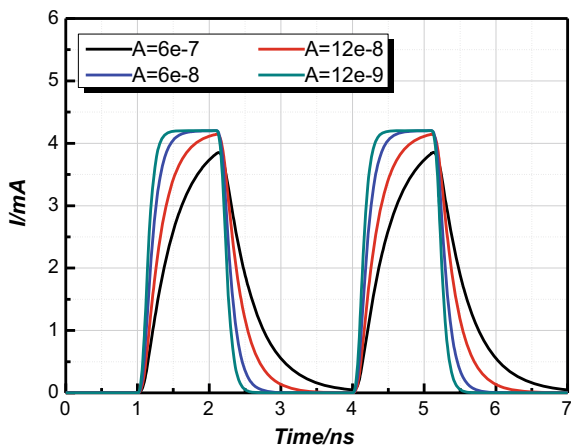
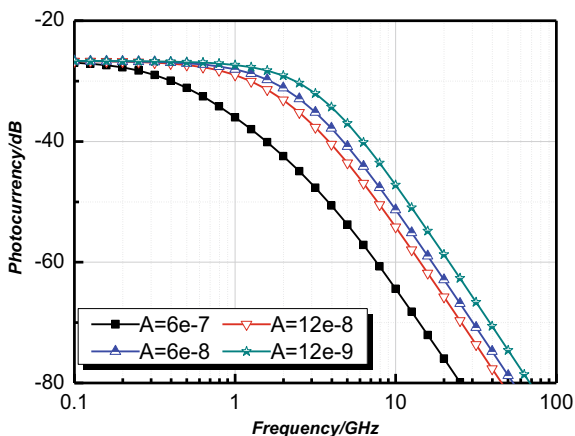


Fig. 1.21 Frequency response at different photosensitive surfaces



1.4 Surface Plasmon Polarization

Surface Plasmon Polarization (SPPs) are surface electromagnetic waves generated by the interaction of light waves incident on a metal surface with freely vibrating electrons on the metal surface [11]. The research on SPPs is increasing dramatically due to their unique properties such as subwavelength characteristics, plasmonic waveguide, and local field enhancement, all of which imply great research value. The emergence of SPPs has led to new physical phenomena in nano-optoelectronic devices, which have greatly improved performance compared to conventional optoelectronic devices.

Two main optical properties of metal nanoparticles are used in the design structure: (1) When incident light is irradiated to metal nanoparticles, metal nanoparticles produce significant surface plasmon effect, which in turn enhances the local equipartition resonance field and eventually causes the enhancement of absorption in the active medium. (2) When the incident light is irradiated to the metal nanoparticles, the metal nanoparticles produce significant surface plasmon effect, which in turn enhances the light scattering and makes the path of the incident light into the the active medium greatly increased, and finally causes the enhancement of the active medium absorption. The designed structure is shown in Fig. 1.22. Figure 1.22a shows the three-dimensional structure, where the substrate is a silicon photodetector (including p, i, and n layers) structure, and the top is antireflective film SiO_2 and spherical metal Ag nanoparticle array with a periodic distribution. Here, the thickness of the antireflective film SiO_2 is denoted by h , the thickness of the silicon substrate is denoted by

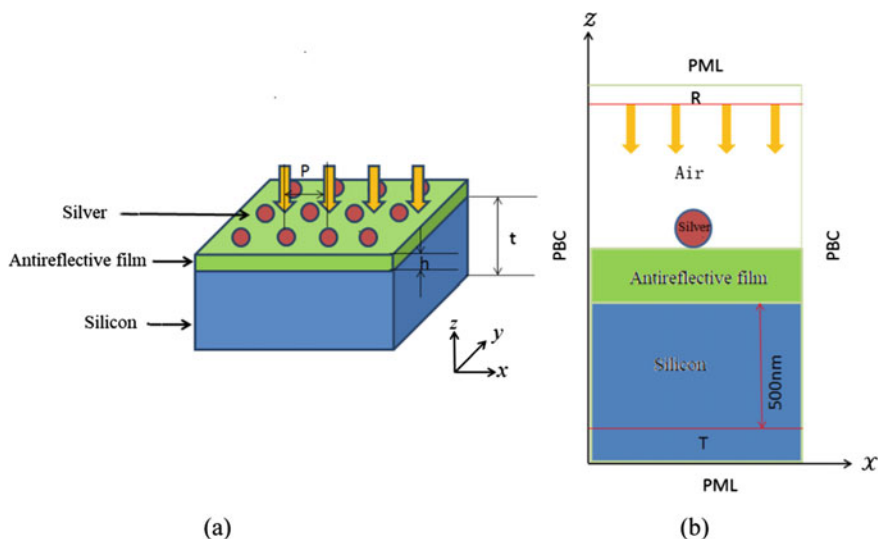


Fig. 1.22 Structure of a silicon thin-film photodetector with an array of Ag nanoparticles attached to the surface. **a** Three-dimensional structure **b** Cross-sectional view of the mathematical model

t , the distance between two adjacent particle spherical centers in the Ag nanoparticle array is denoted by P , and the diameter of the spherical Ag nanoparticles is denoted by d . Figure 1.22b shows the cross-sectional view of the 3D model in the x - z plane, and the periodic boundary conditions (PBC) are set around the model (i.e., axial x direction and axial y direction), and the perfectly matched layer (PML) boundary conditions are set at the top and bottom (i.e., axial z direction). The plane wave is incident perpendicular to the upper surface of the photodetector along the negative direction of the axis z (i.e., the angle of incidence at this point is 0°). The incident light source is the AM1.5 solar spectrum in the wavelength range of 400–1100 nm [12], and the optical constants of Ag in the model are taken from reference [13].

1.4.1 Effect of Different Incident Light Directions on the Light Absorption Performance of Silicon Substrates

With the structure parameters $P = 110$ nm and $d = 100$ nm, when the light with wavelength in the range of 600–800 nm is incident to the silicon thin film photodetector with thickness of 500 nm, the reflection efficiency, transmission efficiency and absorption efficiency of the incident light wave are calculated as shown in Fig. 1.23.

As shown in Fig. 1.23a, b, for the presence of spherical Ag nanoparticle arrays and anti-reflection film SiO_2 on the surface of the silicon thin-film photodetector, the values of the reflection efficiency and transmission efficiency are lower and the corresponding values of absorption efficiency are higher. In Fig. 1.23c, d, when there are no spherical Ag nanoparticle arrays and anti-reflective film SiO_2 on the surface of the silicon thin-film photodetector, the values of light reflection efficiency as well as transmission efficiency are higher, and the corresponding values of light absorption efficiency are lower. Compared with Fig. 1.23b–d, the silicon thin-film photodetector in Fig. 1.23a has the highest light absorption efficiency, and the photodetector composite structure light absorption efficiency remains above 65% in the range of 0° to 65° , which does not change significantly with the incident angle when the incident angle exceeds 65° . Such results show that the light absorption efficiency of silicon thin-film photodetector can be improved to a certain extent at different incidence angles by introducing spherical Ag nanoparticle arrays and anti-reflective film SiO_2 on the surface of silicon thin-film photodetector. It is also found that, compared with bare silicon photodetector, the light absorption efficiency of silicon thin-film photodetector can be improved up to 60% at wavelength of 700 nm with incidence angle from 0° to 65° .

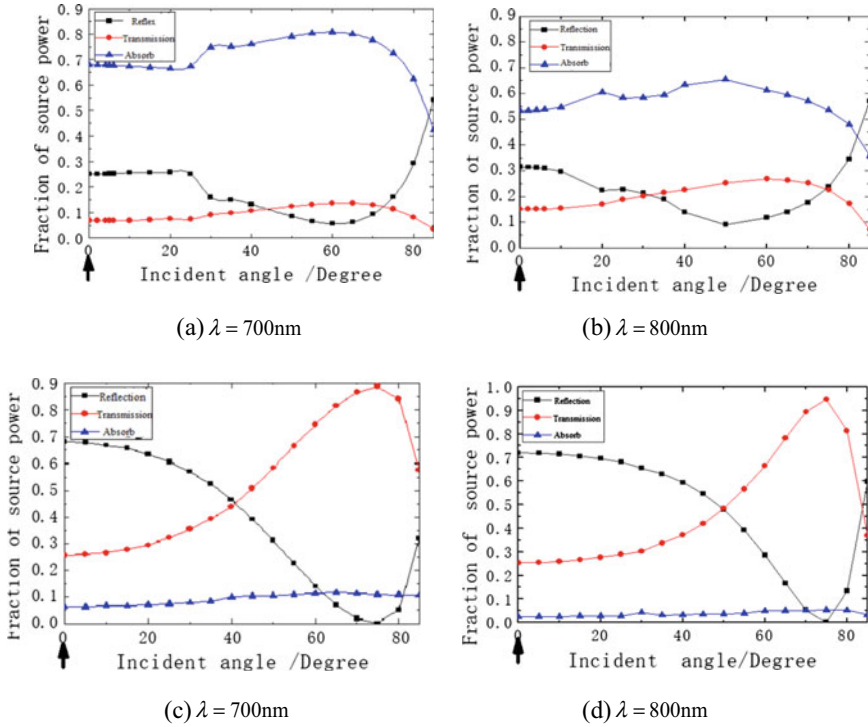
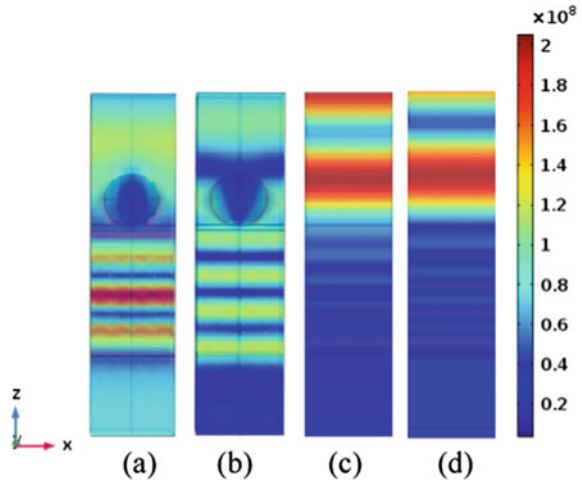


Fig. 1.23 Variation of absorption efficiency, reflection efficiency and transmission efficiency with incident angle for $P = 100\text{ nm}$ and $d = 110\text{ nm}$

1.4.2 Electric Field Modulus Distribution on the x - z Cross Section of the Photodetector

Figure 1.24 shows the electric field modulus distribution on the cross-section of the photodetector. The angles corresponding to (a)–(d) in the figure are the incident angles (i.e., 0°) indicated by the arrows in Fig. 1.23a–d. Figure 1.24c, d are the electric field intensity distributions when only the silicon thin film photodetector is available. This indicates that the majority of the incident light is reflected back. Figure 1.24a, b show the electric field intensity distribution when the spherical metal Ag nanoparticle array and the anti-reflective film SiO_2 are added to the top of the silicon thin film photodetector. As we can see, the electric field intensity around the Ag nanoparticles is significantly enhanced, indicating that the metallic nanoparticles have produced localized surface plasmon resonance. Due to the strong scattering effect of metal nanoparticles, most of the incident light is scattered to the inside of the silicon thin film photodetector. Compared with bare silicon photodetectors, its reflection efficiency and transmission efficiency are significantly reduced, and the absorption efficiency is improved. In Fig. 1.24b, the coupling efficiency between the

Fig. 1.24 Distribution of electric field modulus on the cross section of the photodetector at the angles indicated by the arrows in Fig. 1.15(a), (b), (c) and (d)



incident light wave and the substrate waveguide mode is relatively low, which leads to little improvement in the light absorption efficiency of the photodetector.

1.5 Signal Detection

The detection methods of optical wireless communication are divided into two categories: intensity modulation/direct detection (as shown in Fig. 1.25) and coherent detection. In this section, direct detection is discussed.

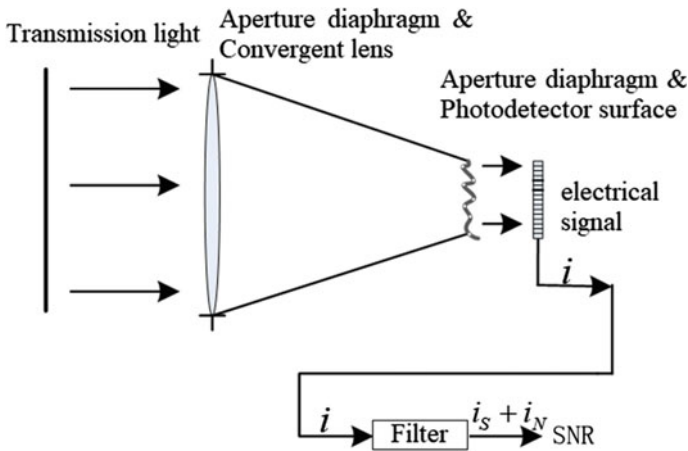


Fig. 1.25 Intensity modulation/direct detection

1.5.1 Direct Detection

The photoelectric direct detection system directly incidents the light signal on the photosensitive surface of the photo detector. The photo detector outputs the corresponding current and voltage responding to the intensity (amplitude) of the incident light radiation. Assume that the light field incident on the surface of the photo detector is

$$E_s(t) = A \cos \omega t. \quad (1.28)$$

The average optical power of the light field is

$$P_s = \overrightarrow{E_s^2}(t) = \frac{A^2}{2}, \quad (1.29)$$

where $\overrightarrow{E_s^2}(t)$ is the time average of $E_s^2(t)$. The output current of the photodetector is

$$I_s = \frac{e\eta}{2h\nu} A^2, \quad (1.30)$$

where $\frac{e\eta}{2h\nu}$ is known as photoelectric conversion constant. The output power of the photo detector is

$$P_0 = \left(\frac{e\eta}{h\nu}\right)^2 P_s^2 R_L, \quad (1.31)$$

where R_L is the load resistance of the photo detector. The photocurrent is proportional to the square of the amplitude of the optical field, and the electrical output power is proportional to the square of the incident optical power. This is known as the square law characteristic of the photo detector. For amplitude modulation

$$E_s(t) = A[1 + d(t)] \cos \omega t, \quad (1.32)$$

the output current of the photo detector is

$$i_s = \frac{1}{2} \alpha A^2 + \alpha A^2 d(t). \quad (1.33)$$

The signal-to-noise ratio of the output signal power can be expressed as

$$(\text{SNR})_p = \frac{\left(\frac{P_s}{P_n}\right)^2}{1 + 2\left(\frac{P_s}{P_n}\right)}, \quad (1.34)$$

where p_s and p_n are the signal and noise power at the output, respectively. If $\frac{p_s}{p_n} \ll 1$, then $(\text{SNR})_p \approx \left(\frac{p_s}{p_n}\right)^2$, and if $\frac{p_s}{p_n} \gg 1$, then $(\text{SNR})_p \approx \frac{1}{2} \left(\frac{p_s}{p_n}\right)$, which indicates that the direct detection method cannot improve the input signal-to-noise ratio. The direct detection method has been widely used for its simplicity, easy implementation, high reliability, and low cost.

1.5.2 Direct Detection Limit

The total output power of all noises in the direct detection system is

$$P_{no} = [i_{NS}^2 + i_{NB}^2 + i_{ND}^2 + i_{NT}^2] \cdot R_L, \quad (1.35)$$

where $i_{NS}^2, i_{NB}^2, i_{ND}^2$ are shot noises caused by signal light, background light, and dark current, respectively, and i_{NT}^2 is the sum of the load resistance and the thermal noise of the amplifier. Then, the system output signal-to-noise ratio is

$$(\text{SNR})_p = \frac{(e\eta/h\nu)^2 P_s^2}{i_{NS}^2 + i_{NB}^2 + i_{ND}^2 + i_{NT}^2}. \quad (1.36)$$

When thermal noise is the main noise source of the direct detection system, the system is limited by the thermal noise, and the signal-to-noise ratio is

$$(\text{SNR})_{\text{thermal}} = \frac{(e\eta/h\nu)^2 P_s^2}{4KT \Delta f / R_L}. \quad (1.37)$$

When shot noise is much greater than thermal noise, the direct detection system is limited by the shot noise, and the signal-to-noise ratio is

$$(\text{SNR})_{\text{shot}} = \frac{(e\eta/h\nu)^2 P_s^2}{i_{NS}^2 + i_{NB}^2 + i_{ND}^2}. \quad (1.38)$$

When background noise is the main noise source of the direct detection system, the system is limited by the background noise, and the signal-to-noise ratio is

$$(\text{SNR})_{\text{back}} = \frac{\eta P_s^2}{2h\nu \Delta f P_B}. \quad (1.39)$$

When the shot noise caused by the incident signal light is the main noise source of the direct detection system, the system is limited by the signal noise, and the signal-to-noise ratio is

$$(\text{SNR})_{\text{limit}} = \frac{\eta P_s}{2h\nu\Delta f}, \quad (1.40)$$

which is known as *the quantum limit sensitivity of direct detection*.

1.6 Optical Amplifier

As the light source is modulated directly, it is difficult to increase the modulation speed by increasing the light source power; however, it is much easier by decreasing the light source power at a cost of limited transmission distance due to weak signal. As a solution to this tradeoff, an optical amplifier is used to directly amplify the modulated optical signal to achieve the combination of high-power light source and high modulation rate. The optical amplifier is an active device that can increase the power of the optical signal while keeping its characteristics unchanged.

1.6.1 Classification of Optical Amplifiers

Optical amplifiers can be divided into two categories based on their different working principles, as shown in Fig. 1.26. These are semiconductor optical amplifiers (SOAs) and fiber amplifiers; fiber amplifiers include erbium-doped fiber amplifier (EDFA) and praseodymium-doped fiber amplifier.

The principle of SOA is the same as that of semiconductor lasers. The phenomenon of population inversion caused by excited transitions between energy levels is used for optical amplification. There are two types of SOAs; one is the FP semiconductor laser amplifier, which uses ordinary semiconductor lasers as optical amplifiers. The other type uses coating of the two end faces of FP lasers with anti-reflection films to eliminate reflections to obtain a broadband output with high output power and low noise. The semiconductor optical amplifier has a gain of more than 30 dB and is

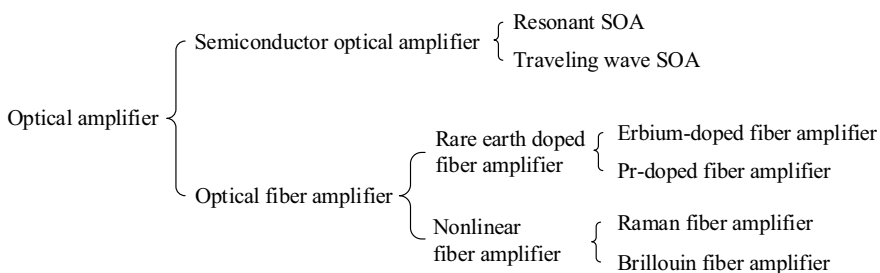


Fig. 1.26 Classification of optical amplifiers

applicable in both windows, at 1310 and 1550 nm. It is a feasible option for optical amplification if the gain can be maintained flat in the corresponding wavelength range.

Because it fully utilizes the existing semiconductor laser technologies, SOA has several advantages, such as mature production process, simple structure, small volume, low power consumption, low cost, long life, and easy integration with other optical devices. Moreover, the working wavelength of SOA covers 1.3–1.6 μm , which is infeasible for EDFA or PDFA. However, the major drawbacks of SOA are the large coupling loss with the optical fiber, sensitivity to ambient temperature, and large noise and crosstalk. Thus, its stability is poor.

Optical fiber amplifier is a new all-optical amplifier for signal amplification. Unlike the traditional SOAs, optical fiber amplifiers provide a “transparent” amplification by amplifying the signal directly in an all-optical fashion without complicated processes such as photoelectric and electrical-optical conversion or signal regeneration.

1.6.2 Erbium-Doped Fiber Amplifier

1.6.2.1 Overview of EDFA

EDFA is formed by erbium-doped fiber with a pump source. The working wavelength band of EDFA is 1530–1560 nm. EDFA can directly amplify optical signals with several advantages, such as high gain, high output power, low noise, fast response, and no requirements for signal encoding formats.

1.6.2.2 Basic Structure of EDFA

EDFA is primarily composed of erbium-doped fiber (EDF), pump light source, optical coupler, isolator, and optical filter.

The structure of EDFA has the following three forms:

- (1) Co-directional pumping: In the co-directional pumping scheme, the pump light and signal light are injected into the EDF from the same end.
- (2) Reverse pumping: The pump light and signal light are input to the doped fiber from different directions, and the two are transmitted in the opposite directions in the EDF.
- (3) Bidirectional pumping: Two pump light sources are used to inject pump light from both ends of the EDF at the same time. For the erbium ions in the EDF to be fully excited, the pump power must be increased.

1.6.2.3 Amplification Principle of EDFA

The basic working principle of EDFA is as follows. With the action of the pump source, the working substance particles transition from a low energy level to a high energy level (usually through another auxiliary energy level). Under a certain pump intensity, the population inversion distribution is obtained, and the light amplification effect is then generated. When the signal light in the working wavelength band is input, it is amplified.

As illustrated in Fig. 1.27, the erbium ions in EDF have three energy levels, where energy level 1 represents the ground state with the lowest energy; energy level 2 is a metastable state which is at an intermediate energy level; and energy level 3 represents the excited state with the highest energy.

1.6.3 Semiconductor Optical Amplifier

Generally, SOA refers to a traveling wave optical amplifier, whose working principle is similar to that of a semiconductor laser. It has a wide working bandwidth and a small gain. SOA is a type of optoelectronic device, in which the semiconductor gain medium under the condition of population inversion generates stimulated radiation and amplifies incident photons. SOA has similar advantages as semiconductor lasers, such as small size, high efficiency, low power consumption, and easy integration with other optoelectronic devices.

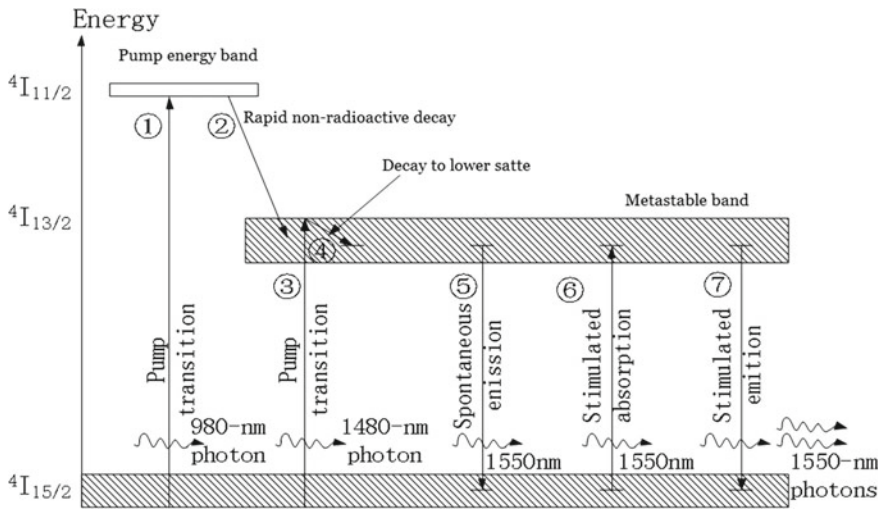


Fig. 1.27 Three-level structure of an erbium atom

1.7 Spatial Light to Fiber Coupling Technology

The commonly used spatial light to fiber coupling methods are single lens coupling, lens group coupling, fiber micro lens coupling, fiber array coupling, and special fiber coupling. Coupling the spatial light into the optical fiber facilitates the application of mature optical fiber communication technology to wireless laser communication.

1.7.1 Single Lens Coupling

The simplest system uses single lens for coupling spatial light into fiber, as shown in Fig. 1.28.

In the wireless laser communication system, the signal light can be regarded as a plane wave incident on the surface of a single lens after a long-distance transmission. Under the influence of turbulence, a convergent light field with randomly distributed wavefront is formed on the back focal plane of the lens, and the optical fiber is placed at the focal plane of the lens to couple the signal light. From the perspective of mode field matching, coupling is a process of matching between the optical field distribution $U(r)$ of the fiber end face and the mode field distribution $\psi(r)$ of the fiber end face. The coupling efficiency η is defined as the ratio between the average optical power coupled into the fiber $\langle P_T \rangle$ and the average optical power in the receiving plane $\langle P_a \rangle$:

$$\eta = \frac{\langle P_T \rangle}{\langle P_a \rangle} = \frac{\left\langle \left| \int_A U_i(\mathbf{r}) U_m^*(\mathbf{r}) dr \right|^2 \right\rangle}{\left\langle \int_A |U_i(\mathbf{r})|^2 dr \right\rangle}, \quad (1.41)$$

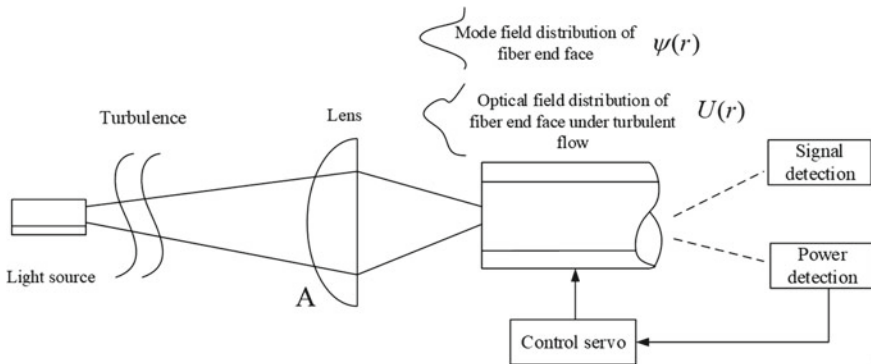


Fig. 1.28 Schematic diagram of coupling spatial light into a fiber with single lens

where $U_i(\mathbf{r})$ is the incident light field distribution on the receiving aperture A , and $U_m^*(\mathbf{r})$ is the fiber mode field distribution converted to the receiving aperture A . Since the optical system has the aperture limitation, the Airy disk is formed at the focal point of the lens, and its optical field distribution is different from the Gaussian distribution of the single-mode fiber mode field. Therefore, even if the wavefront distortion, jitter, and flicker caused by the turbulence are not considered, the theoretical maximum coupling efficiency of space light to single-mode fiber is 81% [14].

Since the light transmission in atmosphere is considerably affected by turbulence, the expression of the coupling efficiency, taking into account the turbulence effects, can use the cross-correlation light function $\Gamma_i(\mathbf{r}_1, \mathbf{r}_2)$ of the incident light field to extend Eq. (1.42) as

$$\eta_T = \frac{1}{\langle P_a \rangle} \iint_A \Gamma_i(\mathbf{r}_1, \mathbf{r}_2) U_m^*(\mathbf{r}_1) U_m(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2. \quad (1.42)$$

Here, the cross-correlation light function $\Gamma_i(\mathbf{r}_1, \mathbf{r}_2)$ of the incident light field uses the weak turbulence cross-correlation function under the Kolmogorov refractive index spectral density, and the cross-correlation function of the incident light field can be expressed using Gaussian approximation [15]:

$$\Gamma_i(\mathbf{r}_1, \mathbf{r}_2) = I_i \exp\left(-\frac{|\mathbf{r}_1 - \mathbf{r}_2|^2}{\rho_c^2}\right), \quad (1.43)$$

where I_i is the intensity of the incident light field, $\rho_c = (1.46C_n^2 k^2 L)^{-3/5}$ is the spatial coherence length, C_n^2 is the atmospheric refractive index structure constant, and L is the communication distance.

As shown in Fig. 1.28, in practice, the coupling fiber and the control servo (including displacement and rotation) are often fixed together to form a closed-loop control with the power detection at the back end to overcome the unstable coupling power caused by turbulence. To address the limited shaping effect of a single lens on the spatial light, a multi-lens system (i.e., lens group coupling), such as the Cassegrain reentrant system shown in Fig. 1.29, can be used to shape the spatial

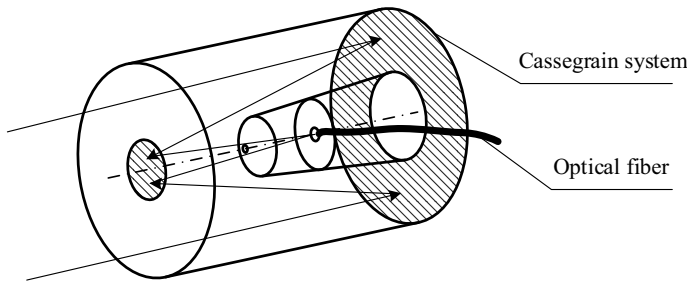


Fig. 1.29 Schematic diagram of spatial light-lens coupling

light. Furthermore, an aspheric lens can be used to focus plane light field distribution to make adjustments to the Airy disk with more concentrated energy; this is called the aspheric lens coupling system.

1.7.2 Array Coupling

As shown in Fig. 1.30, the light irradiated on the N effective array units is respectively coupled into N discrete optical fibers of equal length, and then the fusion taper technology is used to synthesize the optical fibers.

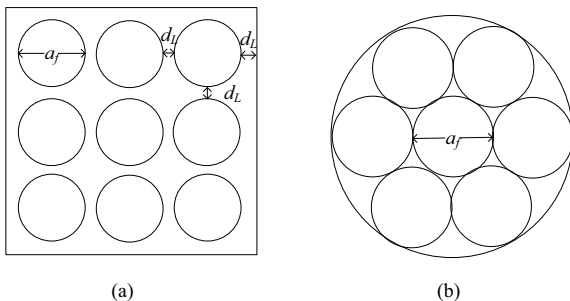
The incident light field changes with the communication distance and turbulence. When a light spot completely covers the lens and lens array, for lens and lens array with the same area, the effective unit in the array has a smaller aperture, which has a strong inhibiting effect on turbulence. When the communication distance exceeds a certain range, the coupling efficiency of an array will be significantly higher than that of lens. By increasing the area of the array to suppress the spot jitter caused by turbulence, the spatial light-fiber coupling of the array structure is suitable for long-distance free space optical communication systems.

1.7.3 Special Fiber Coupling

There are many types of special optical fibers. Tapered fibers, self-focusing lenses, and fibers with larger core diameters are often used in spatial light-fiber coupling systems. Among them, the conical fiber refers to a special type of fiber whose diameter changes linearly with the length of the fiber. Figure 1.31 shows the light path, where the meridian light enters the small end from the large end of the light cone, and δ is the taper angle of the tapered fiber. Using the principle of total reflection, the tapered fiber can couple the spatial light of the incident light into the tapered fiber.

According to the condition of total reflection, if the incident light can be emitted from the other end of the fiber, the fiber length must satisfy the following formula:

Fig. 1.30 Arrangement of lens array (a) rectangular array and (b) circular array



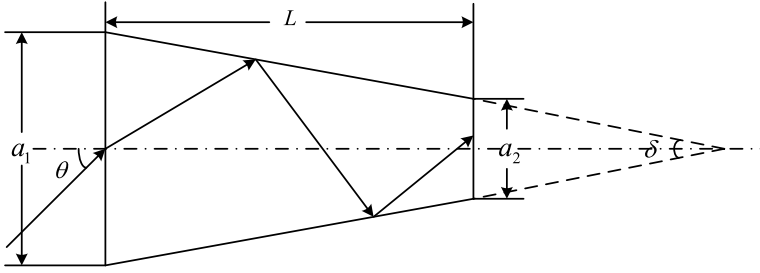


Fig. 1.31 Tapered fiber coupling

$$l \geq \frac{1}{2} \frac{2(a_2 - a_1) \cos \theta}{\left[\frac{a_1}{a_2} \left[1 - \left(\frac{n_2}{n_1} \right)^2 \right]^{\frac{1}{2}} - \sin \theta \right]}, \quad (1.44)$$

where a_2 is the radius of the small mouth, a_1 is the radius of the large end, n_2 is the refractive index of the fiber sheath, n_1 is the refractive index of the core, and θ is the incident angle. Equation (1.44) shows that to condense the light of the tapered fiber, the fiber length must exceed a certain value to achieve the coupling effect.

1.8 Optical Antenna and Telescope

An optical antenna is essentially an optical telescope. Optical telescopes comprise the optics and the mechanical device. The optics consists primarily of the objective lens and eyepiece of the telescope. Based on the different objective lens, telescopes can be divided into refracting and reflecting telescopes. An objective lens, mirror, and lens to correct the phase difference in front of it comprise a catadioptric telescope.

1.8.1 Refractor Telescope

Galilean telescope uses a lens to gather and converge light at one point to produce an enlarged image. Since the principle that light bends when passing through glass (lens refraction) is applied in a Galilean telescope, as shown in Fig. 1.32, it is also known as refracting telescope.

Compared with reflecting telescopes, the refracting telescope has a better image quality and a larger field of view. Refracting systems are mostly adopted by small and medium-sized astronomical telescopes and used in many special instruments.

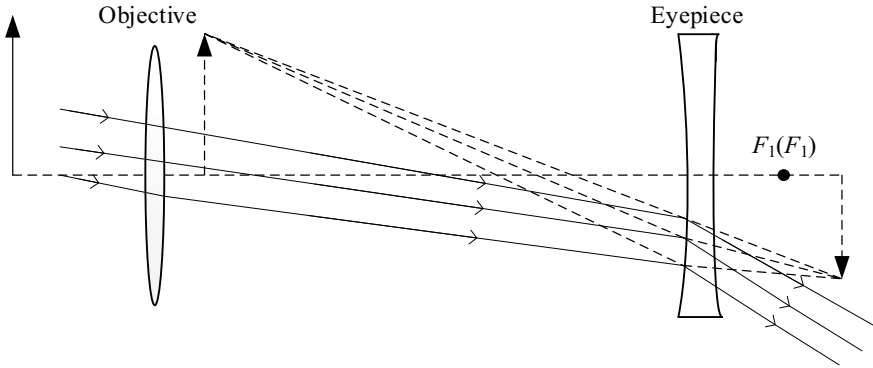


Fig. 1.32 Galilean telescope

1.8.2 Reflecting Telescope

The objective lens of a reflecting telescope is composed of a reflecting mirror (concave surface). Its most important feature in terms of optical performance is that it has no chromatic aberration. When a parabolic surface is adopted in objective lens, spherical aberration can also be eliminated. However, to reduce the influence of other aberrations, the available field of view is small.

For reflecting telescopes, the commonly used objective lens systems include Newtonian system, main focus system, and Cassegrain system. The first two have only one curved mirror, called a simple objective lens system, whereas the others have two or more curved mirrors, called complex objective lens systems. The most famous reflection objective lens systems are the double reflection objective lens systems, namely Cassegrain and Gregorian systems.

1.8.2.1 Cassegrain System

The Cassegrain system is composed of two mirrors, as shown in Fig. 1.33, with the primary mirror parabolic and the secondary mirror hyperboloid. The resulting image is inverted, and the tube length of this structure is relatively short.

Fig. 1.33 Cassegrain system

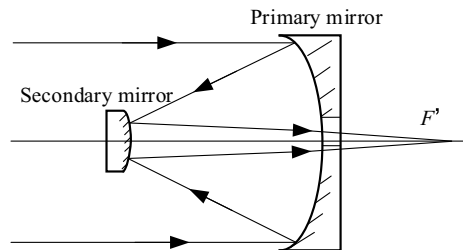
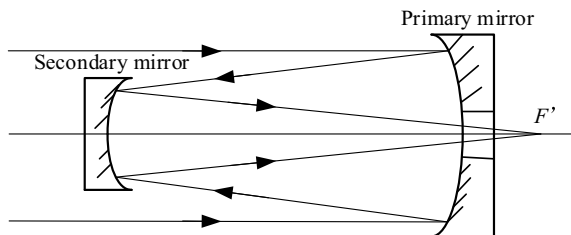


Fig. 1.34 Gregorian system

1.8.2.2 Gregorian System

The Gregorian system is also composed of two reflecting surfaces, as shown in Fig. 1.34. The primary mirror is a parabolic surface, and the secondary mirror is replaced by an ellipsoidal surface, resulting in a positive image. This structure has a relatively long tube.

1.8.3 Catadioptric Telescope

A telescope that uses an objective lens system in which light passes through a lens and then forms images by a mirror is called a catadioptric telescope. There are two basic types of catadioptric telescopes: Schmidt telescope and Maksutov telescope. Based on these two types, catadioptric telescopes of the Becker, Maksutov-Cassegrain, and super Schmidt systems are further introduced. In catadioptric telescopes, the image is formed by a spherical mirror, and the refractor, which is used to correct spherical aberration, is also called a correcting mirror.

1.8.3.1 Schmidt Telescope

The Schmidt telescope is composed of a spherical primary mirror and a Schmidt correcting mirror, as shown in Fig. 1.35. The correcting mirror is a transmission element, with a flat surface on one side and an aspheric surface on the other.

1.8.3.2 Maksutov Telescope

The Maksutov telescope consists of a spherical primary mirror and a negative meniscus thick lens, as shown in Fig. 1.36. Since the catadioptric telescope has a large field of view and relative aperture, it has a wide range of applications in astronomical observations.

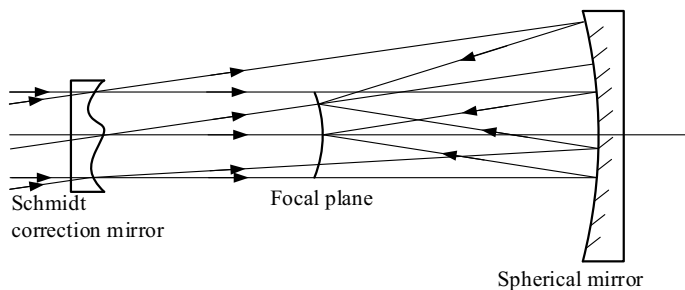
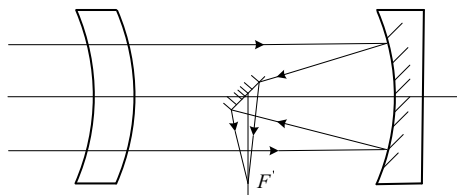


Fig. 1.35 Catadioptric telescope

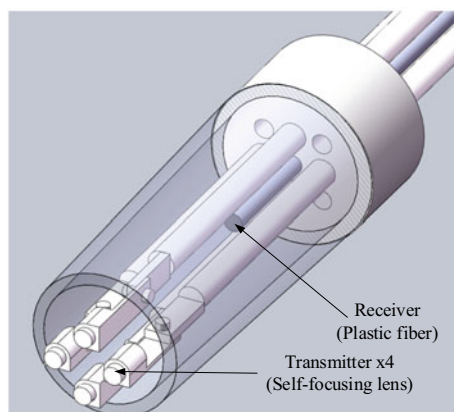
Fig. 1.36 Maksutov telescope



1.8.4 Integrated Transceiver Optical Antenna

Blind areas in the reflective secondary mirror of the Maksutov antenna leads to a decreased transmission efficiency. After the receiving and transmitting optical paths are split by the beam splitter, a loss of efficiency occurs. Figure 1.37 illustrates an integrated transceiver optical antenna, which uses fiber-coupled array for transmission and fiber coupling for reception. Figure 1.38 shows the distribution diagram of light intensity emitted by this antenna.

Fig. 1.37 Integrated transceiver optical antenna based on Maksutov antenna



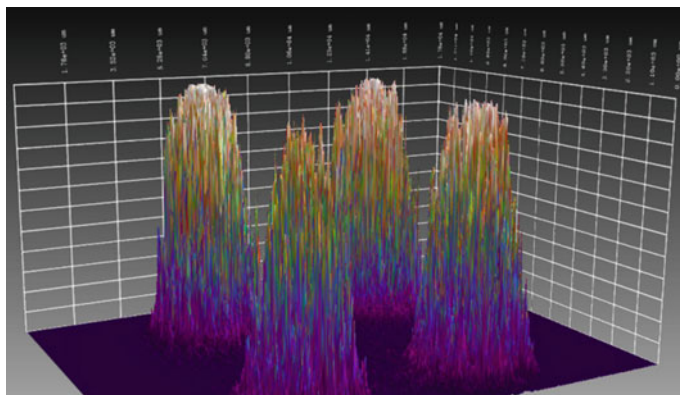


Fig. 1.38 Distribution of light intensity of integrated transceiver optical antenna

1.9 Summary and Prospects

This chapter provides a brief introduction to the optical wireless communication system. The optical wireless communication system is composed of a transmitter, receiver, and channel. The current optical wireless communication is primarily point-to-point communication. By increasing the transmission power to increase the transmission distance, the transmitting antenna is designed based on the existing optical telescope. Future development directions include (1) investigation of point-to-multipoint and multipoint-to-multipoint optical communication mechanisms; (2) research on new photodetectors, especially with high sensitivity and near the quantum limit; (3) research on the aspheric optical antenna; and (4) integrated design optical-mechanical-electrical systems with low cost and high reliability.

1.10 Questions

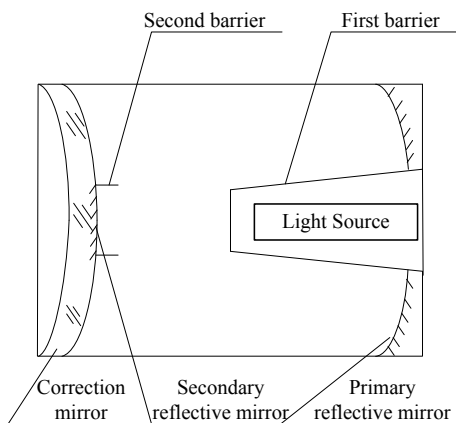
- 1.1 What are the components of the optical wireless communication system? What are their functions?
- 1.2 What are the similarities and differences between LED and LD? How to correct the nonlinear characteristics of LED and LD?
- 1.3 What is the principle of photodetection?
- 1.4 How many types of optical modulation technologies can be divided? What are the characteristics of each?
- 1.5 What are the advantages of passive modulation?
- 1.6 What are the types of optical amplifiers? Briefly describe its working principle.
- 1.7 What is the function of spatial light-fiber coupling? What are the types of coupling? Explain their working principles respectively.

- 1.8 How many types of optical antennas can be generally divided into?
- 1.9 Can direct detection improve the signal-to-noise ratio of the system? Try to analyze.

1.11 Exercises

- 1.1 As shown in Fig. 1.39, the fold-back Cassegrain system is widely used as a transmitting antenna in wireless laser communication because of its long focal length, small size, light weight and easy production. The Cassegrain telescope generally adopts a double reflection design, which is composed of a parabolic primary mirror, a hyperboloid secondary mirror and a correction mirror. The double-reflective structure of the system determines that the Cassegrain system has serious optical energy loss when used as a transmitting antenna. In order to solve the loss problem, an off-axis transmission scheme can be used.
 - (1) Please draw the receiving and luminous circuit diagram of the Cassegrain system, and try to analyze the cause of the loss.
 - (2) Please explain why off-axis emission can solve the problem of emission loss and the drawbacks of this scheme, and draw the light path diagram of off-axis emission.
- 1.2 Consider a Gaussian beam transmission [16] with spot radius $\omega(z) = \omega_0 \sqrt{1 + (\lambda z / \pi \omega_0^2)^2}$ at point z , where ω is the waist radius, λ is the wavelength, z is the transmission distance. The light intensity at z is expressed as $I(z, \rho) = I_0 \frac{\omega_0^2}{\omega^2(z)} \exp[-2\rho^2 / \omega^2(z)]$, and the optical power received by the antenna can be expressed as $P_S(z) = \pi \eta_2 R_2^2 I(z) = \frac{2\eta_1 \eta_2 R_2^2 P_1}{0.865 \omega^2(z)}$, where η_2 is the receiving antenna efficiency, and R_2 is the receiving antenna radius. The

Fig. 1.39 Fold-back Cassegrain system



bit error rate of the system can be expressed as

$$\begin{aligned}\eta_{BER} &= \frac{1}{\sqrt{\pi}} \int_{s2\sqrt{2}\sigma}^{\infty} e^{-x^2} dx = \frac{1}{2} \operatorname{erfc} \frac{\langle i_s \rangle(z)}{2\sqrt{2}\sigma(z)} \\ &= \frac{1}{2} \operatorname{erfc} \left[\frac{1}{4} \sqrt{\frac{2S_i \eta_1 \eta_2 R_2^2 P_1 \pi^2 \omega_0^2}{0.865 \Delta f (\pi^2 \omega_0^4 + \lambda^2 z^2)}} \right]\end{aligned}$$

From the above formula, it is easy to know that under the condition of ignoring the influence of atmospheric factors, the bit error rate of the communication system is inversely proportional to the transmitted optical power, the transmitting aperture radius, the receiving aperture radius, the efficiency of the transceiver antenna, and the sensitivity of the detector, and proportional to the communication distance and measurement bandwidth, namely $\eta_{BER} = \eta_{BER}(S_i, R_2, P_1, \omega_0, \eta_1, \eta_2, z, \Delta f)$.

- (1) If $P_1 = 30 \text{ mW}$, $\lambda = 1.55 \text{ m}$, $\omega_0 = 15 \text{ mm}$, $\eta_1 = \eta_2 = 0.7$, $R_2 = 50 \text{ mm}$, $\Delta f = 3 \text{ GHz}$, find the bit error rate;
- (2) If $P_1 = 3 \text{ mW}$ and other parameters are the same as above, find the limit communication distance.

- 1.3 The Cassegrain antenna has two features: (1) It has a large aperture, does not produce chromatic aberration and has a wide band range; (2) When an aspheric lens is used, the aberration ability is strong. Therefore, the Cassegrain antenna is often used as an optical transmitting and receiving antenna in laser space communication [17]. Assuming that the beam transmitted by the transmitting antenna is a Gaussian beam in fundamental mode, its distribution can be expressed by the following formula

$$E_0(r_0) = \sqrt{\frac{2}{\pi \omega^2}} \exp\left(-\frac{r_0^2}{\omega^2}\right) \exp\left(\frac{ikr_0^2}{2R}\right)$$

where ω is the waist size and R is the curvature of the wave surface at the antenna. the intensity distribution at the observation point (r, θ) is

$$I(r, \theta) = \frac{k^2}{r^2} \times \left| \int_b^a \sqrt{\frac{2}{\pi \omega^2}} \exp\left(-\frac{r_0^2}{\omega^2}\right) \exp\left(ik \frac{r_0^2}{2} \left(\frac{1}{r} + \frac{1}{R}\right)\right) J_0(kr_0 \sin \theta) r_0 dr_0 \right|^2$$

where a is the radius of the antenna's primary mirror and b is the radius of the secondary mirror. The gain of the optical antenna is

$$G(r, \theta) = \frac{I(r, \theta)}{I_0} = \frac{8k^2}{\omega^2} \times \left| \frac{\int_b^a \exp\left(-\frac{r_0^2}{\omega^2}\right) \exp\left(ik \frac{r_0^2}{2} \left(\frac{1}{r} + \frac{1}{R}\right)\right)}{J_0(kr_0 \sin \theta) r_0 dr_0} \right|^2$$

where $I_0 = 1/(4\pi\gamma^2)$. Assuming that $\alpha = a/\omega$, $\gamma = b/a$ and $X = ka \sin \theta$, then

$$\beta = (ka^2/2) \left(\frac{1}{r} + \frac{1}{R} \right)$$

where α represents the ratio of the aperture of the antenna's primary mirror to the laser beams light waist, γ is the occlusion ratio, and X represents the pointing angle factor of the optical antenna. Therefore, the antenna gain can be written as

$$G_T(\alpha, \beta, \gamma, X) = (4\pi A/\lambda^2) g_T(\alpha, \beta, \gamma, X)$$

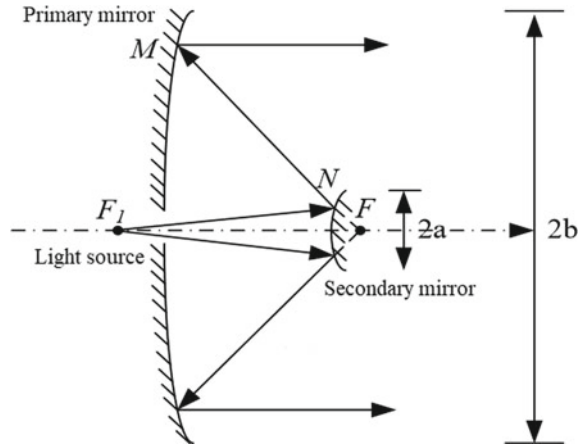
where $A = \pi a^2$ and g_T is the antenna gain efficiency factor. The available antenna gain efficiency factor is

$$g_T(\alpha, \beta, \gamma, X) = 2\alpha^2 \times \left| \int_{r^2}^1 \exp(i\beta u) \exp(-\alpha^2 u) J_0(Xu^{1/2}) du \right|^2$$

Try to discuss the maximum gain of the antenna and its conditions.

- 1.4 A Cassegrain antenna system is shown in Fig. 1.40 [18]. The antenna gain factor is

Fig. 1.40 Cassegrain antenna system



$$g_T(\alpha, \beta, \gamma, X) = 2\alpha^2 \times \left| \int_{r^2}^1 \exp(i\beta u) \exp(-\alpha^2 u) J_0(Xu^{1/2}) du \right|^2$$

where $\alpha = a/\omega$, $\gamma = b/a$, $X = ka \sin \theta$, $\beta = (ka^2/2)(\frac{1}{r} + \frac{1}{R})$, and R is the radius of curvature of the light wave front, and k is the wave number of the light wave. For an on-axis point with $X = ka \sin \theta = 0$, the antenna main axis gain factor is

$$g_T(\alpha, \beta, \gamma, 0) = \left(\frac{2\alpha^2}{\beta^2 + \alpha^4} \right) \times \left\{ \begin{array}{l} \exp(-2\alpha^2) + \exp(-2\alpha^2\gamma^2) \\ -2\exp[-\alpha^2(\gamma^2 + 1)] \cos[\beta(\gamma^2 - 1)] \end{array} \right\}$$

- (1) Find the relationship between the optimal shielding ratio and the aperture of the optical antenna system.
 - (2) Find the divergence angle of the far field of the beam.
- 1.5 Design the transmitting antenna of the optical wireless communication system [19]. Using Cassegrain optical antenna, the focal length of the system is 500 mm, the back intercept is 60 mm, the main lens aperture is 100 mm, and the system field of view is 1 mrad. Try to determine the system blocking rate, secondary mirror aperture value, primary mirror apex radius of curvature, secondary mirror apex radius of curvature, primary mirror the eccentricity of the secondary mirror, the eccentricity of the secondary mirror, the hollow diameter of the primary mirror, and the distance between the vertices of the two mirrors.
 - 1.6 Analyze the coupling efficiency of the spatial light-fiber coupling system when there is lateral movement under the condition of small hole coupling [20].
 - 1.7 A space optical coupling system [21]. The diameter of the receiving antenna is 150 mm, the spot diameter is less than 1.5 mm, the diameter of the image square hole is no more than 12.79° , and the working wavelength is 400–660 nm. Try to analyze the coupling efficiency.
 - 1.8 Assuming that the diameter of the collimated beam at the transmitting end is 10 mm, the transmission distance is 1 km, the spot radius at the receiving end is 0.05 m, and the mode field diameter is 9 m, try to analyze the influence of different atmospheric turbulence on the coupling efficiency [22].
 - 1.9 The bit error rate of the optical receiver is given by the Q function [23]

$$BER = \int_{\gamma} \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw$$

where γ is the signal-to-noise ratio output by the receiver $\gamma = \frac{m_1 - m_0}{\sigma_1 - \sigma_0}$, m_1 represents the signal level at logic “1”, m_0 represents the signal level at logic

“0”, σ_1 represents the noise level at logic “1”, and σ_2 represents logic The noise level at “0”.

- (1) If the influence of thermal noise is considered, find the expression of the bit error rate of the system.
 - (2) If the code interference is considered, find the expression of the system error rate.
- 1.10 Find the relationship between the divergence angle of the output beam of the semiconductor laser and the efficiency of light passing through the small hole [24].
 - 1.11 In the satellite optical communication system, in order to obtain an extremely narrow beam divergence angle and reduce the volume of the optical terminal, a coaxial two-mirror reflecting telescope is generally used as an optical antenna. The major disadvantage of this structure is that the light energy loss caused by the secondary mirror shielding is relatively large, resulting in low emission efficiency of the optical communication terminal. In order to compensate this loss and improve the transmission efficiency, the traditional solution is to increase the output power of the transmitting laser, but this method will increase the power consumption of the terminal, which will put pressure on the energy distribution of the satellite platform. As shown in Fig. 1.41, the phase recovery algorithm is used to design a diffractive optical element that can eliminate the occlusion of the secondary mirror [25], and the energy distribution of the output beam before and after the design in the far field is compared under different occlusion ratios and cut-off ratios.
 - 1.12 Try to analyze the effects of spherical aberration, coma, astigmatism, curvature of field and distortion of the Cassegrain antenna on the beam quality [26].
 - 1.13 Try to analyze the influence of Cassegrain antenna secondary mirror blocking on antenna gain and antenna transmission efficiency [27].
 - 1.14 The signal-to-noise ratio of the direct detection system can be expressed as

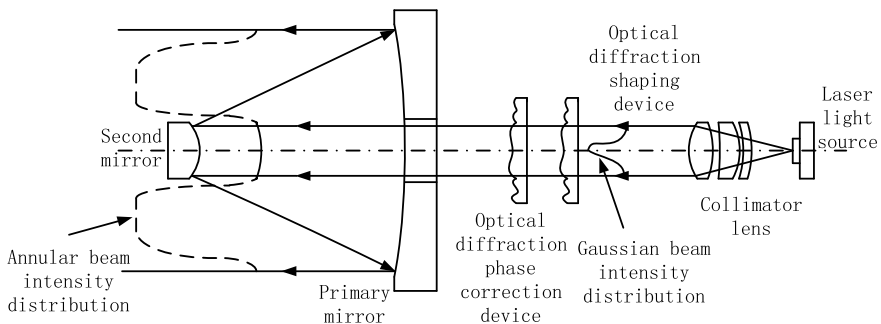


Fig. 1.41 Coaxial two-mirror reflecting telescope

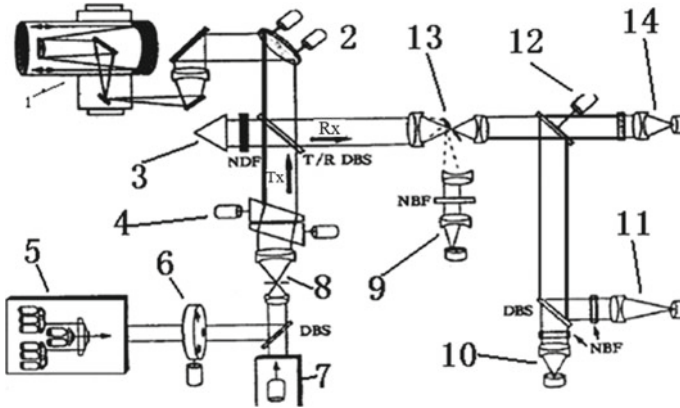


Fig. 1.42 Integrated transceiver antenna with an aiming advance unit

$$(SNR) = \frac{\left(\frac{e\eta}{h\nu}\right)^2 P_s^2}{i_{NS}^2 + i_{NB}^2 + i_{ND}^2 + i_{NT}^2}$$

where i_{NS}^2 , i_{NB}^2 , i_{ND}^2 are the signal light, background light and the shot noise caused by the dark current, respectively. i_{NT}^2 is the sum of the load resistance and the thermal noise of the amplifier. Discuss the influence of thermal noise, shot noise, and background noise on the signal-to-noise ratio, and analyze the detection limits in different situations.

- 1.15 A wireless laser communication system is shown in the Fig. 1.42. This is an integrated transceiver antenna with an aiming advance unit, which is suitable for satellite-borne laser communication. The wedge mirror is used to correct the beam pointing and control the advance at different speeds. This system includes a beacon laser (7) to indicate the optical path for reception.

1-Universal joint and telescope; 2-two-axis rotating mirror; 3-angle mirror; 4-pointing advance mechanism; 5-beam combiner; 6-light divergence switch; 7-beacon laser; 8-aperture; 9-searching detector; 10-tracking detector; 11-pointing advancement mechanism; 12-search/communication switch; 13-split mirror; 14-communication detector.

Please describe how the system works.

- 1.16 Suppose an integrated transceiver antenna of the Maksutov telescope shown in Fig. 1.43. The Maksutov antenna is used as the antenna, and the transmitting beam is separated from the receiving beam by the reflector (B). The system has a certain energy loss. The Maksutov antenna also affects the received light energy. Please describe the working principle of the system.

17. He W, Yang H, Jiang P (2014) Research on transmission characteristics of cassegrain optical antenna. *Laser Infrared* 44(3):280–284
18. Ma X, Wang B, Yang H et al (2014) Optimal design of Cassegrain optical antenna system. *Laser Infrared* 44(4):410–413
19. Li X, Chen Z, Gu L (2009) Design of Cassegrain optical receiving antenna[J]. *Commun Technol* 03:38–39+42
20. Chen X, Zhang L, Wu Z (2013) Coupling efficiency of space laser with single-mode fiber and photonic crystal fiber[J]. *Chin Optics* 02:208–215
21. Wei D, Wang F, Fang D, Wang X (2013) Sunlight fiber coupling optical system design[J]. *Appl Opt* 02:220–224
22. Xiong Z, Ai Y, Shan X et al (2013) Optical fiber coupling efficiency and compensation analysis for space optical communication[J]. *Infrared Laser Eng* 09:2510–2514
23. Yao J, Yang S (1993) Noise analysis of optical amplifier and sensitivity calculation of optical preamplifier receiver[J]. *Acta Optics* 07:611–618
24. He J, Li X (2006) Power coupling efficiency of semiconductor laser beam collimation system[J]. *Appl Opt* 01:51–53
25. Yu J, Tan L, Ma J et al (2009) A new method to improve the transmission efficiency of satellite optical communication terminals[J]. *China Laser* 03:581–586
26. Zhang Y, Ai Y (2005) Discussion on the drawbacks of Cassegrain antennas based on space optical communication[J]. *Infrared Laser Eng* 05:60–63
27. Ma X, Wang B, Yang H et al (2014) Optimal design of Cassegrain optical antenna system [J]. *Laser Infrared* 04:410–413

Chapter 2

Coherent Optical Communication



Compared to intensity modulation/direct detection (IM/DD), coherent optical communication systems can achieve a detection sensitivity gain of approximately 20 dB (homodyne detection can reach 23 dB), allowing for longer distance transmission under the same power. Furthermore, coherent optical communication systems are more conducive to the suppression of atmospheric turbulence and channel fading. This chapter introduces the concepts behind coherent optical communication.

2.1 Basic Principles of Coherent Optical Communication

2.1.1 Fundamentals

The coherent optical communication system is summarized in Fig. 2.1. Light emitted from the laser is the carrier, and the signal is modulated using direct or external modulation (amplitude, frequency, phase, or polarization). At the receiving antenna, the spatial optical signal is coupled into the optic fiber through the fiber coupler; then, the signal is generated by the local oscillator laser and the light wave is superimposed with the received signal in the mixer. An intermediate frequency (IF) signal is generated from the balanced detector output, and the baseband signal can be obtained directly during homodyne detection.

In Fig. 2.2, the modulated signal is transmitted to the receiving end and mixed with the optical signal generated by the local oscillator laser; then, the IF signal is obtained and sent to the balance detector for detection. After the electrical signal output from the balance detector is amplified, filtered, and demodulated using the IF, it is restored to the digital signal at the transmitting end. The electric fields of the signal light (E_S) and the local oscillator light (E_L) can be expressed as follows:

$$E_S(t) = A_S(t) \cos(\omega_s t + \varphi_s) \quad (2.1)$$

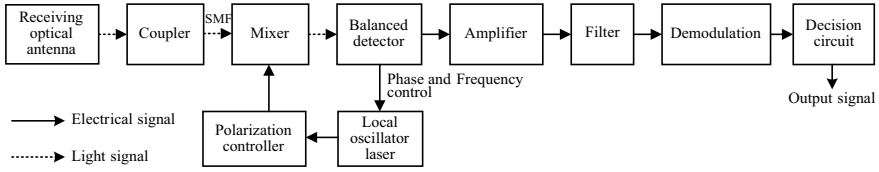


Fig. 2.1 Flowchart of the coherent optical communication system

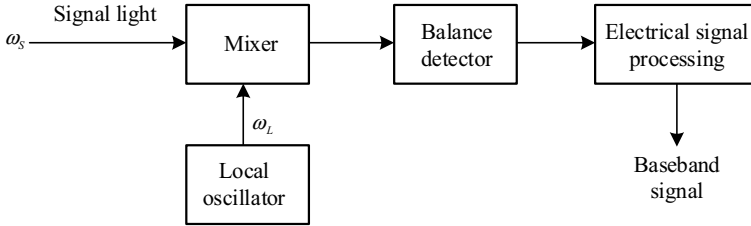


Fig. 2.2 Diagram of coherent detection [1]

$$E_L(t) = A_L \cos(\omega_L t + \varphi_L) \quad (2.2)$$

where A_L , ω_L , and φ_L are the amplitude, central angular frequency, and phase of the local oscillator light, respectively, and A_S , ω_S , and φ_S are the amplitude, central angular frequency, and phase of the signal light, respectively. Assuming that the signal and local oscillator lights have the same polarization direction and are incident on the optical mixer in parallel, then the incident light power is $[E_S(t) + E_L(t)]^2$, and the IF current signal output (i_p) from the mixer can be expressed as

$$\begin{aligned} i_p &= \alpha P(t) = \alpha [E_S(t) + E_L(t)]^2 \\ &= \alpha \left\{ A_S^2 \cos^2(\omega_S t + \varphi_S) + A_L^2 \cos^2(\omega_L t + \varphi_L) \right. \\ &\quad + A_S A_L \cos[(\omega_S + \omega_L)t + (\varphi_S + \varphi_L)] \\ &\quad \left. + A_S A_L \cos[(\omega_S - \omega_L)t + (\varphi_S - \varphi_L)] \right\} \end{aligned} \quad (2.3)$$

In Eq. (2.3), $\alpha = en/h\nu$; the first and second terms are cosine squared terms, the time average value in the integer period is 1/2, and the sum is $(A_S^2 + A_L^2)/2$, equivalent to the direct current (DC) component of the detector output. The third term is the sum frequency term, which can be divided into two situations: (1) the time average value corresponding to the sum frequency term in the integer period is zero and (2) in a non-integer period, the frequency corresponding to the sum frequency term is very high and the photodetector cannot respond under normal circumstances. The fourth term is the difference frequency term, and because its rate of change is slower than that of the light field, it can be regarded as a constant. After filtering the

DC and sum-frequency terms through a band-pass filter, the IF photocurrent output (i_{IF}) from the photodetector can be expressed as [2]

$$i_{IF} = \alpha A_S A_L \cos[(\omega_S - \omega_L)t + (\varphi_S - \varphi_L)] \quad (2.4)$$

The frequency spectrum and waveform of heterodyne detection are shown in Fig. 2.3.

Using the average optical power, Eq. (2.4) can be expressed as [2]

$$i_{IF} = 2\alpha\sqrt{P_S P_L} \cos[(\omega_S - \omega_L)t + (\varphi_S - \varphi_L)] \quad (2.5)$$

Here, $P_S = A_S^2/2$ is the average optical power of the signal light and $P_L = A_L^2/2$ is the average optical power of the local oscillator light. When restoring the baseband signal, the received optical signal carrier frequency is transformed into a signal f_{IF} , and the IF signal is transformed into a baseband signal. This detection method is called heterodyne detection.

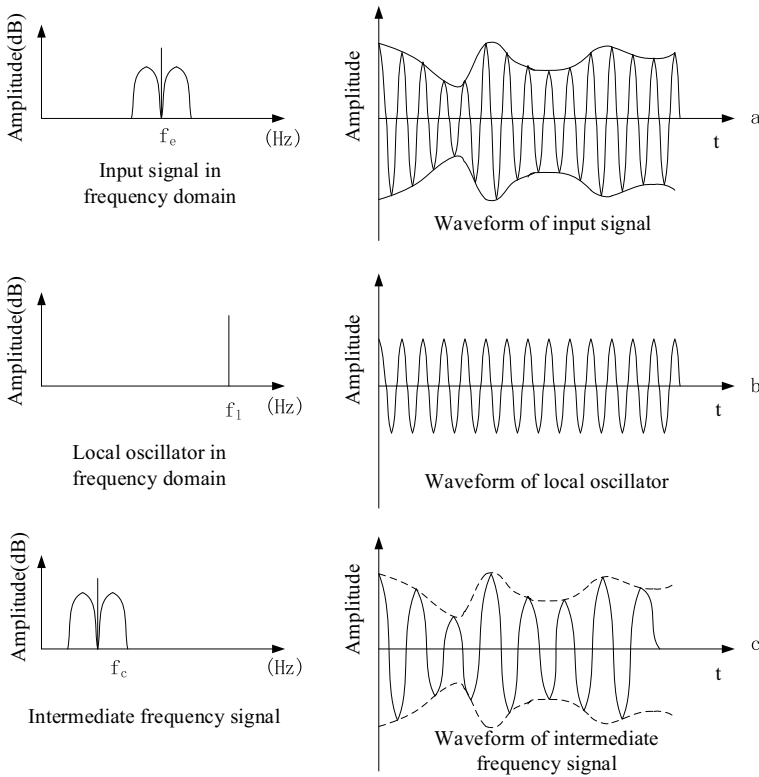


Fig. 2.3 Heterodyne detection frequency spectrum and waveform [3]

2.1.2 Homodyne Detection

When the local oscillator frequency ω_L is the same as the signal light frequency ω_S (at this time $\omega_{IF} = 0$), it is referred to as homodyne detection. From Eq. (2.5), the photocurrent output by the balanced detector can be expressed as [2]

$$i_{IF} = 2\alpha\sqrt{P_S P_L} \cos(\varphi_S - \varphi_L) \quad (2.6)$$

Considering that the optical phase of the local oscillator is locked on the signal optical phase ($\varphi_S = \varphi_L$), the output photocurrent during homodyne detection can be expressed as

$$i_{IF} = 2\alpha\sqrt{P_S P_L} \quad (2.7)$$

Other coherent optical communication systems exist. For example, Fig. 2.4 shows the principle diagram of auto-coherent detection using differential phase shift keying (DPSK) modulation. The received optical signal is coupled to the single-mode fiber through the reflective receiving antenna and enters the DPSK demodulator after being amplified by the erbium doped optical fiber amplifier. The signal is divided into two channels in the demodulator; and, with the exception of a 1-bit delay, the channels remain consistent. This information is then converted into intensity information, and the optical signal is converted into an electrical signal by the photodetector.

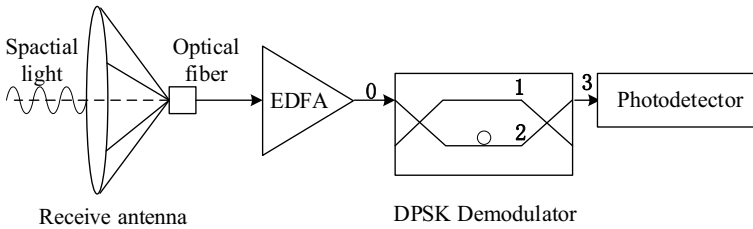


Fig. 2.4 Auto-coherent (self-homodyne) detection using differential phase-shift keying (DPSK) modulation. EDFA stands for erbium doped optical fiber amplifier [4]

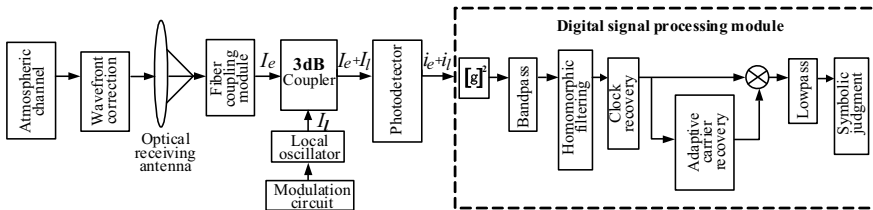


Fig. 2.5 Subcarrier intensity modulation heterodyne detection system [5]

Figure 2.5 shows the heterodyne detection system of subcarrier modulation. The square term is constructed in the signal processing unit. Owing to the local laser, the heterodyne gain of this scheme is in the range 18–23 dB [5].

2.1.3 Heterodyne Detection

In the case of heterodyne detection, ω_L differs from ω_S and the heterodyne current generated by the balanced detector can be expressed as [2]:

$$i_{IF} = 2\alpha\sqrt{P_S P_L} \cos[\omega_{IF}t + (\varphi_S - \varphi_L)] \quad (2.8)$$

Equation (2.8) shows that.

- (1) The output IF current is proportional to P_L . Even if the received optical signal power P_S is small, a sufficiently large output current can be obtained by increasing P_L . The local oscillator light is used for optical amplification in coherent detection, and sensitivity is improved because the system obtains a mixing gain.
- (2) Because $\omega_S - \omega_L$ remains constant (ω_{IF} or 0) in coherent detection, it is necessary for the signal and local oscillator light sources have high frequency stability, a narrow spectral width, and a specific frequency tuning range.
- (3) In both homodyne and heterodyne detection, the system must maintain phase lock and polarization direction matching between the received and local oscillator signal.

Optical heterodyne detection is holographic detection technology. The information carried by the amplitude E_S , frequency $\omega_S = \omega_{IF} + \omega_L$ (ω_L is known, ω_{IF} can be measured), and phase φ_S of the light field can be detected.

2.1.4 Detection of an Amplitude Modulated Signal

If a signal with a modulation frequency of Ω is loaded on the amplitude of a light wave with a frequency of ω_S , the amplitude-modulated light wave can be expressed as [6]

$$\begin{aligned}
E_S(t) &= A_0 \left[1 + \sum_{n=1}^{\infty} \cos(\Omega_n + \varphi_n) \right] \cos(\omega_S t + \varphi_S) \\
&= A_0 \cos(\omega_S t + \varphi_S) + \sum_{n=1}^{\infty} \frac{m_n A_0}{2} \cos[(\omega_S + \Omega_n)t + (\varphi_S - \varphi_n)] \\
&\quad + \sum_{n=1}^{\infty} \frac{m_n A_0}{2} \cos[(\omega_S - \Omega_n)t + (\varphi_S - \varphi_n)],
\end{aligned} \tag{2.9}$$

where A_0 is the average amplitude of the amplitude modulation (AM) wave, Ω_n and φ_n are the angular frequency and initial phase of the n -th harmonic component, respectively, and m_n is the AM coefficient. If the AM signal $E_S(t)$ is coherent with the local oscillator light $E_L(t)$, the instantaneous intermediate-frequency current can be expressed as [6]

$$\begin{aligned}
i_{IF} &= \alpha A_0 A_L \cos[(\omega_L - \omega_S)t + (\varphi_L - \varphi_S)] \\
&+ \alpha A_L \sum_{n=1}^{\infty} \frac{m_n A_0}{2} \cos[(\omega_L - \omega_S - \Omega_n)t + (\varphi_L - \varphi_S - \varphi_n)] \\
&+ \alpha A_L \sum_{n=1}^{\infty} \frac{m_n A_0}{2} \cos[(\omega_L - \omega_S + \Omega_n)t + (\varphi_L - \varphi_S + \varphi_n)],
\end{aligned} \tag{2.10}$$

The signal converted by the photodetector is proportional to the instantaneous IF current. The modulated signal carried on the amplitude of the light wave is transferred to the frequency current without $\omega_{IF} = \omega_L - \omega_S$ distortion (frequency and phase modulated coherent detection use a similar process). Figures 2.6 and 2.7 show the instantaneous IF electrical signal spectrum following AM signal light and coherent detection.

For AM signals, the photocurrent signal obtained using homodyne detection is the original shape of the light wave modulation signal; however, during homodyne detection, small changes in the amplitude of the local oscillator will directly impact the signal spectrum and cause signal distortion. Consequently, homodyne detection has a greater requirement for local oscillator amplitude stability.

Fig. 2.6 Amplitude modulated (AM) light wave signal spectrum [6]

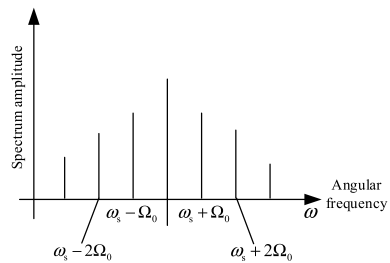
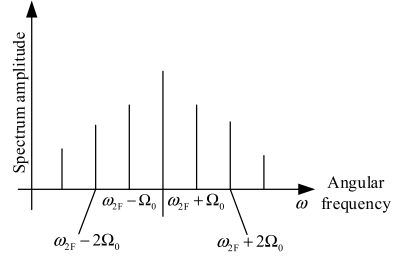


Fig. 2.7 Coherent detection instantaneous intermediate frequency electrical signal spectrum [6]



2.1.5 Dual-Channel Balanced Detection

To reduce the influence of intensity noise on the system, the dual-channel balanced detection technique is used, as shown in Fig. 2.8. This method is based on the structural composition of the balanced heterodyne detection system, using four virtually identical photodetectors to form two sets of balanced heterodyne detectors at the receiving end. The difference frequency signal is derived from the difference in the photocurrent signal received by the detector [8] as

$$\begin{cases} I : I_0 - I_{180} = 2k_1k_4 \cos\left[\omega_{IF} - \left(\varphi + \frac{\pi}{4}\right)\right] \\ Q : I_{90} - I_{270} = 2k_2k_3 \sin\left[\omega_{IF} - \left(\varphi + \frac{\pi}{4}\right)\right] \end{cases} \quad (2.11)$$

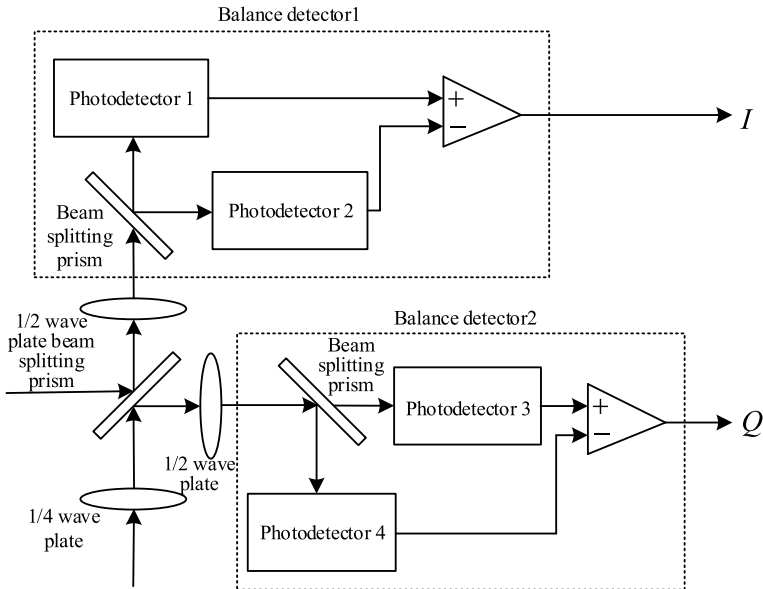


Fig. 2.8 Dual-channel balanced detection [7]

where k_1, k_2, k_3 , and k_4 are the components of the local oscillator light and the signal light in their respective directions, and ω_{IF} is the IF. When the two photocurrents are subtracted, the DC component is completely canceled, and the intensity noise related to the DC component is eliminated. The alternating current (AC) term is directly proportional to the square root of the local oscillator; therefore, the influence of intensity noise is much smaller.

2.2 Coherent Modulation and Demodulation

2.2.1 Optical Modulation

In a coherent optical communication system, the transmitter adopts a direct (or external) approach to modulate the optical carrier amplitude, frequency, or phase. In terms of digital modulation, three basic forms are available: amplitude shift keying (ASK), frequency shift keying (FSK), and phase shift keying (PSK) [9, 10].

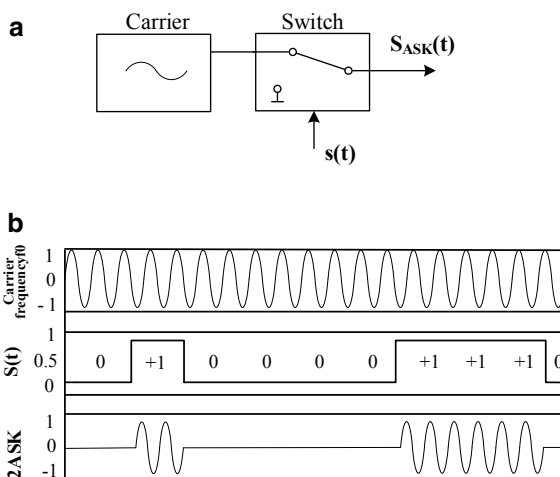
2.2.1.1 ASK Modulation

In ASK modulation, frequency and phase are constants, amplitude is a variable, and information bits are transmitted via the carrier wave amplitude. Because the binary amplitude keying (2ASK) modulation signal has only two levels (0 or 1), the multiplication result is equivalent to turning the carrier frequency on or off. This means that when the modulated digital signal is “1,” the carrier is transmitted, and when the modulated digital signal is “0,” the carrier is not transmitted. This principle is shown in Fig. 2.9a, where $s(t)$ is the baseband rectangular pulse. Generally, the carrier signal is represented by a cosine signal, and the modulation signal converts the digital sequence into a unipolar baseband rectangular pulse sequence. Then, the on/off keying function is used to multiply the output with the carrier and move the spectrum to the vicinity of the carrier frequency. The 2ASK modulation waveform is shown in Fig. 2.9b.

2.2.1.2 PSK Modulation

Figure 2.10 shows that, during PSK modulation, the phase of the carrier changes with the state of the modulated signal. If two carriers with the same frequency start to oscillate at the same time, they are considered to be in-phase; that is, they reach their peak, zero, and nadir values at the same time. A phase inversion occurs when one carrier reaches its peak positive value at the same time as the second carrier reaches its peak negative value. Generally, a signal oscillation is equal to 360° . If one wave is half an oscillation (180°) away from the other wave, the waves are in

Fig. 2.9 Binary amplitude shift keying (2ASK) modulation [10] **a** ASK modulation principle **b** 2ASK output waveform



antiphase. When transmitting digital signals, the “1” code controls the 0° phase and the “0” code controls the 180° phase.

2.2.1.3 FSK Modulation

As the earliest modulation method used in digital signal transmission, FSK uses digital signals to modulate the carrier frequency. In FSK modulation, amplitude and phase are constants, frequency is a variable, and information bits are transmitted via the carrier wave frequency, as shown in Fig. 2.11a. The signals transmitted by FSK are binary (0 and 1), whereas in MFSK, M symbols are represented by M frequencies. The 2FSK waveform is shown in Fig. 2.11b. This method is relatively easy to implement, has good anti-noise and anti-attenuation performance, is stable and reliable, and is the best choice for medium- and low-speed data transmission.

2.2.2 Coherent Demodulation

There are two demodulation methods for coherent detection: synchronous and asynchronous demodulation [11]. In homodyne detection, the optical signal is directly demodulated into a baseband signal and the frequency and phase of the local oscillator light must be the same as those of the signal light. Because of this, synchronous demodulation is predominantly used in homodyne detection, as shown in Fig. 2.12. Synchronous demodulation appears simple; however, the actual technology is highly complex. In heterodyne detection, phase and frequency matching between the signal

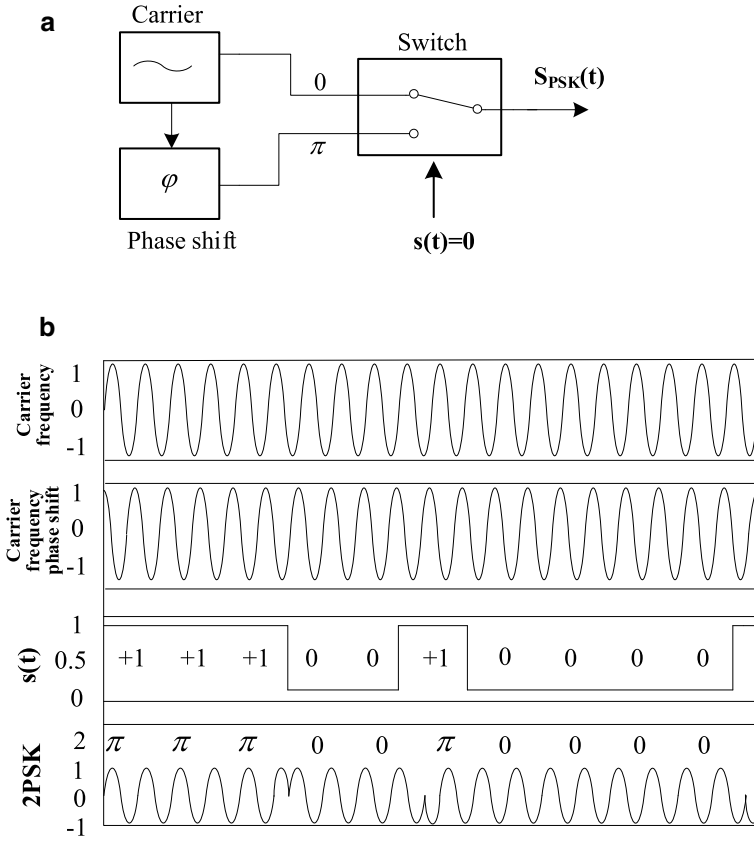


Fig. 2.10 Binary phase shift keying (2PSK) modulation [10] **a** PSK modulation principle **b** 2PSK output waveform

and local oscillator light is not required; hence, synchronous or asynchronous demodulation can be used. Synchronous demodulation requires the recovery of the IF (ω_{IF}), and therefore requires an electric phase-locked loop. Conversely, asynchronous demodulation simplifies the receiver design and is easier to implement. Homodyne and heterodyne synchronous and asynchronous signal demodulation methods share the principles and implementation methods of synchronous and asynchronous demodulation in radio technology, as shown in Fig. 2.13.

2.2.2.1 Heterodyne Synchronization Constant

Figure 2.14 shows the processes performed by the heterodyne synchronous demodulation receiver. The photocurrent generated by the balance detector passes through

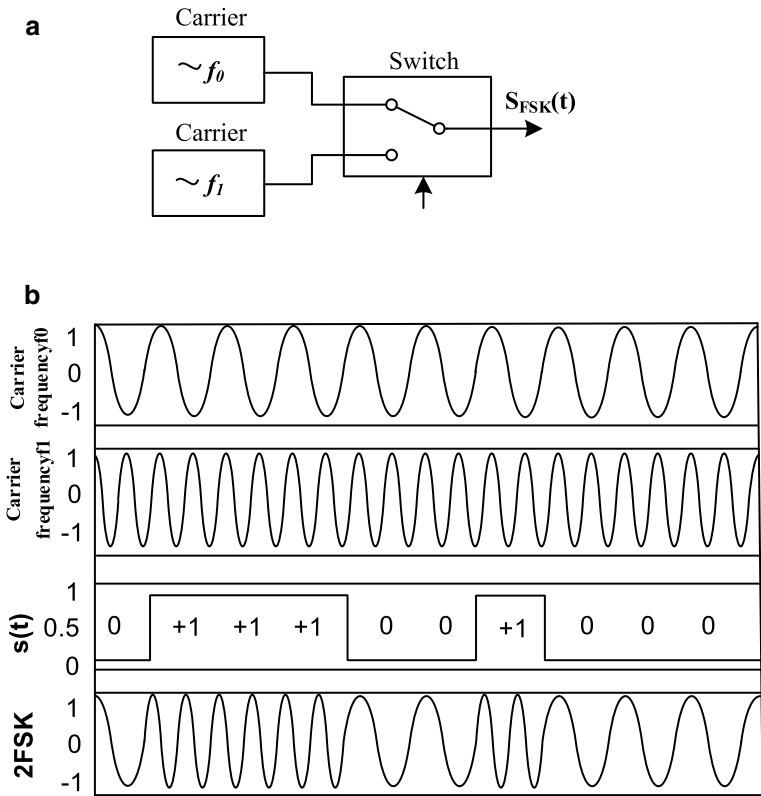


Fig. 2.11 Binary frequency shift keying (2FSK) modulation [10] **a** FSK modulation principle **b** 2FSK output waveform

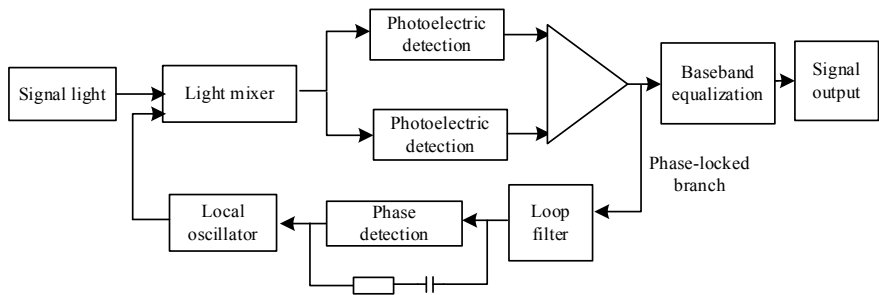


Fig. 2.12 Homodyne synchronous demodulation [11]

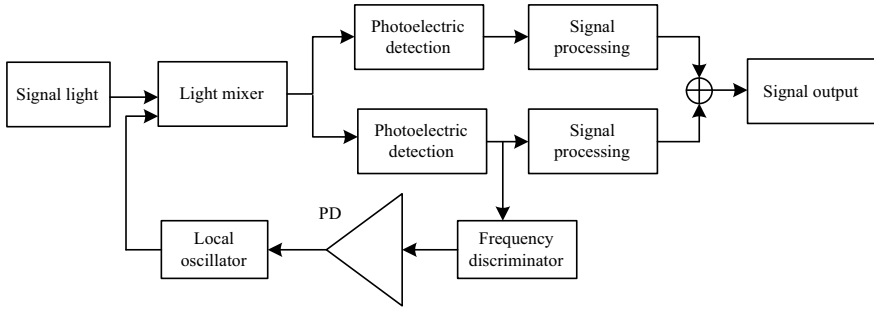


Fig. 2.13 Heterodyne asynchronous demodulation [11]

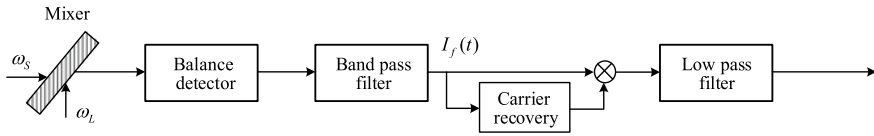


Fig. 2.14 Process diagram of the heterodyne synchronous receiver [12]

the band-pass filter at the IF between the signal and local oscillator light frequencies. When noise is not considered, the signal passes through the frequency band. The filter can be written as [12]

$$I_f(t) = 2\alpha\sqrt{P_S P_L} \cos(\omega_{IF}t - \varphi), \quad (2.12)$$

where φ is the phase difference between the local oscillator and signal lights. For synchronous demodulation, we multiply $I_f(t)$ and $\cos(\omega_{IF}t)$, and use a low-pass filter to obtain the baseband signal as follows [12]:

$$I_d = \frac{1}{2} \left[2\alpha\sqrt{P_S P_L} \cos(\varphi) + i_C \right] \quad (2.13)$$

where i_C is the in-phase Gaussian random noise with a zero mean. Equations 2.2–2.13 shows that only the in-phase noise component affects the performance of the synchronous heterodyne receiver. Synchronous demodulation requires the recovery of an IF microwave carrier ω_{IF} , and there are several ways to achieve this; however, all methods require the use of an electric phase-locked loop. Commonly used phase-locked loops include square and the Costas loops.

2.2.2.2 Carrier Recovery

To achieve coherent demodulation, the receiver must recover a coherent carrier with the same frequency and phase as the transmitted modulated carrier, referred to as carrier recovery. Carrier recovery methods can use a pilot frequency or a direct approach. The pilot method inserts a pilot frequency in a specified position while sending a useful signal, enabling the receiver to extract the carrier from the pilot frequency. The direct method does not send a pilot frequency, and the receiver extracts the carrier directly from the transmitted signal.

The pilot frequency can be inserted into the frequency or time domain. Frequency domain insertion means that the pilot signal is inserted while sending useful information, whereas carrier information is transmitted at a set time in time domain insertion. Using the time domain insertion pilot method, the inserted pilot signal and transmitted information are distinguished in time and the digital signals are grouped into frames. The pilot signal is then sent in a specified time interval in a set time sequence. Time domain insertion is discontinuous and has a short duration; hence, it is not possible to use a narrow band filter to extract the coherent carrier using this method.

Direct carrier recovery methods can be classified as square transformation, square ring, and special phase-locked loop methods. They all perform nonlinear processing of the received signal and then extract the carrier component. The special phase-locked loop method eliminates modulation and denoising, and can identify the phase error between the suppressed carrier component in the received signal and the output signal of the local voltage-controlled oscillator, thereby extracting the coherent carrier. Special phase-locked loops can be in-phase quadrature, inverse modulation, or decision phase-locked loops.

When the direct method is used to extract the carrier, the power signal-to-noise ratio is larger because the pilot power is not occupied at the transmitting end. This can prevent mutual interference between the pilot and the signal, which is common when using the pilot method, and pilot phase errors caused by an imperfect channel. However, some modulation systems, such as single sideband modulation, cannot use direct methods. The insert pilot method uses a separate pilot signal, which can extract the carrier signal, and it can also be used for automatic gain control. In the modulation system, where the direct method cannot be used, only the insert pilot method can be used. It needs to consume power during insertion. Therefore, under the condition of the same total power, compared to the direct method, the power signal-to-noise ratio of the inserted pilot is smaller.

Carrier synchronization has the following basic requirements: a small synchronization error, long hold time, short setup time, and the power band occupied by the synchronization should be as small as possible. Carrier synchronization error distorts the received signal waveform, decreases the signal-to-noise ratio, and increases the bit error rate (BER).

2.2.2.3 Outer Difference Step Demodulation

Figure 2.15 shows the processes performed by external difference step demodulation receiver. Because it does not require IF microwave carrier recovery, the receiver design can be simplified. The receiver uses envelope detection and a low-pass filter to transform the band-pass filtered signal $I_f(t)$ into a baseband signal, and the signal current is determined as follows [12]:

$$I_d = \left[(i_C + 2\alpha\sqrt{P_S P_L} \cos \varphi)^2 + (i_S + 2\alpha\sqrt{P_S P_L} \sin \varphi)^2 \right]^{1/2}, \quad (2.14)$$

where i_C and i_S are Gaussian random noise components, and i_S is the current fluctuation due to shot noise. Shot noise is caused by the dispersion of current-forming carriers in semiconductor devices. At low and intermediate frequencies, shot noise has nothing to do with frequency and shows white noise characteristics. The step demodulation receiver differs from the heterodyne synchronous receiver in that the in-phase and quadrature components of the receiver noise affect the signal output, and the sensitivity of the external difference step receiver is slightly reduced (by approximately 0.5 dB). Furthermore, no synchronous receiver is required for frequency carrier recovery, and the line width requirements of the signal light source and the local oscillator light source are lower. Therefore, the outer differential step demodulation scheme is often used in coherent optical communication systems.

2.2.3 System Performance

2.2.3.1 Signal-to-Noise Ratio of Optical Heterodyne Detection

A core problem with optical heterodyne detection systems is improving the system signal-to-noise ratio. The signal-to-noise ratio at the output of the IF filter can be expressed as [12]

$$\left(\frac{S}{N} \right)_{IF} = \frac{P_{IF}}{N_P} = \frac{\alpha^2 M^2 P_S P_L R_L}{M^2 e \Delta f_{IF} [\alpha(P_S + P_L + P_B) + I_D] R_L + 2kT \Delta f_{IF}}, \quad (2.15)$$

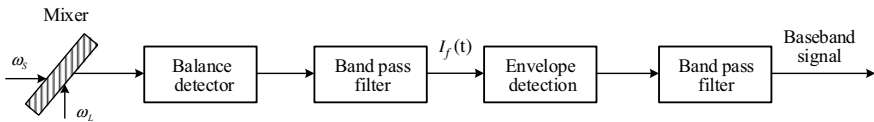


Fig. 2.15 Process diagram of the external difference step receiver [12]

where M is the internal gain of the mixer, P_B is the background radiation power, I_D is the dark current of the optical mixer, and is α the responsivity. The first term in the denominator is shot noise, and the second term is the thermal noise caused by the internal resistance of the mixer and the load resistance of the preamplifier. When the vibration power P_L is sufficiently large, the shot noise caused by the local vibration in the denominator is much larger than all other noises, and Eq. (2.15) becomes [12]

$$\left(\frac{S}{N}\right)_{IF} = \frac{\eta P_S}{h\nu \Delta f_{IF}}. \quad (2.16)$$

This is the maximum signal-to-noise ratio that the optical heterodyne detection system can achieve, and is called the quantum detection limit or quantum noise limit of optical heterodyne detection. From Eq. (2.16), the conditions for realizing quantum noise detection can be derived. For systems where the main noise source is thermal noise, the following conditions must be met to achieve quantum noise limit detection [12]:

$$\frac{e\eta R_L}{h\nu} P_L \Delta i_{IF} > 2kT \Delta f_{IF}.$$

From this we derive

$$P_L > \frac{2kTh\nu}{e^2\eta R_L}. \quad (2.17)$$

Equation (2.17) shows that increasing the optical power of the local oscillator is beneficial to suppress all other noises, except that caused by the signal light, thereby obtaining a high conversion gain. Because the local oscillator itself also produces noise, a larger P_L is not always better. When the optical power of the local oscillator is sufficiently large, the shot noise generated by the local oscillator will be greater than all other noises. If the optical power of the local oscillator is further increased, then the signal-to-noise ratio of the optical heterodyne detection system will be reduced. However, if the local oscillator optical power is too small, the IF conversion gain will decrease. In an actual optical heterodyne detection system, the optical power of the local oscillator should be selected to obtain the best IF signal-to-noise ratio and largest IF conversion gain.

2.2.3.2 Noise Equivalent Power of Optical Heterodyne Detection

In the quantum noise limit formula (i.e., Eq. (2.16)), if $(S/N)_{IF} = 1$, then the noise equivalent power NEP of optical heterodyne detection can be obtained as follows

$$NEP = \frac{h\nu \Delta f_{IF}}{\eta}. \quad (2.18)$$

Equation (2.18) is sometimes referred to as the sensitivity of optical heterodyne detection, and it is the theoretical limit of optical heterodyne detection. If the detector has a quantum efficiency of $\eta = 1$ and a measurement bandwidth $\Delta f_{IF} = 1$ Hz, then the theoretical sensitivity limit of optical heterodyne detection is one photon.

In practice, the sensitivity of optical heterodyne detection is restricted by many factors. For example, thermal and amplifier noise introduced by the optical mixer itself and load resistance will reduce the sensitivity of optical heterodyne detection, even if NEP increases. Therefore, the detection sensitivity of an actual optical heterodyne system can only approach the theoretical limit.

2.3 Factors Affecting Detection Sensitivity

2.3.1 Phase Noise

In a coherent optical detection system, the phase noise of the emitting laser and the local oscillator laser are the main factors leading to decreased sensitivity. Therefore, the homodyne and heterodyne signal phase φ_S , and the local oscillator optical phase φ_L must be relatively stable to reduce sensitivity loss. Phase instability in the photoelectric detection process leads to current instability, thereby reducing the signal-to-noise ratio.

2.3.2 Intensity Noise

For Eq. (2.15), if the intensity noise $\delta_1 = RP_L r_l$ is increased, r_l is the parameter related to the local oscillator laser intensity noise (RIN). If the RIN spectrum is consistent with the receiver bandwidth Δf_{IF} , then Δf_{IF} can be approximated by $2r_l$ (RIN) Δf_{IF} , and the signal-to-noise ratio can be expressed as [12]

$$\left(\frac{S}{N}\right)_{IF} = \frac{P_{IF}}{N_P} = \frac{\alpha^2 M^2 P_S P_L R_L}{M^2 e \Delta f_{IF} [\alpha(P_S + P_L + P_B) + I_D] R_L + \sigma_I + 2kT \Delta f_{IF}}. \quad (2.19)$$

If the receiver works under the limitation of shot noise ($\sigma_S^2 = 2q(I + I_d)\Delta f_{IF}$, q is electronic charge), then the local oscillator optical power P_L should be sufficiently large. Equation (2.19) shows that when P_L increases, intensity noise also increases. It will become as large as shot noise, and the signal-to-noise ratio will decrease unless the signal power P_S is also increased to offset the increase in receiver noise. The increase in P_S is the power penalty caused by the intensity noise of the local oscillator laser. If the system is under the limit of shot noise, then both the dark current I_D and thermal noise in Eq. (2.19) can be ignored. The power penalty δ_1 (expressed in dB) caused by intensity noise can be expressed as [12]

$$\delta_1 = 10 \lg[1 + (\eta P_L(RIN))/h\nu], \quad (2.20)$$

and the BER can be expressed as [13]

$$\text{SNR} = \frac{1}{4} \left[\text{erfc} \left(\frac{I_I - I_D}{\sqrt{2}\sigma_1} \right) + \text{erfc} \left(\frac{I_D - I_0}{\sqrt{2}\sigma_0} \right) \right], \quad (2.21)$$

where I_D is the judging current, I_1 and I_0 represent the current of codes “1” and “0”, respectively. σ_1 and σ_0 represent the variance of codes “1” and “0”, respectively, and $\text{erfc}(\cdot)$ is the error function. When $I_D = \frac{\sigma_0 I_1 - \sigma_1 I_0}{\sigma_1 + \sigma_0}$, the BER is the lowest.

2.3.3 Polarization Noise

In heterodyne detection, the signal light is first mixed with the local oscillator light to obtain the IF signal, whereas the baseband signal is obtained directly in homodyne detection. During this process, it is necessary for the signal and local oscillator lights to maintain the same polarization state. The depolarization effect of the atmospheric channel causes the polarization state of the signal light to change over time; hence, random polarization noise is generated when the signal and local oscillator lights are mixed. When the signal light is orthogonal to the local oscillator light, the signal disappears. Consequently, polarization noise must be suppressed in a coherent optical communication system.

2.3.4 Key Technologies of Coherent Optical Communication Systems

Compared to the IM/DD system, successful coherent optical communication requires that signal and local oscillator light sources use lasers with a high-frequency stability and spectral purity [14]. In incoherent optical communication systems, the IF is generally $2 \times 10^8 - 2 \times 10^9$ Hz, whereas the coherent optical carrier frequency at 1550 nm is approximately 2×10^{14} Hz. The IF is multiplied $10^{-6} - 10^{-5}$ by the optical carrier frequency; hence, the light source frequency stability must be better than 10^{-8} . Furthermore, the signal and local oscillator light sources must meet strict matching conditions during mixing to obtain a high mixing efficiency. Matching approaches include spatial phase, wavefront, and polarization direction matching [15].

2.4 Spatial Phase Conditions for Optical Heterodyne Detection

2.4.1 Spatial Phase Difference Conditions

Figure 2.16 shows the spatial relationships involved in optical heterodyne detection. To study the influence of the non-coincidence of the two light wavefronts on heterodyne detection, we assume that the signal and the local oscillator lights are both plane waves, and there is an angle θ between their wavefronts. To simplify the analysis, we assume that the photosensitive surface of the optical mixer is a square with a side length d . The incident of the local oscillator light is perpendicular to the surface of the mixer, and the signal and the local oscillator lights have a mismatch angle at the wavefront; hence, the incident of the signal light is oblique to the mixer surface, and the same wavefront reaches the surface of the optical mixer at different times. This is equivalent to proceeding at speed v_x in the direction of x , so that the wavefront phase difference occurs at different points on the surface of the optical mixer. The expression of the light field of the local oscillator and signal lights, respectively, can be written as [16]

$$E_L(t) = A_L \cos(\omega_L t + \phi_L) \quad (2.22)$$

$$E_S(t) = A_S \cos(\omega_s t + \phi_s - \frac{\omega_s}{v_x} \cdot x), \quad (2.23)$$

where $\omega_s/v_x = k_x$ is the component of the signal light wave vector k in the direction of x . From Fig. 2.18, we know that $k_x = k \sin \theta = (\omega_s/c) \sin \theta$; hence, $v_x = c/\sin \theta$, where c is the speed of light. The signal light field can be expressed as [16]

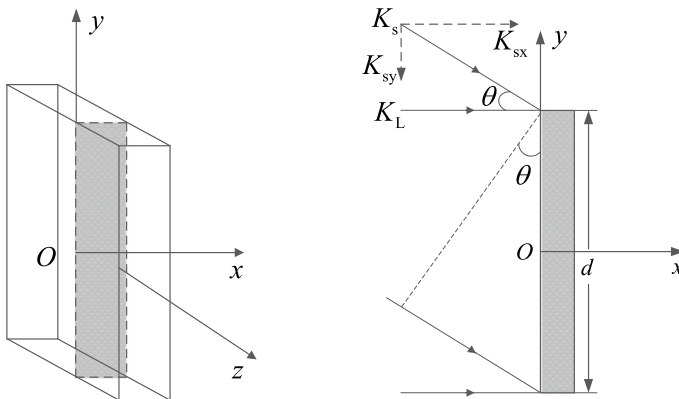


Fig. 2.16 Spatial relationship of optical heterodyne detection [16]

$$E_S(t) = A_S \cos\left(\omega_S t + \phi_S - \frac{2\pi \sin \theta}{\lambda_S} \cdot x\right). \quad (2.24)$$

The total electric field incident on the optical mixer is

$$E_t(t) = E_s(t) + E_L(t). \quad (2.25)$$

The instantaneous photocurrent output following optical mixing can be expressed as [12]

$$\begin{aligned} i_P(t) &= \frac{G\beta}{d^2} \int_{-d/2}^{d/2} \int_{-d/2}^{d/2} \left[A_S \cos(\omega_S t + \phi_S - \frac{2\pi \sin \theta}{\lambda_S} x) + A_L \cos(\omega_L t + \phi_L) \right]^2 dx dy \\ &= \frac{G\beta}{d^2} \int_{-d/2}^{d/2} \int_{-d/2}^{d/2} \left\{ A_S^2 \cos^2\left(\omega_S t + \phi_S - \frac{2\pi \sin \theta}{\lambda_S} x\right) + A_L^2 \cos^2(\omega_L t + \phi_L) \right. \\ &\quad + A_S A_L \cos\left[(\omega_L - \omega_S)t + (\phi_L - \phi_S) + \frac{2\pi \sin \theta}{\lambda_S} x\right] \\ &\quad \left. + A_S A_L \cos\left[(\omega_L + \omega_S)t + (\phi_L + \phi_S) + \frac{2\pi \sin \theta}{\lambda_S} x\right] \right\} dx dy. \end{aligned} \quad (2.26)$$

The instantaneous IF current output of the IF filter is [12]

$$i_{IF} = \frac{G\beta}{d^2} \int_{-d/2}^{d/2} \int_{-d/2}^{d/2} \left\{ A_S A_L \cos\left[(\omega_L - \omega_S)t + (\phi_L - \phi_S) + \frac{2\pi \sin \theta}{\lambda_S} x\right] \right\} dx dy. \quad (2.27)$$

After integration, the instantaneous IF current output from the optical mixer is [12]

$$i_{IF} = G\beta A_S A_L \cos[(\omega_L - \omega_S)t + (\phi_L - \phi_S)] \frac{\sin(\omega_S d / 2v_x)}{\omega_S d / 2v_x}. \quad (2.28)$$

Since $v_x = c / \sin \theta$, the instantaneous intermediate frequency current is related to the mismatch angle θ . The IF current reaches its maximum value when the factor in Eq. (2.28) is

$$\frac{\sin(\omega_S d / 2v_x)}{\omega_S d / 2v_x} = 1,$$

which requires $\omega_S d / 2v_x = 0$ and $\sin \theta = 0$, therefore, $\theta = 0$. It is difficult to adjust the θ angle to zero in practice, because the IF output is generally less than the maximum value. To maximize the IF output, the factor $\sin(\omega_S d / 2v_x) / (\omega_S d / 2v_x)$ should be as close to 1 as possible. If the permitted IF output is 10% lower than the

If output at $\theta = 0$, then $\omega_S d / 2v_x$ must be equal to or less than 0.8 radians. According to this requirement, the mismatch angle θ can be expressed as

$$\theta \leq \frac{\lambda_S}{4d}. \quad (2.29)$$

The mismatch angle θ is proportional to λ_S and inversely proportional to the size of the optical mixer d ; that is, the longer the wavelength and the smaller the size of the optical mixer, the larger the allowable mismatch angle. It is apparent that the spatial collimation requirements of coherent detection are strict, and the shorter the wavelength, the stricter the spatial collimation. Consequently, optical heterodyne detection in the infrared band is more advantageous compared to heterodyne detection in the visible light band. Because of the strict spatial collimation requirements, optical heterodyne detection has good filtering performance, another important feature of the system.

To ensure that the signal and local oscillation light waves are spatially coherent on the light mixing surface, the two beams must be as close to parallel as possible. This relatively strict requirement presents certain difficulties in the realization of optical heterodyne detection. Currently, the most promising means of addressing this issue is by using the Airy disk principle method, as shown in Fig. 2.17.

According to physical optics, a normally incident light wave will form a diffraction spot on the detector at the focal plane after passing through a lens with an area of A_r . The area $\lambda^2 f^2 / A_r$ of the largest peak in the diffraction spot is called the Airy spot area, and this area determines the diffraction-limited field of view of the receiving system. If it is represented by a solid angle of Ω_{dl} , then

$$\Omega_{dl} \approx \frac{\lambda^2 f^2}{A_r} \cdot \frac{1}{f} = \frac{\lambda^2}{A_r}. \quad (2.30)$$

If it is expressed by a plane angle, then

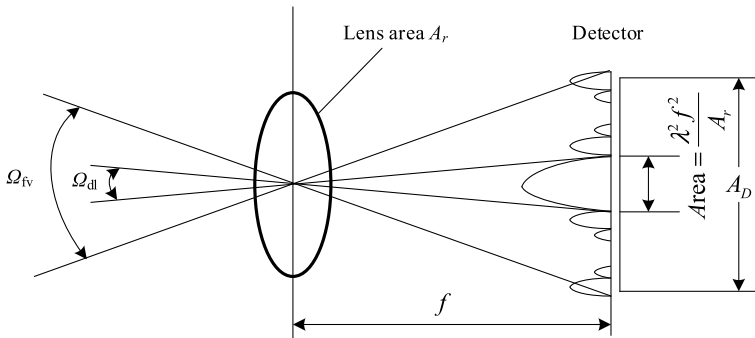


Fig. 2.17 Airy disk principle in an optical lens antenna [17]

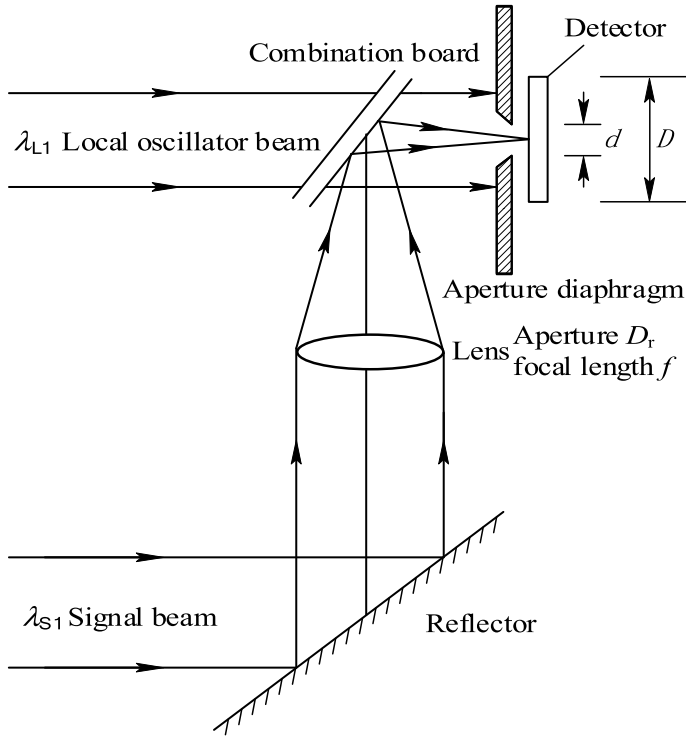


Fig. 2.18 Schematic diagram of the Airy disk principle [18]

$$\theta_{dl} \approx \frac{\lambda}{D_r}, \quad (2.31)$$

where D_r is the lens diameter. A schematic diagram of the Airy disk principle is shown in Fig. 2.18. A lens is used to focus the signal light onto the surface of the optical mixer, and the effective area of the optical mixer is the area of the Airy disk. At the same time, the local oscillator light is irradiated on the Airy spot, and light mixing can occur. As the temperature D_r/f increases, the requirements for mismatch angles become wider.

2.4.2 Frequency Conditions

2.4.2.1 Good Monochromaticity

To obtain high-sensitivity coherent detection, it is necessary for the signal and local oscillator lights to have high monochromaticity and frequency stability. In terms of physical optics, optical heterodyne detection (or coherent detection) is the result of

interference between two light waves after superposition, and this is heavily dependent on the monochromaticity of both the signal and local oscillator beams. Light with only a single wavelength is called monochromatic light, and monochromaticity means that the beam is comprised of light within a very narrow frequency or spectral line. The color of light is determined by its frequency, and this is generally determined by the light source. Fully monochromatic light is difficult to produce because emissions from any light source have a certain frequency range, and the intensity of light for each frequency is different. A high degree of monochromaticity is an important characteristic of lasers; however, their excited state is always comprised of a certain energy level width, and consequently, the laser beam width $\Delta\nu$ cannot be zero. Generally, narrower values of $\Delta\nu$ mean greater light monochromaticity. To obtain laser output with good monochromaticity, the laser must operate in a single longitudinal mode (single frequency) by using a short cavity structure or other mode selection technologies.

2.4.2.2 Small Frequency Drift [19]

The frequency drift of the signal and local oscillator lights must be limited within the allowable range to prevent performance deterioration of the optical heterodyne detection system. If the relative drift of the signal and local oscillator lights is large, then the difference between the two frequencies may exceed the bandwidth of the IF filter. If this occurs, the pre-amplification and IF amplifying circuit following the optical mixer cannot apply the IF signal. In optical heterodyne detection, special measures are required to stabilize the frequencies of the signal and the local oscillator lights. This is an important reason why optical heterodyne detection is more complicated than the direct detection method.

2.4.3 Polarization Conditions

The essence of optical heterodyne detection is the interference between the local oscillator and signal lights on the photosensitive surface of the photodetector. According to the difference in the polarization state of the light wave during the propagation process, the electric field distribution of the local oscillator and signal lights, respectively, can be expressed as [20]

$$E_L(t) = \hat{e}_L A_L \cos(\omega_L t + \phi_L) \quad (2.32)$$

$$E_S(t) = \hat{e}_S A_S \cos(\omega_S t + \phi_S), \quad (2.33)$$

where \hat{e}_L and \hat{e}_S are the unit vectors of the polarization directions of the local oscillator and signal lights, respectively. By incorporating Eqs. (2.32) and (2.33) into Eq. (2.3), the IF current can be simplified as follows [16]

$$i_{IF} = \alpha \hat{e}_S \hat{e}_L A_S A_L \cos[(\omega_S - \omega_L)t + (\phi_S - \phi_L)] \quad (2.34)$$

when the polarization directions of the two beams of light are the same, $\hat{e}_L \hat{e}_S = 1$, and when the polarization directions of the two beams have an angle of θ , Eq. (2.34) becomes [20]

$$i_{IF} = \alpha \cos \theta |\hat{e}_S| |\hat{e}_L| A_S A_L \cos[(\omega_S - \omega_L)t + (\phi_S - \phi_L)] \quad (2.35)$$

where $\cos \theta = \hat{e}_L \hat{e}_S / |\hat{e}_L| |\hat{e}_S|$ represents the matching influence of the polarization directions of the two beams on the heterodyne signal, and its value is between 0 and 1. When the polarization directions of the two beams are the same; that is, $\theta = 0^\circ$, the heterodyne output IF current is the largest. Conversely, when the polarization directions of the two beams are perpendicular; that is, $\theta = 90^\circ$, the heterodyne output IF current is the smallest.

In summary, the spatial, frequency, and polarization matching conditions are all important to realize successful optical heterodyne detection. To obtain the maximum IF output, the wavefronts of the signal and local oscillator light waves must match. To perform effective mixing, the signal and local oscillator lights must be monochromatic, have the same polarization, and their frequencies must be extremely stable.

2.5 Adaptive Optical Wavefront Correction

The atmospheric turbulence causes a random fluctuation in the wave front of the laser signal during propagation, which results in beam expansion, phase fluctuation, beam bending and drift. It increases the communication bit error rate and reduces the stability of the communication link. For coherent optical communication systems, the spatial angular collimation cannot be guaranteed when the signal light containing aberrations and the local oscillation light are mixed, thus reducing the mixing efficiency and signal-to-noise ratio of the system and the impact of turbulence on the performance of coherent optical communication systems is more obvious in the case of long-distance and strong turbulence.

As a comprehensive new optical technology, adaptive optics (AO) covers several disciplines, such as optics, communication, control, computer, and mechanics. It is used to correct the random distortion of wave front caused by the change of external environment during the propagation of light beam in real time, and suppress the effect

of atmospheric turbulence on the signal light. Then the beam quality is improved. Therefore, adaptive optics is considered to be the most effective and promising method to suppress atmospheric turbulence.

2.5.1 Wavefront Distortion Correction System

An AO system can be implemented with or without wavefront detection. The AO system without wavefront detection consists of a wavefront corrector, wavefront controller and performance evaluation function module. The AO system with wavefront detection consists of a wavefront sensor, wavefront controller and wavefront corrector. Figure 2.19 shows a schematic diagram of an AO system with wavefront detection. Here, the wavefront sensor measures the wavefront error in real time and transmits the measured wavefront error to the wavefront controller, and the wavefront controller receives the aberration beam information from the wavefront sensor and calculates and obtains the control voltage of the wavefront corrector through a certain algorithm. The wavefront corrector receives the control voltage from the wavefront controller to correct the wavefront aberration and improve the communication quality.

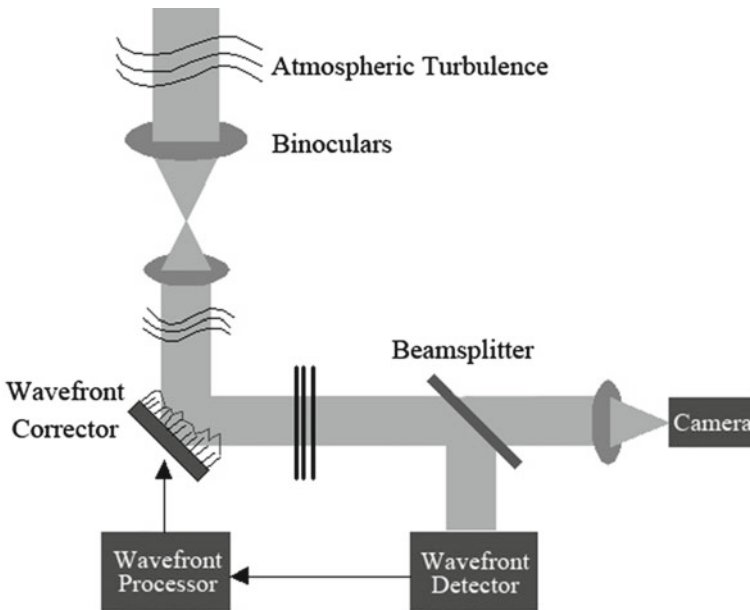


Fig. 2.19 Schematic diagram of an adaptive optics system with wavefront detection

2.5.2 Wavefront Measurement and Correction

The wavefront sensor is the eye of the adaptive optics system to detect the wavefront aberration of the system servo loop. By measuring the phase aberration of the optical wavefront on the pupil surface in real time, it provides a real-time voltage control signal to the wavefront corrector. Then, an image close to the diffraction limit after closed-loop correction is obtained. Figure 2.20 shows the detection schematic of the Shack-Hartmann wavefront sensor, which consists of a microlens array and an electrically coupled device. The microlens array splits a completed spot into multiple tiny sub-spots, each of which is focused onto the focal plane and imaged onto the detection target surface of a charge coupled device. The amount of wavefront distortion is estimated by comparing the actual focal position of the subaperture with the ideal focal position.

The wavefront corrector is used to compensate the phase of the wavefront with the completed aberration. It corrects the wavefront aberration by changing its own aperture shape and thus changing the beam's optical path difference. At present, a commonly used wavefront corrector is based on the position shift of the reflecting mirror, which has excellent characteristics of fast response, large deformation displacement, wide working spectrum, high optical utilization, and multiple implementation methods. The aperture shape distribution of piezoelectric deflection mirror, few-cell-count deflection mirror and multi-cell-count deflection mirror is shown in Fig. 2.21. The piezoelectric deflection mirror is electrically connected with a 4-point driver driven by two pairs of independent piezoelectric ceramics. The

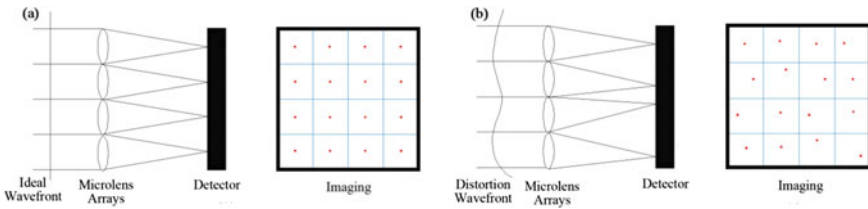


Fig. 2.20 Principle of Shack-Hartmann wavefront detector **a** ideal wavefront; **b** distorted wavefront

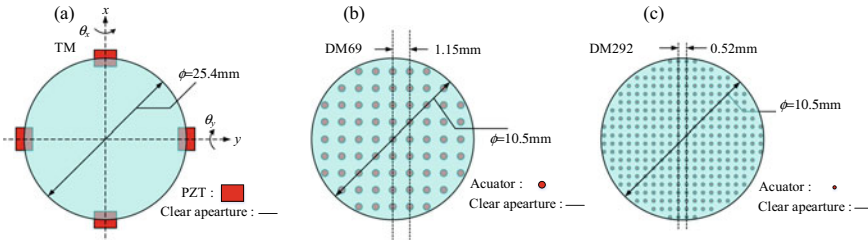


Fig. 2.21 Wavefront corrector surface type **a** Piezoelectric deflector; **b** Few-cell deflector; **c** Multi-cell deflector

two piezoelectric ceramics located in x or y axis change their own deformations by voltage to change the wavefront tilt in the direction of x or y axis as the roll axis. The few-cell deflection mirror and the multi-cell deflection mirror are electromagnetic continuous aperture deflection mirrors with 69 and 292 driving cells, respectively.

For free-space optical communication systems, adaptive optics mainly uses the phase-conjugate principle for feedback closed-loop control to achieve beam correction and compensation. The principle is as follows: When a laser beam with a specified shape wavefront is emitted and disturbed by atmospheric turbulence, a random fluctuation occurs and wavefront aberrations are produced. The aberrated light field with phase error, $E_{\text{atmosphere}}$, is expressed as

$$E_{\text{atmosphere}} = |E|e^{i\varphi}, \quad (2.36)$$

where φ represents the fluctuation of the initial beam phase due to the atmospheric turbulence disturbance. The wavefront distortion is measured in real-time by the wavefront sensor and transmitted to the wavefront controller. Then, the wavefront controller calculates the error signal based on the received wavefront distortion and obtains the control signal by a control algorithm. After that, the control is sent to the driver unit of the wavefront calibrator. Finally, the wavefront calibrator produces a certain deformation to form a calibration front. The optical field E_{DM} with phase fluctuation generated by the corrector is

$$E_{DM} = |E|e^{-i\varphi}. \quad (2.37)$$

From Eq. (2.37), it can be seen that the wavefront corrector face shape deformation can produce a wavefront with the same wavefront shape as the aberrated beam detected by the wavefront sensor, however, the propagation direction is opposite, that is, the wavefront conjugate. Then the aberrated wavefront is corrected by the wavefront corrector, and the phase error is compensated by the superposition of the two optical fields, i.e., the corrected beam has nearly the same phase information as the beam at the transmitter.

Coherent mixing efficiency is an important criterion in coherent detection, which reflects the degree of matching between the local oscillation light and the signal light. The advantage that the signal-to-noise ratio of coherent detection can reach the quantum noise limit is derived under the assumption that the mixing efficiency is 100%, however, the mismatch of both the amplitude and phase of the signal light and the local oscillation light will lead to a decrease in the mixing efficiency. The mixing efficiency η_{mixing} in coherent detection systems is usually defined as

$$\eta_{\text{mixing}} = \frac{\left[\int_U A_S A_L (\cos(\Delta\varphi)) dU\right]^2 + \left[\int_U A_S A_L (\sin(\Delta\varphi)) dU\right]^2}{\int_U |E_S|^2 dU \cdot \int_U |E_L|^2 dU}, \quad (2.38)$$

where e is the electron power, η is the quantum efficiency, U is the detector area, h is Planck's constant, ν is the carrier light frequency, B is the detector bandwidth, and $\Delta\varphi$ is the phase difference between the signal light with aberrated wavefront phase and the local oscillation light.

For the IF electrical signal noise output after mixing, there are mainly detector scattering noise, relative intensity noise and detector thermal noise, as the intensity of the local oscillation light is much higher than the signal light, the noise of the local oscillation light scattering noise dominates. Therefore, the expression of signal-to-noise ratio can be obtained as follows

$$\text{SNR} = \frac{e\eta \int_U |E_S|^2 dU}{h\nu B} \cdot \frac{[\int_U A_S A_L (\cos(\Delta\varphi)) dU]^2 + [\int_U A_S A_L (\sin(\Delta\varphi)) dU]^2}{\int_U |E_S|^2 dU \cdot \int_U |E_L|^2 dU}. \quad (2.39)$$

The system BER can be expressed as

$$\text{BER} = \frac{1}{2} \cdot \text{erfc}(\sqrt{\text{SNR}}). \quad (2.40)$$

From Eqs. (2.38), (2.39) and (2.40), we can see that the light intensity undulation and phase distortion caused by atmospheric turbulence will cause the phase mismatch between the signal light and the local oscillation light, thus reducing the coherent mixing efficiency and system signal-to-noise ratio and increasing the BER.

For adaptive optics systems, the incremental proportional-integral algorithm is mainly used for calibration. The adaptive optics correction model based on the direct slope method with a proportional-integral controller is shown in Fig. 2.22. According to the model, the determination of the system command matrix is first performed using the push-pull method, followed by the detection of the wavefront slope S_r using a wavefront sensor and its conversion to a voltage value V_e using the direct slope method. This value is then transmitted to the incremental proportional-integral controllers k_p, k_i for control to obtain the final deformation mirror drive voltage value V_{dm} and send it to the driver to drive the deformation mirror operation.

After correction using the incremental proportional integration algorithm, the value of the deformation mirror driver voltage V_{dm} at moment k is

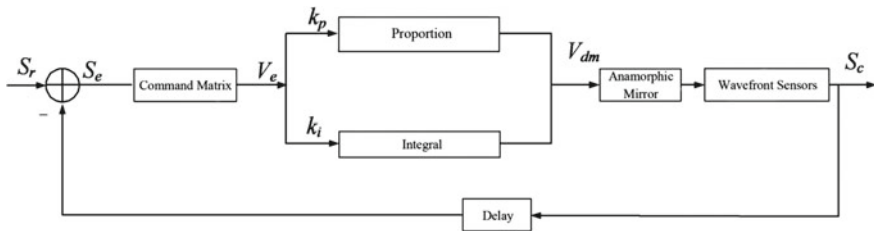


Fig. 2.22 Proportional integration algorithm calibration model for adaptive optics system

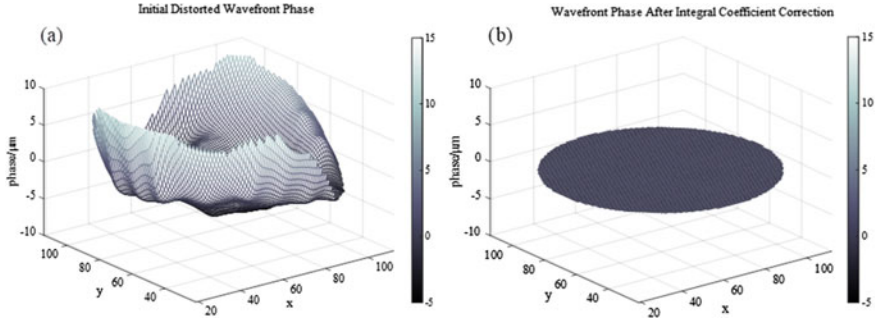


Fig. 2.23 Adaptive optical wavefront correction phase diagram **a** before correction **b** after correction

$$\begin{aligned} V_{dm}(k) &= V_{pid} + V_{dm}(k-1) \\ &= k_p[V_e(k) - V_e(k-1)] + k_i V_e(k) + V_{dm}(k-1), \end{aligned} \quad (2.41)$$

where $S_e(k)$ is the wavefront slope at moment k , and $V_e(k)$ is the corresponding drive voltage obtained using the direct slope method.

Figure 2.23 shows the phase distribution before and after the adaptive optical wavefront correction. It can be seen from the figure that the proportional integration algorithm has obvious correction effect. The wavefront is also very close to the plane wave form compared with the initial state, and the degree of concavity and surface non-flatness are greatly reduced. This demonstrates that the proportional integration algorithm in the adaptive optics system can effectively correct the wavefront distortion and improve the communication performance of the wireless laser communication system.

2.5.3 Wavefront-Free Measurement System

In the conventional adaptive optics system, the wavefront sensor is complex and expensive, and in the strong turbulent environment, the laser beam is affected by strong turbulence will constantly flicker and phase interruption, which makes the wavefront information, obtained by the wavefront sensor inaccurate and cannot accurately correct the wavefront distortion. For the above shortcomings, adaptive optics without wavefront sensor has been proposed. Adaptive optics without wavefront sensor does not require a wavefront sensor, thus avoiding the disadvantages of traditional AO systems. Figure 2.24 shows the block diagram of the adaptive optics system without wavefront sensor. After the distorted wavefront affected by atmospheric turbulence is emitted by the deformation mirror, the optical signal (such as far-field spot light intensity distribution, spatial light-to-fiber coupling power) is

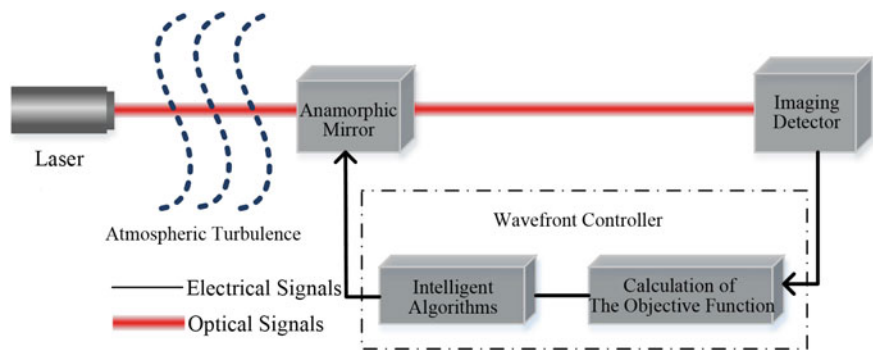


Fig. 2.24 Block diagram of adaptive optics system without wavefront sensor

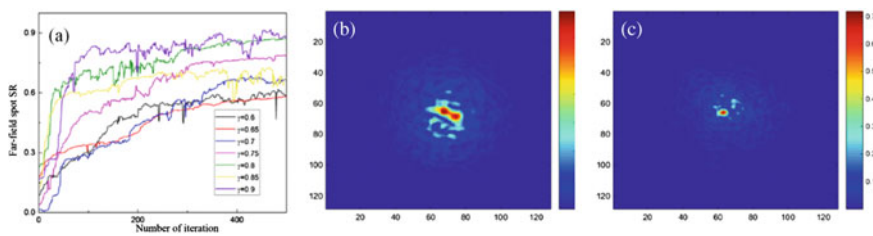


Fig. 2.25 SPGD algorithm based wavefront correction **a** Light intensity SR variation curve **b** Light spot before correction **c** Light spot after correction

collected by the imaging detector, the objective function is calculated by the wavefront controller, and the calculated voltage value is applied to the deformation mirror using an optimization algorithm to complete the correction of the distorted wavefront.

Wavefront aberration correction by adaptive optics without wavefront detection using stochastic parallel gradient descent algorithm. Figure 2.25 shows the correction curve using the stochastic parallel gradient descent algorithm and the spot before and after the correction, before the correction of the far-field spot scattering serious and energy does not converge, after the correction of the adaptive optics technology based on the stochastic parallel gradient descent algorithm without wavefront detection, the spot center light intensity increased significantly, and more convergence of energy.

2.6 Summary and Prospects

Some of the research challenges facing optical wireless communication include improving the transmission distance without excessive power consumption and increasing channel capacity. Improving detection sensitivity has proven to be an effective means of suppressing atmospheric turbulence and increasing channel

capacity. Coherent detection is currently a key research area of optical wireless communications.

To realize effective coherent optical communication, numerous technical challenges need to be addressed. These include (1) wavefront correction; (2) improving light source monochromaticity; (3) atmospheric degradation of light polarization; and finally, (4) the high-speed movement of communication terminals should consider the impact of Doppler frequency shift on coherent detection. Resolving these challenges will help to facilitate the practical application of coherent optical communication technology, and future development should continue to focus on resolving these technical challenges. For example, (1) low-cost narrow-band light sources have demonstrated great potential for enhancing coherent optical communications; (2) new coherent principles, such as subcarrier coherent and homodyne detection, have been shown to improve transmission sensitivity; and (3) simple methods of wavefront correction have proven to be effective in suppressing turbulence.

2.7 Questions

- 2.1 What is coherent optical communication? What are their characteristics?
- 2.2 What is homodyne detection? What is heterodyne detection? What are the similarities and differences between the two?
- 2.3 Briefly describe what is frequency shift keying? What is amplitude shift keying?
- 2.4 Briefly describe the method of coherent detection.
- 2.5 What factors are related to the SNR of optical heterodyne detection?
- 2.6 The coherent optical communication system requires the signal light and the local oscillator light to meet strict matching conditions when mixing. What aspects does this matching include?
- 2.7 Briefly describe the spatial phase conditions, frequency conditions, and polarization conditions of optical heterodyne detection.

2.8 Exercises

- 2.1 As shown in Fig. 2.4, it is a self-homodyne coherent detection system. The light fields in the two optical paths are coherent with the other after a 1-bit delay. The two coherent input light fields are

$$E_{S1} = \frac{\sqrt{2G}}{2} A_S \exp[-i(2\pi f_C t + \phi_S)],$$

$$E_{S2} = \frac{\sqrt{2G}}{2} A_S \exp[-i(2\pi f_C t + \phi_S + \Delta\phi)],$$

where $\Delta\phi$ is the phase difference of adjacent bits

$$\Delta\phi = |\phi_n - \phi_{n-1}| = \begin{cases} 0 & \text{"1" code} \\ \pi & \text{"0" code} \end{cases}.$$

Suppose that two beams of coherent input have the same frequency, constant phase difference, and same polarization direction, and the optical power detected by the photodetector is $E_{S3} = P_{S1} + P_{S2} + 2\sqrt{P_{S1}P_{S2}} \cos \Delta\phi$. Try to analyze the coherent gain and signal-to-noise ratio of this detection method.

- 2.2 As shown in Fig. 2.5, it is a subcarrier modulation heterodyne detection system. The transmitting end of the optical wireless communication system performs subcarrier QPSK intensity modulation on the baseband signal, signal optical power is $P_e(t) = P_s[1 + k_p \cos(\omega_c t + \varphi_j)]$, P_s is the average power of signal light; k_p is the degree of optical modulation; the sub-carrier phase is $\varphi_j = j \cdot \frac{\pi}{4}$, $j = 1, 3, 5, 7$, and ω_c is the angular frequency of the sub-carrier at the transmitter. Supposing that the local oscillator optical power is $P_l(t) = P_l[1 + k_p \cos(\omega_l t)]$, find the heterodyne gain and bit error rate of the system.

- 2.3 The coherent optical detection system needs a phase-locked loop, as shown in the Fig. 2.26. The transfer function of the phase-locked loop is

$$H(j\omega) = \frac{\theta_{VCO}(j\omega)}{\theta_N(j\omega)} = \frac{F(j\omega)A_{IF}^4 K(j\omega)^{-1}}{1 + F(j\omega)A_{IF}^4 K(j\omega)^{-1}}$$

$$= \frac{(A_{IF}^4 K j\omega \tau_2 + A_{IF}^4 K)/\tau_1}{(j\omega)^2 + (A_{IF}^4 K j\omega \tau_2 + A_{IF}^4 K)/\tau_1} = \frac{\omega_n^2 + 2\zeta\omega_n j\omega}{\omega_n^2 + 2\zeta\omega_n j\omega + (j\omega)^2},$$

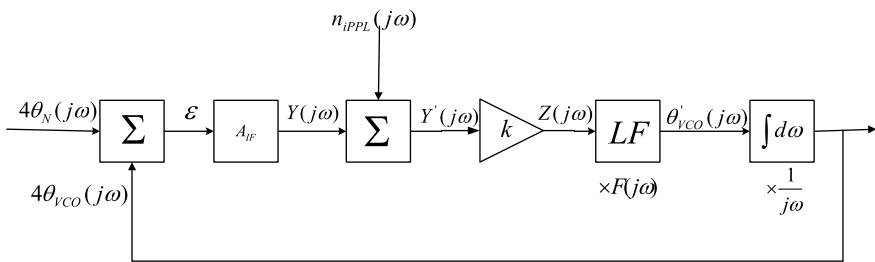


Fig. 2.26 Coherent optical detection system with a phase-locked loop

where $\omega_n^2 = A_{IF}^4 K / \tau_1$ and $\zeta = \frac{A_{IF}^2 \tau_2}{2} \sqrt{K / \tau_1}$.

Power spectral density of phase error is.

$$G_\varepsilon(f) = |1 - H(j\omega)|^2 G_{PN}(f) = \frac{16}{\pi f^2} \cdot \frac{\omega^4 / \omega_n^4}{[1 - (\omega / \omega_n)^2]^2 + [2\zeta(\omega / \omega_n)]^2}.$$

Find the phase noise variance of the loop.

2.4 Analysis and discussion of quantum noise in optical heterodyne detection.

[Hint] The generalized rate equation describing the laser can be expressed as

$$\dot{n} = (G - r)n + R_s + F_n \quad (2.42)$$

$$\dot{N} = P - Gn - S + F_N \quad (2.43)$$

$$\dot{\phi} = \frac{\alpha}{2}(G - r) + F_\phi \quad (2.44)$$

where n is the number of photons in the LD, G is the gain, r is the cavity loss, R_s is the rate of spontaneous radiation to the cavity mode, F_n , F_N , and F_ϕ are Langevin noise sources, and N is the number of carriers in the active region, P is the pumping rate, S is the carrier coincidence rate, and the α linear broadening factor. Consider the first-order approximation

$$n(t) = n_0 + \delta n(t)$$

Then, Eqs. (2.42)–(2.44) become

$$\delta \dot{N} + G_e \delta n + \Gamma_N \delta N = F_N, \quad G_e = -[G_n n_0 + G_0] \quad (2.46)$$

$$\delta \dot{\phi} - \frac{\alpha}{2} G_N \delta N = F_\phi$$

Consider the signal light field:

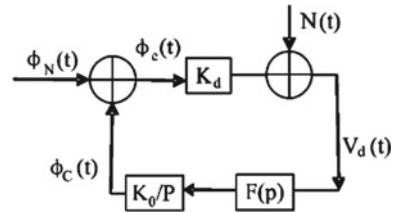
$$E_1(t) = \sqrt{2P_1(t)} \exp[j(\omega_1 t + \phi_1(t))]$$

Local oscillator light field:

$$E_2(t) = \sqrt{2P_2(t)} \exp[(j\omega_2 t + \phi_2(t))]$$

The output current of the photodetector:

Fig. 2.27 Phase-locked loop of coherent detection system



$$\begin{aligned}
 I(t) &= R|E_1(t) + E_2(t)|^2 \\
 &= R \left\{ P_1(t) + P_2(t) + 2\sqrt{P_1(t)P_2(t)} \cos[\Delta\omega t + \varphi_1(t) - \varphi_2(t)] \right\} + n_s(t)
 \end{aligned}$$

By approximating the above formula, the frequency spectrum of the noise and the noise change incoherent detection can be analyzed.

- 2.5 A coherent detection system as shown in Fig. 2.5.
- (1) Find the BPSK modulation, the system's bit error rate, and the upper limit of the bit error rate in the case of white noise.
 - (2) Find the bit error rate under DBPSK modulation.
- 2.6 A coherent detection system shown in Fig. 2.5. Analyzes the output noise characteristics in the case of large input signal-to-noise ratio and Gaussian noise.
- 2.7 Consider a coherent optical communication using double-balanced detection,
- (1) Find the corresponding bit error rate of the double-balance detector under the quantum limit condition.
 - (2) Find the signal-to-noise ratio of the system using double-balanced detection when only shot noise and thermal noise are considered.
 - (3) Try to analyze the common-mode rejection ratio of the double-balanced detector.
- 2.8 As shown in Fig. 2.27 in the phase-locked loop of the coherent detection system, find the minimum phase variance of the loop.
- 2.9 Supposing a coherent light detection system, only the shot noise generated by the signal light is considered when the thermal noise is very small. Find the quantum limit of the coherent detection.
- 2.10 Using Jensen's inequality to simplify the channel capacity to get its lower limit, try to estimate the relationship between the pulse position modulation order when the maximum transmission rate is reached and the reconstruction time of population inversion.
- 2.11 Try to analyze the influence of the balanced detector on the sensitivity of coherent optical communication.
- 2.12 Try to analyze the requirements for the consistency of balanced detectors in coherent optical communications.

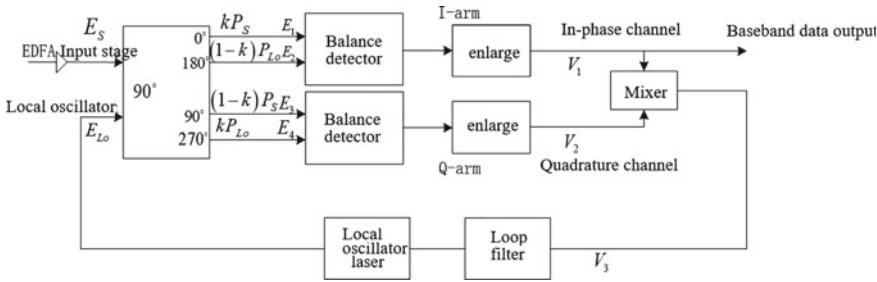


Fig. 2.28 Optical phase-locked loop in coherent optical communication

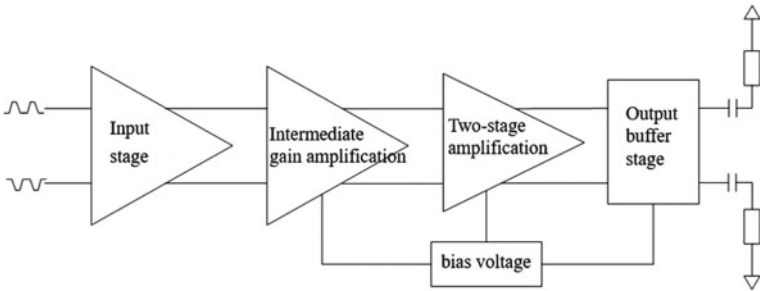


Fig. 2.29 Pre-amplifier of balanced detector

- 2.13 Try to compare the signal-to-noise ratio between the balanced detector and the single-tube detection.
- 2.14 Try to analyze the output intermediate frequency current of the double balance detection in Fig. 2.8.
- 2.15 Try to analyze the improvement of the coherent optical communication signal-to-noise ratio by the double-balanced detector and the influence of the consistency coefficient on the detection signal-to-noise ratio.
- 2.16 Figure 2.28 shows an optical phase-locked loop in coherent optical communication. Analyzes its working process and discuss the relationship between phase noise and bit error rate.
- 2.17 Try to analyze the working process of the pre-amplifier of the balanced detector as shown in Fig. 2.29.

References

1. Xiao X, Chen M (2011) The design of the external differential step demodulation receiver for coherent optical communication. Popul Sci Technol 4:21–22
2. Wang Q, Hu Y, Lin C (1994) Photoelectric detection technology. Publishing House of Electronics Industry

3. Lin C (2010) High-frequency electronic circuits. Publishing House of Electronics Industry
4. Zhao F (2011) Research on the receiving performance of inter-satellite optical communication system based on single-mode fiber coupling and self-difference detection, p 1. Harbin Institute of Technology
5. Ke X, Chen J (2014) Experimental research on non-optical heterodyne detection of 1 km atmospheric laser communication system. *J Appl Sci* 32(4):379–384
6. Fan C, Cao L (2006) Principles of communication, 6th ed. National Defense Industry Press
7. Huai Y (2013) Simulation study of double-balanced heterodyne laser detection system. *Softw J* 34(4):132–134
8. Zhou L (2011) 900 optical mixer in coherent optical communication, p 4. University of Electronic Science and Technology of China
9. Ke X, Chen J (2012) Research on influencing factors and key technologies of wireless optical heterodyne detection system. *Semicond Optoelectron* 33(4):548–557
10. Yu J, Chi N, Chen L (2013) Coherent optical communication technology based on digital signal processing, p 10. People's Posts and Telecommunications Press
11. Liu Z, Zhou Y, Hu L, et al (2008) Optical fiber communication, 2nd ed, p 12. Xidian University Press
12. Qin Y (2012) Coherent optical communication link heterodyne receiving technology research, p 6. University of Electronic Science and Technology of China
13. Zhu Y, Wang J, Lu L (2011) Principle and technology of optical communication, 2nd ed. Science Press
14. An Y, Zeng X, Feng Z (2010) Photoelectric detection and signal processing. Science Press
15. Griadi RM, Kabo S (1982) Optical communication (translated by Chen Zhenguo and others). People's Posts and Telecommunications Press
16. Wang L (2011) Research on coherent optical communication heterodyne detection technology. Xidian University
17. Lu X et al (2013) Phase-locked loop analysis based on QPSK modulation in coherent optical communication. *J Chang Univ Sci Technol* 36(3):49–52
18. Li L (1992) Quantum noise in optical heterodyne detection. *Opt Commun Res* 2:21–26
19. Bi G (1990) Intensity modulation nonlinear heterodyne detection optical communication system and its performance. *Chin J Electron* 02:69–75
20. Jacobaen G, Garrets I (1988) Optical ASK heterodyne receiver: comparison of a theoretical model with experiment. *Electron Lett* 22(3):170–171

Chapter 3

Modulation, Demodulation, and Coding



Modulation in optical wireless communication is the process of loading information onto the light wave. The modulator is an electro-optic converter, which changes the parameters of the output beam, such as intensity, frequency, phase, polarization state, and beam/photon orbital angular momentum, with the signal. This chapter introduces the modulation, demodulation, and coding of the optical signal.

3.1 Modulation

Modulation is using one signal (modulation signal) to control another signal as a carrier (carrier signal) so that a certain parameter of the carrier signal changes with the modulation signal. As shown in Fig. 3.1, modulation techniques can be divided into active and passive modulation. Active modulation is conducted at the optical transmitter, whereas passive modulation is conducted at the opposite side, where the light source and modulation process are separated. Active modulation includes internal and external modulation: parameters of the light source are modulated in the former, and parameters of the light wave are modulated in the latter.

3.1.1 Basic Concepts

In optical wireless communication, the information carried by a laser beam, including language, text, images, and symbols, is transmitted through transmission channels, such as atmosphere or free space, and received, identified, and restored by the optical receiver. The process of loading information onto a light wave carrier is called modulation, and the device to complete this process is called a modulator. The process of loading information onto a laser beam is called laser modulation driving technology, where the laser is the carrier wave, and the low frequency information as a controller

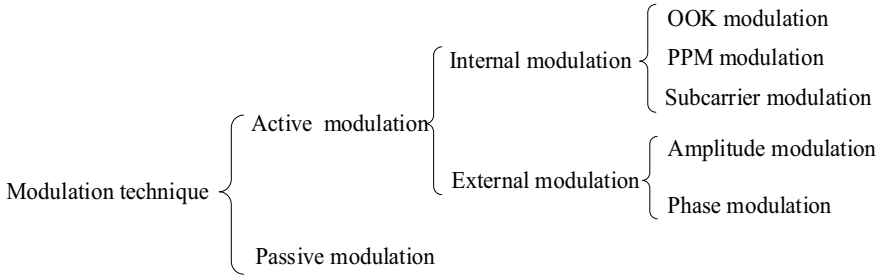


Fig. 3.1 Classification of optical modulation technology

is called the modulation signal. In optical communication, the light wave is used as a carrier for information transmission.

According to the nature of modulation, optical modulation can be divided into amplitude modulation, intensity modulation, frequency modulation, phase modulation, and pulse modulation. Depending on the working principle of the modulator, modulation can be classified as electro-optic, acousto-optic, magneto-optic, etc. In addition, it can be divided into analog and digital modulation.

3.1.2 Analog and Digital Modulation

A baseband signal is the original electrical signal without modulation via spectrum shifting or transformation. The spectrum of baseband signal is located in low frequencies and has a low-pass form; that is, the spectrum starts from near zero frequency. The baseband signal can be divided into digital and analog baseband signal based on the characteristics of the original electrical signal.

In analog modulation, a sinusoidal signal is generally used as the carrier. Combined with the carrier, the spectrum of baseband signal is transformed for the transmission in the channel. For linear modulation, the spectrum of the modulated signal is a shift or linear transformation of the spectrum of the baseband signal (e.g., amplitude modulation). In nonlinear modulation (e.g., frequency and phase modulation), there is no linear relationship between the modulated signal and the baseband signal.

Digital modulation moves the spectrum of the baseband signal to a higher frequency band, which is more suitable for baseband signal transmission, and loads the baseband signal onto a parameter of the high-frequency carrier. The noise introduced by analog modulation in the transmission process cannot be completely eliminated at the receiver, but this is possible with digital modulation.

Based on the number of levels of the baseband signal, digital modulation can be divided into binary and multilevel digital modulation. Depending on the parameters of the information bearing carrier, digital modulation can be divided into amplitude-shift keying (ASK), frequency-shift keying (FSK), and phase-shift keying (PSK).

Additionally, other digital modulations can be formed based on the combinations of the mentioned schemes. For example, quadrature amplitude modulation (QAM) is a combination of ASK and PSK.

3.1.3 Direct and Indirect Modulation

Depending on the relationship between the modulated signal and the light source, optical modulation can be divided into direct and indirect modulation.

(1) Direct modulation

Direct modulation is also known as internal modulation, where the light source parameters, such as light intensity, are directly controlled by the modulation signal, and the modulated optical signal changed by the modulation signal is thus obtained. An advantage of direct modulation is that the circuit is simple and easy to implement; however, the transmission rate is limited.

(2) Indirect modulation

Indirect modulation is also known as external modulation, where an external modulator is used to modulate a certain parameter of the optical carrier. The modulated object is the light wave emitted by the light source, and the parameters of the light source remain unchanged during the modulation.

3.1.4 Internal and External Modulation

According to the relative relationship between the modulator and laser, modulation can be divided into internal and external modulation. Internal modulation refers to the loading of modulation signal in the process of laser oscillation. The modulation signal is used to change the oscillation parameters of the laser to change the characteristics of the laser output to achieve the modulation. Here, the “modulation signal” corresponds to the source in the communication.

In external modulation, a modulator is placed on the optical path outside the laser after the laser shape is generated, and the physical characteristics of the modulator are changed by the modulation signal. As the laser beam passes through the modulator, some parameters of the light wave are modulated. External modulation is usually implemented by electro-optic, acousto-optic, or magneto-optic crystal. Compared with internal modulation, external modulation is preferred owing to its higher modulation rate (by approximately one order of magnitude) and much wider modulation bandwidth.

3.2 External Modulation

In external modulation, a certain parameter of the light wave (output laser beam) is modulated by the modulation signal after the formation of the light wave, but the laser remains unchanged.

3.2.1 *Electro-Optic Modulation*

The electro-optic modulation [1] is a type of modulation scheme that uses an electro-optic crystal to make the amplitude, phase, and other parameters of the laser field change regularly with the modulation signal in the electric domain. Its physical basis is the electro-optic effect.

3.2.1.1 Electro-Optic Effect

The electro-optic [2] effect refers to the phenomenon that optical properties of materials change when the matter is placed in an electric field. Due to external electric field, some isotropic transparent materials exhibit optical anisotropy and refractive indexes of materials change. The electro-optic effects include Pockels effect [3] and Kerr effect [4].

The propagation law of light waves in a medium is affected by the refractive index distribution of the medium, which is closely related to its dielectric constant. The refractive index of a crystal can be expressed by the power series of the applied electric field; that is, $n = n_0 + \gamma E + bE^2 + \dots$, where the change of refractive index caused by γE is called the linear electro-optic effect or the Pockels effect. The change of the refractive index caused by the quadratic term bE^2 is called the Kerr effect. For most electro-optic crystal materials, the effect of γE is more significant than that of bE^2 [5].

(1) Pockels effect (or linear electro-optic effect)

According to the theory of crystal optics, the influence of the direction of electric field and light transmission on the refractive index of crystal is complex. Depending on the relationship between the applied electric field and the light direction, the Pockels effect can be classified into two categories: longitudinal Pockels effect (the electric field is parallel to the light direction) and transverse Pockels effect (the electric field is perpendicular to the light direction).

(a) Longitudinal Pockels effect

Potassium dihydrogen phosphate (KDP) crystal is a negative uniaxial crystal with light transmission band of 178 nm–1.45 μm . Considering the case of cutting perpendicular to the z-axis (optical axis) in the principal axis coordinate system consistent

with the axial direction of the crystal, when the electric field is incident parallel to the z-axis, the refractive index ellipsoid equation of the KDP crystal is [6]

$$\frac{x^2}{n_0^2} + \frac{y^2}{n_0^2} + \frac{z^2}{n_e^2} + 2\gamma_{63}xyE_z = 1, \quad (3.1)$$

where γ_{63} is the electro-optic coefficient of KDP crystal, and $n_o = n_x = n_y$, $n_e = n_z$ is the principal axis refractive index. To find a new coordinate system (x', y', z') , the refractive index ellipsoid equation does not contain any cross terms, where (x', y', z') is the direction of the principal axis of the ellipsoid after the electric field is applied; this is usually called the induction principal axis. Then, $n_{x'}$, $n_{y'}$, $n_{z'}$ is the principal refractive index in the new coordinate system. Since x-axis and y-axis are symmetrical, the x-coordinate and y-coordinate can be rotated by 45° around z-axis. In the new coordinate system (x', y', z') , Eq. (3.1) can be converted to [6]

$$\left(\frac{1}{n_0^2} + \gamma_{63}E_z\right)x'^2 + \left(\frac{1}{n_0^2} - \gamma_{63}E_z\right)y'^2 + \frac{1}{n_e^2}z'^2 = 1. \quad (3.2)$$

Thus, the three principal refractive indices along the principal axis of the new ellipsoid are [6]

$$\begin{cases} n'_x = n_o - \frac{1}{2}n_o^3\gamma_{63}E_z \\ n'_y = n_o + \frac{1}{2}n_o^3\gamma_{63}E_z \\ n_z = n_e. \end{cases} \quad (3.3)$$

It can be seen from Eq. (3.3) that the electric field parallel to the optical axis changes the KDP crystal from a uniaxial crystal to a biaxial crystal. The cross-section of the refractive index ellipsoid in the plane of $z = 0$ changes from a circle to an ellipse, and the length of its principal axis is related to the magnitude of the applied electric field E_z . As a result, the two equal amplitude and linearly polarized light beams vibrating in the directions of the induction principal axes x' and y' have different propagation speeds. The resulting phase difference is [6]

$$\delta = \frac{2\pi}{\lambda}(n'_x - n'_y)l = \frac{2\pi}{\lambda}n_o^3\gamma_{63}E_zl = \frac{2\pi}{\lambda}n_o^3\gamma_{63}U, \quad (3.4)$$

where λ is the wavelength in vacuum, l is the length of light passing through the crystal, and U is the applied voltage. It can be seen from Eq. (3.4) that the phase delay caused by the longitudinal electro-optic effect does not depend on the length l but is only determined by the nature of the crystal γ_{63} and the applied voltage U . In the electro-optic effect, the voltage required for the phase difference to reach π is called the half-wave voltage. The half-wave voltage of commonly used crystals is in the order of 3–10 kV.

(b) Transverse Pockels effect

In transverse Pockels effect, light propagates in the x' direction perpendicular to the z -axis of the electric field. The linearly polarized light beams vibrating along the two principal vibration directions z and y' have different propagation speeds. From Eq. (3.3), it can be seen that the phase difference produced after passing through a crystal with a length of l is [6]

$$\begin{aligned}\delta &= \frac{2\pi}{\lambda}(n_{x'} - n_z)l = \frac{2\pi}{\lambda}|n_o - n_e|l + \frac{\pi}{\lambda}n_o^3\gamma_{63}E_z l \\ &= \frac{2\pi}{\lambda}(n_{x'} - n_z)l = \frac{2\pi}{\lambda}|n_o - n_e|l + \frac{\pi}{\lambda}n_o^3\gamma_{63}\frac{l}{h}U,\end{aligned}\quad (3.5)$$

where h is the crystal thickness in the direction of electric field, and U is the applied voltage. The transverse electro-optic effect of the KDP crystal causes the phase difference of light wave passing through the crystal, including two items: The first is the phase delay caused by the natural birefringence of the crystal, which is independent of the applied electric field. The second is the phase delay caused by the applied electric field, which is related to the applied voltage U and the crystal size (l/h). Thus, the half-wave voltage can be reduced by appropriate selection of the crystal size.

(2) Kerr effect (quadratic electro-optic effect)

When a linearly polarized light passes through the crystal in the direction perpendicular to the electric field, it is decomposed into two linearly polarized lights that vibrate along and perpendicular to the electric field, respectively. The phase delay of the two linearly polarized lights is proportional to the square of the electric field intensity [7], which is called the Kerr effect:

$$\delta = 2\pi Kl \frac{U^2}{h^2}, \quad (3.6)$$

where K is the Kerr constant of matter, h is the distance between plates, l is the length of light passing through the medium, and U is the applied voltage. The half-wave voltage of Kerr response is generally of magnitude of tens of thousands of volts.

3.2.1.2 Electro-Optic Intensity Modulation

The intensity of light can be controlled by the electro-optic effect of the crystal. Figure 3.2a is a schematic diagram of a typical electro-optic intensity modulation device. It consists of two polarizers with vertical polarization direction and a uniaxial electro-optic crystal placed between them. The vibration direction of the polarizer is parallel to the x - and y -axis.

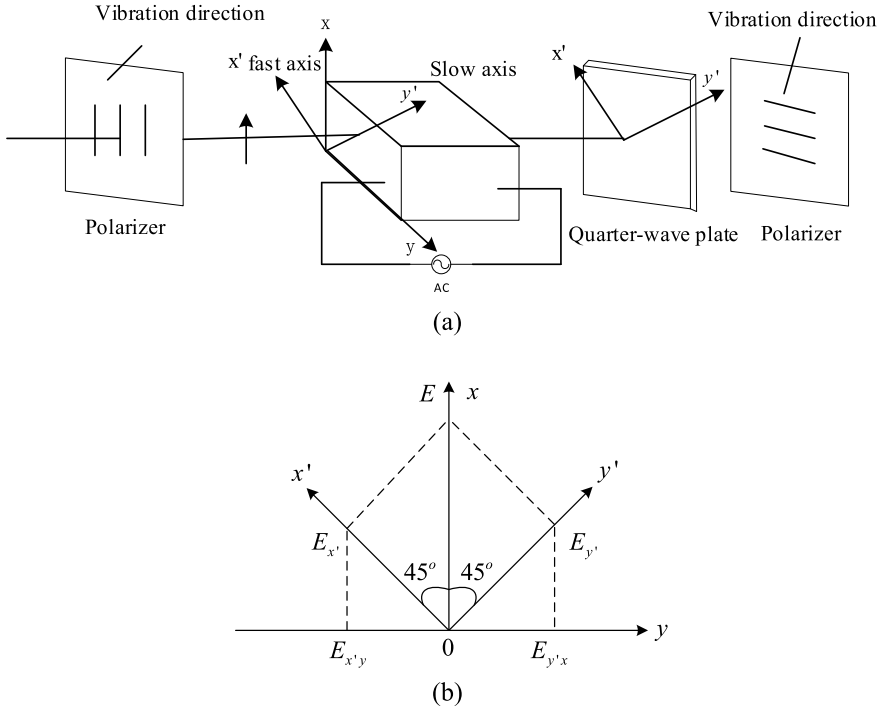


Fig. 3.2 Schematic diagram of light intensity modulation device: **a** light intensity modulator and **b** polarizer

When an electric field is applied along the z -axis direction on the electro-optic crystal, the induced birefringence axis generated by the electro-optic effect forms a 45° angle with the axis. Assuming that x' is the fast axis and y' the slow axis, if the voltage applied to the electro-optic crystal at a certain time is U , the amplitude of the electric vector of the linearly polarized laser in the x direction incident on the crystal is E , and the amplitudes of the electric vectors decomposed on x' and y' axes are $E_{x'}$ and $E_{y'}$. After passing through the crystal, the electric vector amplitudes along the x' and y' axes are both $E_{x'} = E_{y'} = (\sqrt{2}/2)E$. At the same time, the two polarized lights vibrating in x' and y' directions produce a phase difference as expressed by Eq. (3.8).

The two linearly polarized lights emitted from the crystal are then analyzed by a polarizer whose vibration direction is parallel to the y -axis. The resulting light amplitudes (see Fig. 3.2b) are $E_{x'y}$ and $E_{y'y}$, respectively, where $E_{x'y} = E_{y'y} = E/2$. Thus, the phase difference between them is $(\delta + \pi)$. The combined amplitude of these two vibrations is [7]

$$E'^2 = E_{x'y}^2 + E_{y'y}^2 + 2E_{x'y}E_{y'y}\cos(\delta + \pi)$$

$$\begin{aligned}
&= \frac{1}{4}(E^2 + E^2) - \frac{1}{2}E^2 \cos \delta \\
&= \frac{1}{2}E^2(1 - \cos \delta).
\end{aligned} \tag{3.7}$$

Since the light intensity is proportional to the square of the amplitude, assuming that the scale coefficient is 1, the light intensity passing through the polarizer can be written as [7]

$$I = E'^2 = E^2 \sin^2 \frac{\delta}{2} = I_0 \sin^2 \frac{\delta}{2}.$$

Namely,

$$I = I_0 \sin^2 \frac{\pi n_0^3 \gamma_{63}}{\lambda} U. \tag{3.8}$$

When the voltage U applied to the crystal changes, the light intensity passing through the polarizer also changes. A part of the $I/I_0 \sim U$ curve and the working situation of the light emphasis system is shown in Fig. 3.2. To select the working point in middle of the curve, a DC bias half-wave voltage $U_{\lambda/2}$ is usually applied to the modulation crystal. Alternatively, it is more convenient to insert a $\lambda/4$ wave plate in the device (see the dotted line in Fig. 3.2a) to add a fixed phase difference of $\pi/2$ between the vibration components along x' and y' . Then, if the applied signal voltage is a sinusoidal voltage (smaller voltage amplitude), $U = U_0 \sin \omega t$, the output light intensity is approximately sinusoidal. This result can be expressed as follows. Because of the additional fixed position difference $\pi/2$, δ in Eq. (3.8) should be replaced by $\Delta = \delta + \pi/2$ as [7]

$$\begin{aligned}
I &= I_0 \sin^2 \frac{\Delta}{2} = I_0 \sin^2 \left[\frac{\pi}{4} + \frac{\pi}{2} \frac{U_0}{U_\pi} \sin \omega t \right] \\
&= I_0 \cdot \frac{1}{2} \left[1 + \sin \left(\pi \frac{U_0}{U_\pi} \sin \omega t \right) \right].
\end{aligned} \tag{3.9}$$

In general, $U_0 \ll U_\pi$, the sine function can be expanded into series, and the first term can be obtained approximately as [7]

$$I/I_0 = \frac{1}{2} + \frac{\pi}{2} \frac{U_0}{U_\pi} \sin \omega t. \tag{3.10}$$

The relative light intensity remains a sine function of angular frequency ω , which is a linear copy of the modulation voltage, to achieve light intensity modulation.

3.2.1.3 Electro-Optic Phase Modulation

Consider the phase modulation device shown in Fig. 3.3. Assuming that the vibration direction of the polarizer is parallel to the y' -axis of the crystal, the vibration direction of the polarized light perpendicularly incident on the $x'y'$ plane of the crystal is parallel to the y' direction. In this case, the electro-optic effect produced by the external electric field no longer modulates the light intensity but changes the phase of the polarized light. After applying an electric field, the light whose vibration direction is parallel to the y' -axis of the crystal passes through the crystal with a length of l , and its phase increases to [7]

$$\Phi = \frac{2\pi}{\lambda} \left(n_0 + \frac{n_0^3}{2} \gamma_{63} E_z \right) l. \quad (3.11)$$

If the sinusoidal modulation electric field $E_z = E_m \sin \omega_m t$ is applied to the crystal (where E_m and ω_m are the amplitude and angular frequency of the modulation field, respectively), and amplitude of the field vector of the light at the input surface of the crystal ($z = 0$) is $U_{in} = A \cos \omega t$, then the amplitude of the field vector at the output surface ($z = l$) can be written as [7]

$$U_{out} = A \cos \left[\omega t + \frac{2\pi}{\lambda} \left(n_0 + \frac{n_0^3}{2} \gamma_{63} E_z \right) l \right]. \quad (3.12)$$

Using the sinusoidal modulated electric field and omitting the constant term, Eq. (3.12) is rewritten as [7]

$$U_{in} = A \cos(\omega t + M_P \sin \omega_m t). \quad (3.13)$$

Here, $M_P = \frac{\pi n_0^3}{\lambda} \gamma_{63} E_m l$ is called phase modulation degree. It can be seen from Eq. (3.13) that the phase of the output field is modulated by the electric field with modulation degree of M_P and angular frequency of ω_m .

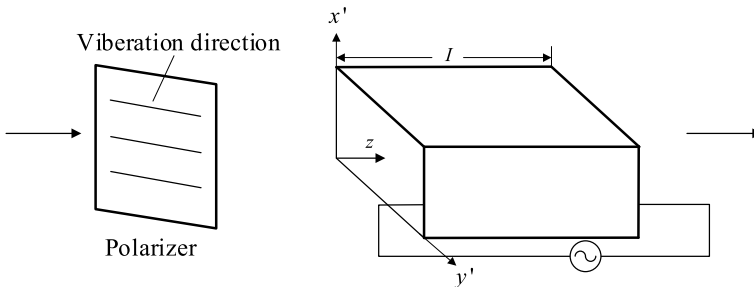


Fig. 3.3 Schematic diagram of phase modulation [7]

3.2.2 Acousto-Optic Modulation

The acousto-optic modulation [8, 9] is a type of modulation method that uses acousto-optic effect of acousto-optic medium to make diffraction intensity, frequency, and direction change regularly with acoustic wave intensity. Its physical basis is the acousto-optic effect.

3.2.2.1 Acousto-Optic Effect

When an ultrasonic wave passes through a medium, elastic strain will be produced due to local compression and elongation of the medium. The strain changes periodically with time and space. When light passes through a medium disturbed by an ultrasonic wave, diffraction will occur. The intensity, frequency, and direction of the diffracted light will change with the change of the ultrasonic field. This phenomenon is called the acousto-optic effect. Raman–Nath diffraction and Bragg diffraction are two common acousto-optic effects. The parameters to measure these two types of diffraction are [6]

$$Q = 2\pi L \frac{\lambda}{\lambda_s^2}, \quad (3.14)$$

where L is the length of acousto-optic interaction, λ is the wavelength of light passing through the acousto-optic medium, and λ_s is the ultrasonic wavelength. When $Q \leq 0.3$, it is the Raman–Nath diffraction, and when $Q \leq 4\pi$, it is the Bragg diffraction. In the middle zone of $0.3 < Q < 4\pi$, the diffraction phenomenon is more complicated, and ordinary acousto-optic devices do not work in this range [6].

3.2.2.2 Raman–Nath Diffraction

As shown in Fig. 3.4, when the ultrasonic frequency is low and the incident light wave parallel to the acoustic wave surface (i.e., perpendicular to the propagation direction of the acoustic field), the length of the acousto-optic interaction is short, and the change of the refractive index can be ignored. Then, the acousto-optic medium can be regarded as a relatively static “plane phase grating”. The speed of sound is much smaller than the speed of light, and the length of a sound wave is much greater than that of a light wave. Therefore, when a light wave passes through the medium in parallel, the wave front of the light passing through the denser part (with a larger refractive index) will lag, while the wave front passing through the sparser part (with a smaller refractive index) will advance. Thus, the wave front of a plane wave passing through acousto-optic medium appears as a concave-convex phenomenon and becomes a wrinkle surface. The secondary waves emitted by the sub-wave sources on the outgoing wave front will interfere coherently to form multi-level diffracted light symmetrically distributed with the incident direction, which is

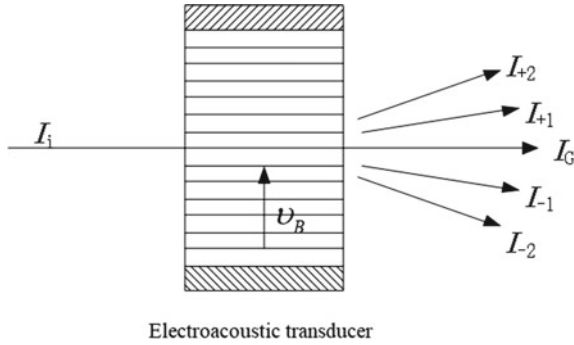


Fig. 3.4 Raman–Nath diffraction [6, 10]

called the Raman–Nath diffraction. The diffraction angle of diffracted light θ at all levels satisfies the following relationship [6]:

$$\lambda_s \sin \theta = m\lambda \quad m = 0, \pm 1, \pm 2 \dots \quad (3.15)$$

In diffraction, maximum points associated with $m = 0, \pm 1, \pm 2 \dots$ exist on both sides of the incident light, and the diffracted light will produce a Doppler effect for the moving acoustic wave. Thus, the frequency of the response light wave is $\omega, \omega \pm \omega_s, \omega \pm 2\omega_s, \dots$ where the zeroth order diffracted light is an extension of the incident light. The extreme light intensity corresponding to the m th order diffraction is [6]

$$I_m = I_i J_m^2(\delta), \quad (3.16)$$

where I_i is the incident light intensity, $\delta = 2\pi(\Delta n)L$ represents the additional phase shift caused by the change of the refractive index after the light passes through the acousto-optic medium, and $J_m(\delta)$ is the Bessel function of the m th order.

3.2.2.3 Bragg Diffraction

When the frequency of an acoustic wave is high, the working length of the acoustic wave is large, and the incident angle between the beam and the acoustic wave surface is oblique, a light wave in the medium will propagate through multiple acoustic wave planes, and the medium has the “volume grating” property. As shown in Fig. 3.5, when the angle between the incident light and the acoustic surface meets certain conditions, the diffracted light at all levels in the medium will interfere with each other, and the diffracted light of higher order will cancel each other. Only the diffraction light of order 0, +1, or −1 (depending on the direction of the incident light) will appear. This is referred to as the Bragg diffraction. If suitable parameters can be selected and

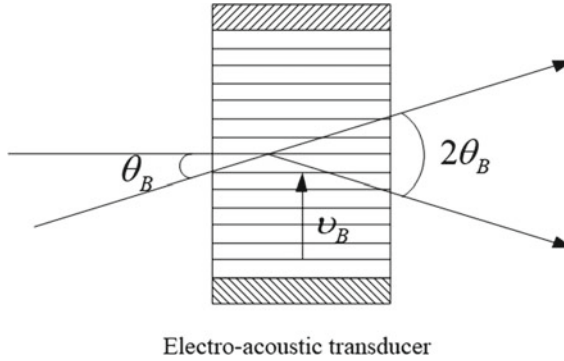


Fig. 3.5 Bragg diffraction [6, 10]

the ultrasonic field is strong enough, almost all energy of the incident light will be transferred to the diffraction maximum of $+1$ or -1 st order. In this way, the energy of the beam will be fully utilized, and the acousto-optic devices made using the Bragg diffraction effect can achieve higher efficiency.

It can be proved that the frequency of the ± 1 order diffraction light of the Bragg diffraction is $\omega \pm \Omega$, and the corresponding zeroth and first order diffraction intensities are [6]:

$$\begin{aligned} I_0 &= I_i \cos^2\left(\frac{\delta}{2}\right) \\ I_1 &= I_i \sin^2\left(\frac{\delta}{2}\right), \end{aligned} \quad (3.17)$$

where δ is the additional phase shift caused by the change of refractive index after light passes through the acousto-optic medium. It is shown that $I_0 = I_1$ when $\delta/2 = \pi/2$. This indicates that the incident power can be converted into the first order diffraction power by properly controlling the incident ultrasonic power.

3.2.2.4 Acousto-Optic Modulation

Acousto-optic modulation is a physical process in which the information is loaded onto an optical frequency carrier by using acousto-optic effect. The modulation signal acts on the electroacoustic transducer in the form of amplitude modulation of electrical signal. The electroacoustic transducer converts the corresponding electrical signal into a variable ultrasonic field. When the light wave passes through the acousto-optic medium, the optical carrier is modulated and becomes an intensity modulated wave carrying information. There are two types of acousto-optic modulators. The Raman–Nath acousto-optic modulator is characterized by a working acoustic source

with frequency of less than 10 MHz, and thus is limited to low frequency operation and narrow bandwidth. The Bragg acousto-optic modulator is characterized by high diffraction efficiency and wide modulation bandwidth.

3.2.3 Magneto-Optic Modulation

The magneto-optic modulation [10] is a type of modulation method that uses the magneto-optic effect of magneto-optic medium to change the polarization direction of linearly polarized light and uses the relative position of polarizer and analyzer to detect the change of light intensity. Its physical basis is the magneto-optic effect [6, 11].

3.2.3.1 Magneto-Optic Effect

In 1811, during the study on the birefringence characteristics of a quartz crystal, Arago found that when a beam of linearly polarized light propagates along the optical axis of a quartz crystal, its vibration plane will turn an angle relative to the original direction, as shown in Fig. 3.6. Because the quartz crystal is a uniaxial crystal, light will not be birefringent when it propagates along the optical axis. Therefore, the phenomenon discovered by Arago was a new phenomenon called optical rotation [10, 12].

In 1846, Faraday discovered that under the action of a magnetic field, a non-rotatory medium also produces optical rotation, which can make the vibrating surface of linearly polarized light rotate. This is the Faraday effect. The structure of the device for observing the Faraday effect is shown in Fig. 3.7. Both ends of a glass rod are polished and placed into the magnetic field of a solenoid. In addition, polarizers P1 and P2 are added to make the beam pass through the polarizer along the magnetic

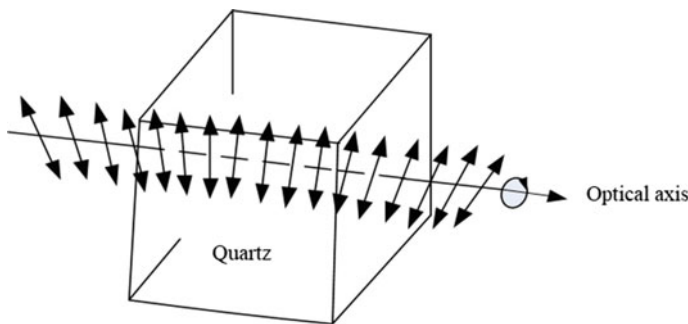


Fig. 3.6 Optical rotations [10]

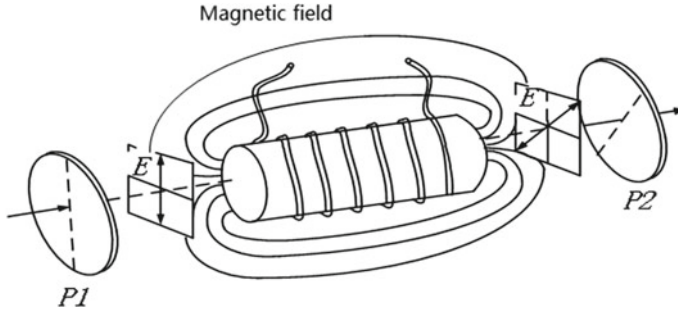


Fig. 3.7 Faraday effect [6, 10]

field direction. The direction of the light vector will rotate, and the rotation angle can be measured by the polarizer.

Later, Verdet studied the Faraday effect and found that the rotation angle of the light vibration plane θ can be expressed as

$$\theta = VBl, \quad (3.18)$$

where V is a constant related to the properties of the matter called Verdet constant, B is the magnetic induction, and l is the length of the light passing through the matter.

3.2.3.2 Magneto-Optic Modulation

The principle of magneto-optic modulation [13] is shown in Fig. 3.8. When there is no modulation signal, there is no external magnetic field in the magneto-optic material. According to Marius' law [6, 7, 10], the intensity of the light beam transmitted from the polarizer is I_0 , and the intensity of the light emitted by the analyzer is

$$I = I_0 \cos^2 \alpha, \quad (3.19)$$

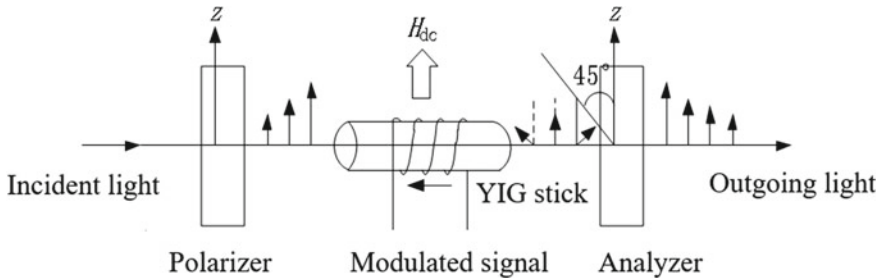


Fig. 3.8 Schematic diagram of magneto-optic modulation

where α is the angle between the polarizer and the optical axis of the polarizer. When the two optical axes are parallel ($\alpha = 0$), the passing light intensity is the maximum; when the optical axes of the two polarizers are perpendicular to each other ($\alpha = \pi/2$), the passing light intensity is zero (extinction). When the modulated AC signal is added to the magnetizing coil outside the magneto-optic material, the alternating magnetic field generated causes the vibration surface of light to rotate alternately by angles of size θ . The output light intensity is

$$I = I_0 \cos^2(\alpha + \theta). \quad (3.20)$$

When α is constant, the output light intensity only varies with θ . Because of the Faraday effect, the signal current makes the rotation of the optical vibration surface into the intensity modulation of light, so information transmission can be performed by using this phenomenon.

3.3 Reverse Modulation

“Passive modulation technology,” also known as reverse modulation technology, uses the method of changing the reverse echo power for modulation, eliminating the tracking and pointing system of the traditional optical wireless communication system and making the system application more flexible [14]. The cat’s eye reverse modulator is a reverse modulation device designed based on the principle of the cat’s eye effect. By combining a reflection modulation device and cat’s eye structure, the echo power of the incident light irradiated to the cat’s eye structure can be modulated by changing the defocus of the modulation device to achieve the purpose of communication.

3.3.1 Cat’s Eye Effect

The cat’s eyes look very bright in the dark because the incident light from the light source is focused on the fundus through the cat’s pupil. The reflection of the fundus causes the light beam to return along its original path. The reflected light is projected into the eye of the observer, and the cat’s eye looks brighter [15–17].

As shown in Fig. 3.9, the working principle of the cat’s eye structure model is as follows. A beam of light parallel to the cat’s eye structure converges at the focal plane through the action of the focusing lens L into a point at the focus F. It is assumed that the plane mirror is located at the focal plane, and the converging beam returns along the original optical path after being reflected by the reflector at the focal plane. Therefore, the light transmitter is the same as the light receiver. The focusing lens L can be regarded as the entrance and exit pupil of the cat’s eye structure. The plane

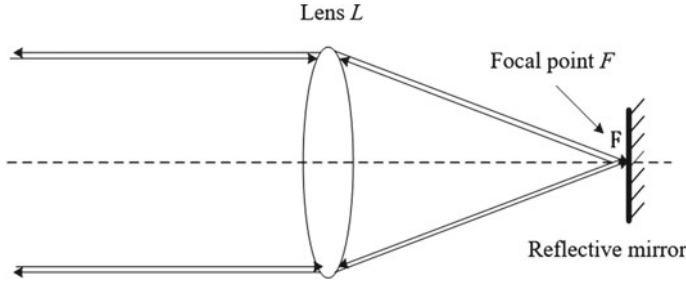


Fig. 3.9 Physical model of cat's eye structure

mirror is located at the focal plane of the focusing lens. The working process is analyzed as follows.

The Fourier transform model of cat's eye effect is shown in Fig. 3.10 [18, 19], where O is the object plane, T is the transformation plane, I is the image plane, (x', y') (u, v), and (x', y') are the spatial coordinates of the three planes, and $\tilde{t}_0(x, y)$ and $\tilde{t}_T(u, v)$ represent the optical amplitude transmissivity of the object plane and transformation plane, respectively. Finally, f represents the focal length of the lens, and F represents the Fourier transform. Denoting the amplitude of the initial incident light field by A_0 , the amplitude distribution on the object plane is

$$\tilde{U}_0(x, y) = A_0 \tilde{t}_0(x, y). \quad (3.21)$$

Let the amplitude transmissivity on the transformation plane T be $\tilde{t}_T = \rho$ ($0 < \rho < 1$; this is the optical reflectivity on the photosensitive surface). The complex amplitude distribution $\tilde{U}_I(x', y')$ on the image plane I can be considered as having undergone two Fraunhofer diffractions, and the transformation of the complex amplitude is a Fourier transform. Thus, the complex amplitude distribution on the image plane I is [20]

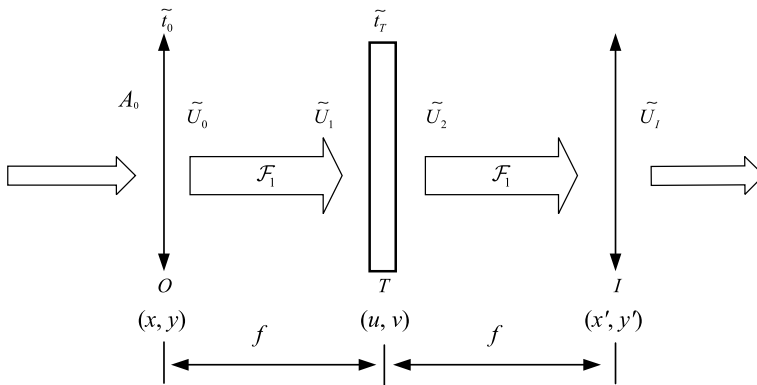


Fig. 3.10 Fourier transform model of cat's eye effect

$$\tilde{U}_I(x', y') = \int \int \int_{-\infty}^{+\infty} A_0 \rho \tilde{t}_0(x, y) \exp \left\{ \frac{-ik}{f} [u(x + x') + v(y + y')] \right\} du dv dx dy,$$

$$\tilde{U}_I(x', y') = \rho A_0 \tilde{t}_0(-x', -y').$$

That is,

$$\tilde{U}_I(x, y) = \rho \tilde{U}_0(-x, -y). \quad (3.22)$$

It can be seen from Eq. (3.22) that the output image is exactly the same as the input image, and the negative sign in the formula represents the inversion of the image. This shows that the light returns along the original path, and only the amplitude decreases $0 < \rho < 1$.

The cat's eye structure is often used as a retroreflector. In combination with a suitable modulation device, a reverse modulator can be formed. Because the cat's eye structure reverse modulator is passively controlled by the light source, only when the light source irradiates the cat's eye structure will the incident light be reflected back to the light source along the original path. In the process of reflection, the reflected light is modulated, and useful information is carried back to the light source; this is also called passive modulation.

3.3.2 Principle of Reverse Modulation

Defocusing can be divided into forward and backward defocusing. When the reflection at the focal plane moves towards the lens, it is called forward defocusing, and vice versa.

As shown in Fig. 3.11 [21, 22], the aperture of the focusing lens is D , the focal length is f , and the defocusing amount is d . The output beam is limited to a certain aperture and angle range: if beyond this range, some light will not be able to exit the

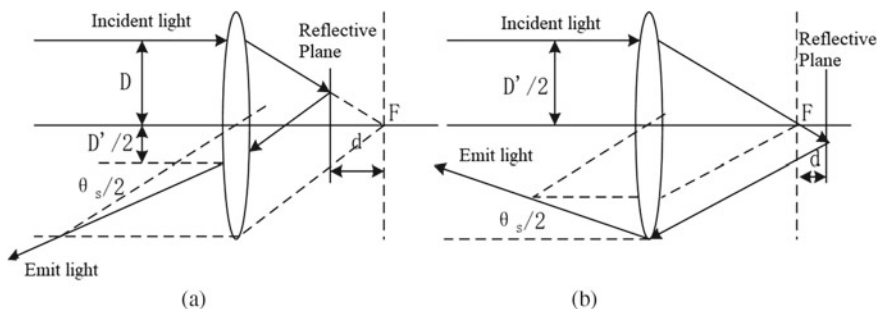


Fig. 3.11 Forward defocus (a) and backward defocus (b)

optical system. We call these the effective aperture D' and divergence angle θ_s of defocusing.

For forward defocusing, the effective aperture is

$$D' = \frac{f - 2d}{f} D \quad (d \ll f), \quad (3.23)$$

and the divergence angle is

$$\theta_s = 2tg^{-1} \frac{dD}{f^2}. \quad (3.24)$$

For backward defocusing, the effective aperture is

$$D' = \frac{f}{f + 2D} D (d \ll f), \quad (3.25)$$

and the divergence angle is

$$\theta_s = 2tg^{-1} \frac{dD}{f(f + 2d)}. \quad (3.26)$$

It can be seen from Eqs. (3.23) and (3.25) that the defocusing amount of reflector is inversely proportional to the effective aperture of the cat's eye.

Suppose that the laser transmitting power is p_t , the effective receiving area is A_r , the distance between the target and the laser transmitting receiver is r , the divergence angle of the beam is θ_0 , the effective cross-sectional area of the cat's eye optical system is A_s , the reflection coefficient is ρ_s , the atmospheric transmittance is τ_a , and the transmittance of the receiving optical system is τ_r . Then the reflected echo power of the cat's eye target received by the receiver is

$$p_r = p_t \frac{16\tau_a \tau_r A_r A_s \rho_s}{\pi^2 \theta_0^2 \theta_s^2 r^4}. \quad (3.27)$$

By introducing $A_r = \pi(D/2)^2$, $A_s = \pi(D'/2)^2$, and Eqs. (3.23)–(3.26) into Eq. (3.27), the reflected power of the cat's eye target can be simplified as

$$p_r = p_t \tau_a \tau_r \rho_s \frac{f^4 D^2}{4\theta_0^2 d^2 r^4}. \quad (3.28)$$

According to Eq. (3.28), the reflected power of cat's eye reverse modulator is related to many factors of the system. It is directly proportional to the fourth power of the lens focal length and the square of the lens diameter and inversely proportional to the square of the defocusing amount and the fourth power of the distance between

the transmitter and the modulation end. When the structural parameters f and D and the distance r between the transmitter and the modulator are fixed, the variable defocus is the most ideal and easiest to realize parameter for modulating the reflected power.

When the mirror is in the focal plane, the incident beam will return to the original path strictly, but it will diverge when out of focus. A modulating retroreflector (MRR) is a method to modulate the reflected light power by adjusting the defocusing amount of the cat's eye structure. The relationship between the defocusing state of the cat's eye reverse modulator and the output power of MRR is shown in Fig. 3.12. The change of power is directly reflected in the change of signal amplitude at the receiving end, which is convenient for the extraction and identification of the signal.

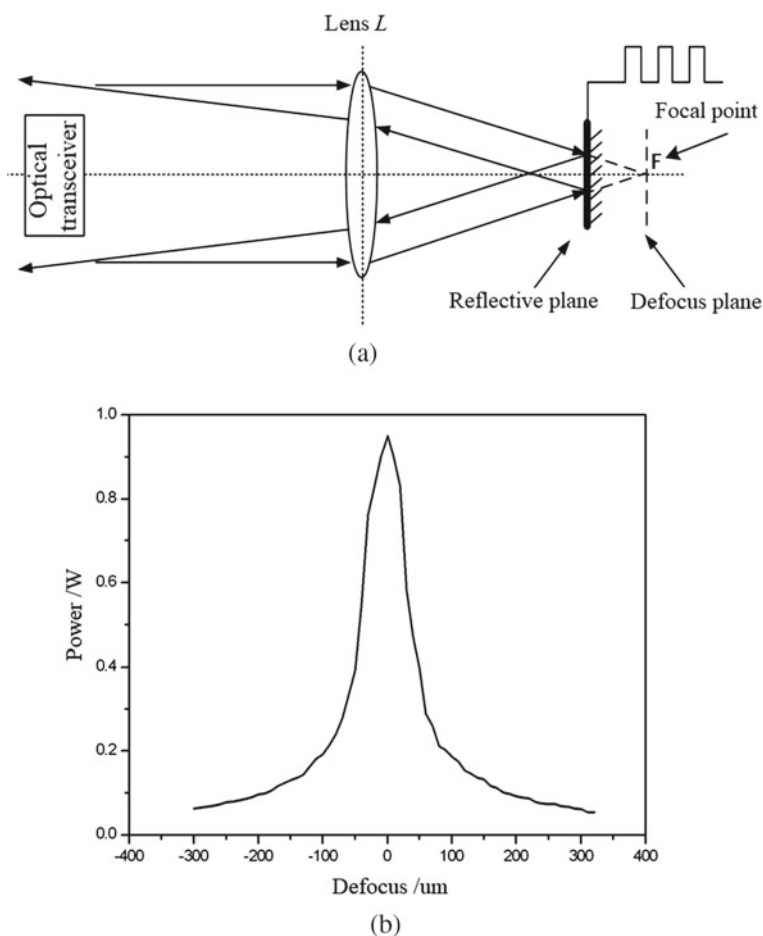


Fig. 3.12 Defocusing principle of the cat's eye effect: **a** diagram of the cat's eye effect defocusing principle and **b** defocusing distance and power curve

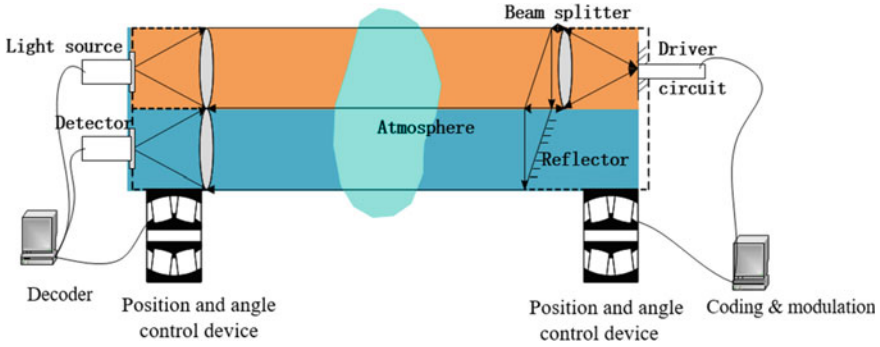


Fig. 3.13 Diagram of modulating retroreflector system

3.3.3 Cat's Eye Reverse Modulation System

The MRR system consists of two unequal terminals [23–25]: the active and passive terminals. A schematic diagram of an MRR system is shown in Fig. 3.13. The active end is the transmitting/receiving terminal of traditional free-space optical communication (FSO) and includes the laser transmitting, laser receiving, signal processing, and control systems. The passive end is the reverse modulation terminal and includes the MRR, information acquisition, signal processing, and control systems; the MRR system includes modulator and retroreflector.

The working process of MRR FSO system is as follows. The transmitting/receiving terminal aims at the reverse modulation terminal through the control system and emits a laser beam. The reverse modulation terminal receives the incident beam by adjusting the angle and position of the controller. When the reverse modulation terminal detects the incident beam, the signal processing system loads the modulation signal to the modulator and modulates the incident light through the change of the modulator state. Then, the reverse reflector returns the modulated beam to the transmitter. The transmitter/receiver obtains the modulated optical signal through the receiving system, converts the optical signal into electrical signal, and finally processes the signal by the signal processing system to demodulate the useful signal.

3.4 Pulse-Like Position Modulation

3.4.1 Pulse-Like Position Modulation

Pulse-like position modulation is a general term of pulse position modulation and includes on-off keying (OOK), pulse position modulation (PPM), digital pulse interval modulation (DPIM), and dual-header pulse interval modulation (DHPIM).

These modulation methods use the time interval between the optical pulse and the reference point as the information carrier for information transmission. The information receiver determines the transmitted information by the position of the optical pulse during a certain period of time; this is referred to as pulse-like position modulation. OOK, PPM, and multichannel subcarrier modulation are the modulation methods for intensity modulation/direct detection in the atmospheric optical wireless communication system. Among them, OOK modulation has higher average transmit power, whereas PPM has a reduced average transmit power but increased bandwidth requirement. The symbol structure, bandwidth requirement, average transmit power, slot error rate, and channel capacity of the mentioned modulation methods are systematically analyzed as follows [26–30].

3.4.1.1 Symbol Structure of Pulse-Like Position Modulation

The OOK modulation is the most widely used and the simplest modulation method in intensity modulation/direct detection system. It uses the presence or absence of optical pulse to transmit information. When an optical pulse in a time slot is detected, transmission information of “1” is recorded; no optical pulse means transmission information of “0”.

PPM maps a group of n -bit binary data into a single pulse signal in a time slot composed of 2^n time slots, where the pulse position is the decimal number corresponding to the binary data.

Multi-pulse position modulation (MPPM) maps the binary data of length M into a symbol of optical pulse, which appears simultaneously in P time slots of an information frame with n time slots; M , n , and P satisfy $C_n^P \geq 2^M$.

Differential pulse position modulation (DPPM) removes “0” after “1” based on PPM, so its symbol length is not fixed.

Dual duration PPM is an improved form of PPM. The symbol length is $2^{M-1} + \alpha - 1$ time slots, and k is the decimal number corresponding to the binary data. If $k < 2^{M-1}$, the pulse is located at the $(k + 1)$ -th time slot, and the pulse width in time slots is $\alpha/2$. If $k \geq 2^{M-1}$, the pulse is located at the $(k + 1)$ -th time slot, and the pulse width in time slots is α .

Dual amplitude PPM uses dual amplitude signals to distinguish the first and second halves of the information. If $k < 2^{M-1}$, the pulse amplitude is A , and 2^{M-1} time slots behind the PPM mapping are removed. If $k \geq 2^{M-1}$, the pulse amplitude is βA , and 2^{M-1} time slots in front of the PPM mapping are removed. The pulse position modulation mode is the same as that of PPM.

Shorten pulse position modulation divides binary data into two parts: the first bit and last $(M - 1)$ bits. Specifically, the first bit remains unchanged, and the last $(M - 1)$ bits are modulated according to the mapping method of the PPM mode. Then the two parts of the data are combined as the modulated data. During demodulation, the two parts of the information are demodulated separately.

Separated double PPM (SDPPM) is based on MPPM and improved with 2 pulses. Assuming that the length of each symbol is N , the number of pulse combinations is C_N^2 .

To avoid inter-symbol interference, the combination of continuous pulses is removed when selecting the pulse combination. Therefore, the number of pulse combinations available for selection is $C_{N-1}^2 - (N - 2)$ satisfying $C_{N-1}^2 - (N - 2) \geq 2^M$.

Pulse interval modulation (PIM) uses the space time slot interval between adjacent optical pulses to represent the transmitted information. The pulse position in the modulation symbol is fixed at the initial position of the symbol, which is called the starting pulse. It is followed by several protection time slots and K space time slots representing the transmitted information. Here, K is the decimal number corresponding to the transmitted binary data. DPIM is a type of PIM, and its symbol structure is approximately the same as that of PIM. When the receiver demodulates the DPIM, symbol synchronization is not required.

DHPIM is an improved form of PIM, with a more complex modulation structure. In DHPIM, each symbol consists of a header slot and subsequent empty slots. The header slot is fixed as $\alpha + 1$ time slots (α is a positive integer), and two cases can occur: if $k < 2^{M-1}$, the head time slot is $(\alpha/2)$ pulse time slots plus $(\alpha/2 + 1)$ time slots for the protection slot; the number of subsequent empty time slots is k , indicating the information to be transmitted. If $k \geq 2^{M-1}$, the head time slot is α pulse time slot and a guard time slot, and the number of subsequent empty time slots is $(2^M - 1 - k)$.

Dual pulse interval modulation (DPPIM) uses a fixed start pulse, a variable marker pulse, and the time interval between the start pulse and the marker pulse to mark the transmitted information. The start pulse width is fixed to one time slot, and the marker pulse change is as follows: if $k < 2^{M-1}$, the marker pulse width is α time slots, and the number of empty time slots between the start pulse and marker pulse is $k - 2^{M-1}$. Finally, a number of empty time slots are added after the marker pulse to ensure that the length of the DPPIM symbol is fixed.

The symbol of dual amplitude pulse interval modulation consists of a start pulse, a guard slot, and m empty slots. The amplitude of the start pulse changes as follows: if $k < 2^{M-1}$, the amplitude of the start pulse is A , and the number of information slots is $m = k$; if $k \geq 2^{M-1}$, the amplitude of the start pulse is βA (β is a positive number), and the number of information slots is $m = k - 2^{M-1}$.

In fixed length digital pulse interval modulation (FDPIM), each symbol consists of a single time slot pulse fixed at the start position, a guard time slot, an information time slot, a double time slot as the marker pulse, and subsequent $(2^M - k)$ empty time slots. The first empty time slot after the pulse is marked as the guard time slot. The other empty time slots do not indicate any information: they ensure a fixed symbol length.

The symbol length of fixed length dual amplitude PIM is fixed to $(2^M + 3)$ time slots. The symbol structure is similar to that of FDPIM, except that the amplitudes of the start pulse and marker pulse in this modulation method are A and βA , respectively, and each pulse is a single time slot. Taking the number of modulation bits $M = 4$ as an example, the symbol structure of various modulation methods is shown in Fig. 3.14.

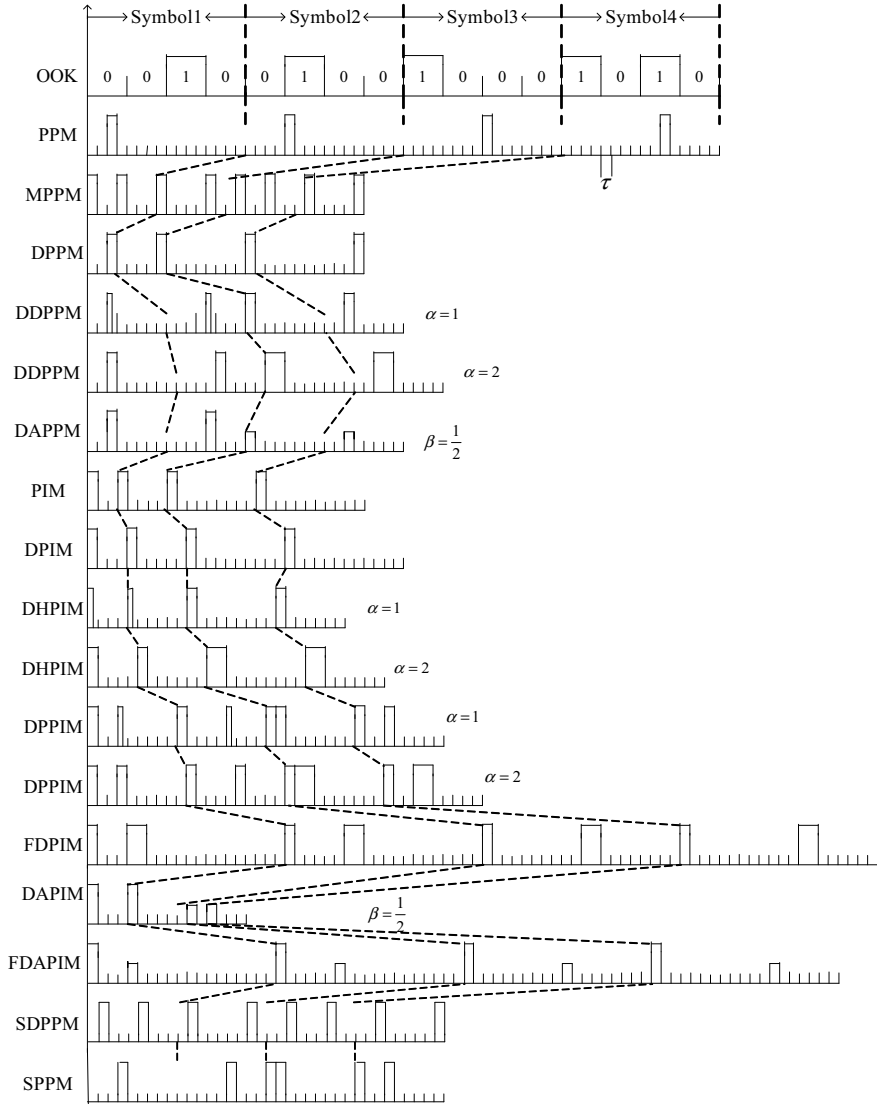


Fig. 3.14 The signal structure of various modulation modes with $M = 4$

3.4.1.2 Performance Analysis of Several Modulation Methods

OOK modulation is commonly used in optical wireless communication systems. Because of its anti-interference inability and low power utilization, PPM, DPIM, DHPIM, and other modulation methods have been proposed [31, 32].

(1) Average symbol length

The symbol length of each modulation mode refers to the number of time slots in an information frame. Some of the symbol lengths of the modulation methods are fixed, and some of them vary depending on the transmission data.

According to the symbol structure of the types of pulse-like position modulation, the average symbol length can be obtained, as shown in Table 3.1.

The symbol lengths of MPPM and SDPPM satisfy $C_{nm}^p \geq 2^M$ and $C_{ns-1}^2 - (ns - 2) \geq 2^M$, respectively, where M represents the number of information bits, T_s represents the slot width, and α represents the pulse width parameter.

(2) Bandwidth requirements and utilization

The optical wireless communication system requires a certain bandwidth to transmit information, and the smaller the bandwidth the better. Assuming that the transmission rate of the source is R_b bit/s, the pulse width duty cycle is 1, and each symbol sends M bits of information, the bandwidth of each modulation method can be estimated as the reciprocal of the slot width. The bandwidth requirements of various modulation methods are shown in Table 3.2.

Table 3.1 Average symbol length

Modulation mode	OOK	PPM	DPPM
Average symbol length	MT_s	$2^M T_s$	$(2^M + 1)T_s/2$
Modulation mode	DPIM	DAPPM	DPPIM
Average symbol length	$(2^M + 3)T_s/2$	$2^{M-1} T_s$	$(2^{M-1} + \alpha)T_s$
Modulation mode	FDPIM	DHPIM	SPPM
Average symbol length	$(2^M + 3)T_s$	$(2^{M-1} + 2\alpha + 1)T_s/2$	$(2^{M-1} + 1)T_s$
Modulation mode	DAPIM	FDPIIM	DDPPM
Average symbol length	$(2^{M-1} + 3)T_s/2$	$(2^M + 4)T_s$	$(2^{M-1} + \alpha - 1)T_s$

Table 3.2 Bandwidth requirements

Modulation mode	OOK	PPM	DPPM
Bandwidth requirements	R_b	$\frac{2^M}{M} B_{OOK}$	$\frac{2^M + 1}{2M} B_{OOK}$
Modulation mode	DPIM	DHPIM	DDPPM
Bandwidth requirements	$\frac{2^M + 3}{2M} B_{OOK}$	$\frac{2^{M-1} + 2\alpha + 1}{\alpha M} B_{OOK}$	$\frac{2^M + 2\alpha - 2}{\alpha M} B_{OOK}$
Modulation mode	DPPIM	DAPPM	DAPIM
Bandwidth requirements	$\frac{2^M + 2\alpha}{\alpha M} B_{OOK}$	$\frac{2^{M-1}}{M} B_{OOK}$	$\frac{2^{M-1} + 3}{2M} B_{OOK}$
Modulation mode	SPPM	FDPIM	FDAPIM
Bandwidth requirements	$\frac{2^{M-1} + 1}{M} B_{OOK}$	$\frac{2^M + 4}{M} B_{OOK}$	$\frac{2^M + 3}{M} B_{OOK}$
Modulation mode	MPPM		
Bandwidth requirements	$\frac{n}{M} B_{OOK}$		

The bandwidth utilization can be defined as $\eta = R_b/B$, and the bandwidth utilization of various modulation methods can be obtained as shown in Table 3.3.

The bandwidth utilizations of SDPPM and MPPM are expressed in the same way, both are M/n , except that n of the former satisfies $C_{n-1}^2 - (n-2) \geq 2^M$, and n of the latter satisfies $C_n^p \geq 2^M$.

(3) Average transmit power

In atmospheric laser communication, the pulse-like position modulation method can be considered to send “1” and “0” sequences with an equal probability. Power P_c is needed when sending “1”, and no power is needed when sending “0”, so the average transmission power can be determined as $P_{ave} = p_1 P_c$. The average transmit power of various modulation methods is shown in Table 3.4.

D. Channel capacity

Channel capacity is the maximum information rate that the channel can transmit without errors. According to the relationship between the average received optical power and the average transmit power $P_R(h) = P_T(\eta A/\lambda L)^2 h$, the telecommunication noise ratio at the receiving end is $\gamma(h) = \eta_z^2 t^2 P_c^2 R(\eta A/\lambda L)^4 h^2 / 2\sigma_n^2$, and the average channel capacity under different turbulence channels is as follows.

Weak turbulence:

$$\langle C \rangle = \frac{B}{\ln 2 \cdot \sigma \sqrt{2\pi}} \int_0^\infty \ln \left(1 + \frac{\eta_z^2 t^2 P_c^2 R}{2\sigma_n^2} \left(\frac{\eta A}{\lambda L} \right)^4 h^2 \right) \cdot h^{-1} \exp \left[-\frac{(\ln h + \sigma^2/2)^2}{2\sigma^2} \right] dh. \quad (3.29)$$

Moderate strong turbulence:

$$\langle C \rangle = \frac{B \cdot 2^{\alpha+\beta-1}}{2 \ln 2 \cdot \pi \Gamma(\alpha) \Gamma(\beta)} G_{6,2}^{1,6} \left(\frac{8t^2 P_c^2 \eta_z^2 R}{\sigma_n^2 (\alpha\beta)^2} \cdot \left(\frac{\eta A}{\lambda L} \right)^4 \middle| \begin{matrix} 1, 1, \frac{1-\alpha}{2}, \frac{2-\alpha}{2}, \frac{1-\beta}{2}, \frac{2-\beta}{2} \\ 1, 0, \frac{2-\alpha-\beta}{4}, \frac{4-\alpha-\beta}{4} \end{matrix} \right). \quad (3.30)$$

Table 3.3 Bandwidth utilization

Modulation mode	OOK	PPM	DPPM
Bandwidth utilization	1	$M/2^M$	$2M/(2^M + 1)$
Modulation mode	DPIM	DHPIM	DDPPM
Bandwidth utilization	$2M/(2^M + 3)$	$\alpha M/(2^{M-1} + 2\alpha + 1)$	$\alpha M/(2^M + 2\alpha - 2)$
Modulation mode	DPPIM	DAPPM	DAPIM
Bandwidth utilization	$\alpha M/(2^M + 2\alpha)$	$M/2^{M-1}$	$2M/(2^{M-1} + 3)$
Modulation mode	SPPM	FDPIM	FDAPIM
Bandwidth utilization	$M/(2^{M-1} + 1)$	$M/(2^M + 4)$	$M/(2^M + 3)$

Table 3.4 Average transmit power

Modulation mode	OOK	PPM	DPPM
Average transmit power	$\frac{P_c}{2}$	$\frac{P_c}{2^M}$	$2M / (2^M + 1)$
Modulation mode	DPIM	DHPIM	DDPPM
Average transmit power	$\frac{2}{2^M+3} P_c$	$M / 2^{M-1}$	$\frac{1+\beta}{2^{M-1}+3} P_c$
Modulation mode	DPPIM	DAPPM	DAPIM
Average transmit power	$\alpha M / (2^M + 2\alpha)$	$M / 2^{M-1}$	$\frac{1+\beta}{2^{M-1}+3} P_c$
Modulation mode	SPPM	FDPIM	FDAPIM
Average transmit power	$\frac{3}{2^M+2} P_c$	$\frac{3}{2^M+4} P_c$	$\frac{1+\beta}{2^M+3} P_c$
Modulation mode	SDPPM	MPPM	
Average transmit power	$\frac{2}{ns} P_c$	$\frac{p}{n} P_c$	

Strong turbulence:

$$\langle C \rangle = \frac{B \cdot 2^a}{2\pi \ln 2 \Gamma(a)} G_{6,2}^{1,6} \left(\frac{8t^2 P_c^2 \eta_z^2 R}{a^2 \sigma_n^2} \cdot \left(\frac{\eta A}{\lambda L} \right)^4 \middle| \begin{matrix} 1, 1, \frac{1-a}{2}, \frac{2-a}{2}, 0, \frac{1}{2} \\ 1, 0, \frac{1-a}{4}, \frac{3-a}{4}, 0 \end{matrix} \right). \quad (3.31)$$

3.4.2 Synchronization Technology

3.4.2.1 Frame and Super Frame

As shown in Fig. 3.15, the beginning of a super frame is a synchronization header to ensure information synchronization. Each frame contains several information segments. Protection time slot can be included to prevent laser overload; in some lasers, protection time slot is not required.

3.4.2.2 Frame Synchronization

To correctly demodulate the information in the PPM frame, the information frame of the receiver must be strictly synchronized in time. Figure 3.16 is the schematic

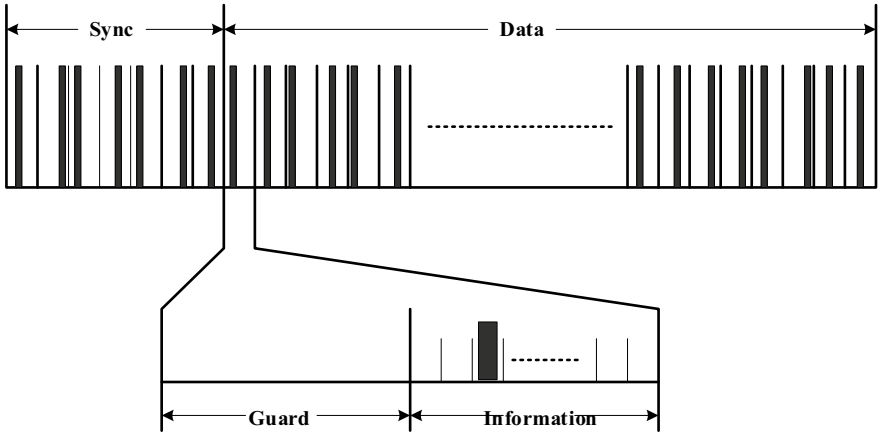


Fig. 3.15 Super frame structure of pulse position modulation

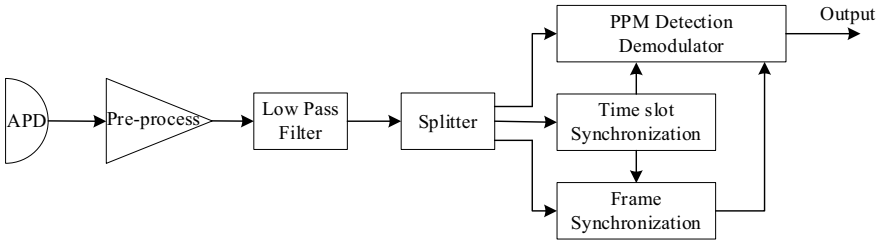


Fig. 3.16 Diagram of receiver composition

diagram of the receiver composition. After the optical pulse signal passes through the atmospheric channel, the noise is superimposed, and the signal amplitude is attenuated. Thus, the signal output by the avalanche photodiode is first processed by the preprocessor. The splitter sends the filtered signals to the demodulation detection unit, the time slot synchronization unit, and the frame synchronization unit, respectively.

3.4.2.3 Time Slot Synchronization

The protection time slot of PPM signal is half of the frame period. Regardless of the PPM pulse time slot, the position of the leading edge of the pulse in the time slot remains unchanged, which satisfies the condition for using digital phase-locked loop (DPLL) to extract the synchronization signal. The block diagram of the time slot synchronization subsystem is shown in Fig. 3.17. The signal output is first sent to the amplitude comparator by the splitter to remove the noise with small amplitude. The width comparator can remove the narrow pulse noise signal, and the pulse extension

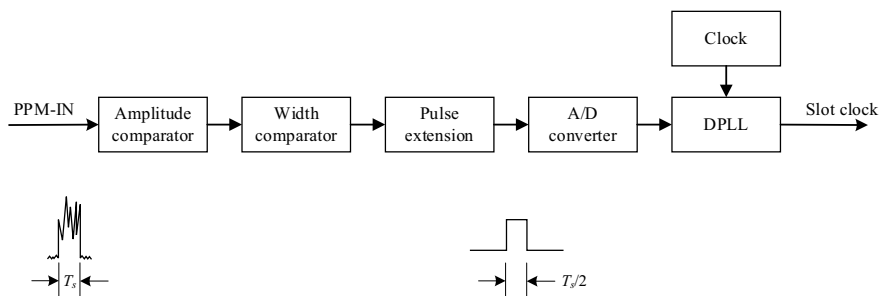


Fig. 3.17 Time slot synchronization

adjusts the duty cycle of the signal to 50%; then, the level signal is output. Then, after AD sampling, the analog signal is converted to digital signal, and then sent to the digital phase-locked loop to extract the synchronized time slot clock [33].

3.5 Direct Drive of Light Source

Direct modulation of the light source requires injecting the signal to be transmitted into the semiconductor laser. By changing some parameters of the laser, the output light wave changes with the modulation signal to achieve the modulation purpose of changing the laser output characteristics. A typical circuit is shown in Fig. 3.18, where the signal input terminal is a single-ended input.

The working process of the circuit is as follows. The AD8138 is used to convert the input single-ended electrical signal to a differential signal. The MAX9375 adjusts the differential signal to a proper range and outputs it to the MAX3738. The MAX3738 drives the laser with current under the action of the input differential signal and sends the light waves with signals.

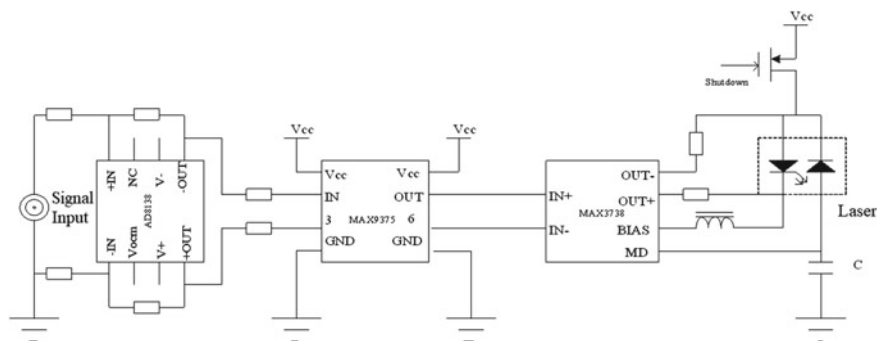


Fig. 3.18 Direct modulation drive circuit

3.5.1 Single-Ended to Differential Converter

When transmitted at a high rate, the signal in differential form is more stable than that in single-ended form, and most devices, such as MAX3738, use differential input and output when transmitting signals. The differential working condition is that two signals of equal amplitude, opposite phase, and centered on an appropriate common-mode voltage drive $IN+$ and $IN-$ at the same time. The ideal way to drive the MAX3738 differentially is to use a differential amplifier such as the AD8138, which can be used for single-ended to differential amplifiers or differential to differential amplifiers. Moreover, it provides common-mode level conversion. The AD8138 and its peripheral circuits [34] are shown in Fig. 3.19.

An AD8138 uses 5 V dual power supply with the main working process as follows. The chip has a dual-end input, $+IN$ is connected to input single-ended signal, $-IN$ is grounded, and two outputs are obtained through operational amplifier, where one is an in-phase output and the other is inverted output. Thus, a pair of signals with the same amplitude and opposite phase are obtained. This process converts the single-ended signal into a differential signal, which facilitates the laser driver unit.

The AD8138 has a unique feature of internal feedback that provides output gain and phase balance to suppress even-order harmonics. It uses two feedback loops to control the differential output voltage and common-mode output voltage, respectively. The differential feedback set by the external resistor controls the differential output voltage, and the common-mode feedback controls the common-mode output voltage, which can be adjusted by applying voltage on the V_{OCM} pin. A small resistor is connected in series to the output to prevent high-frequency ringing in the impulse response [34].

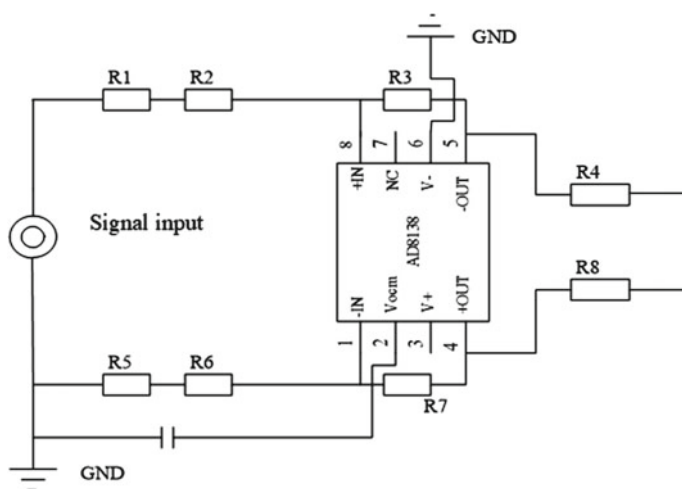
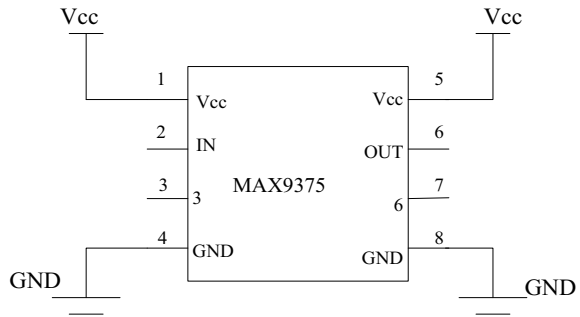


Fig. 3.19 AD8138 and its peripheral circuits

Fig. 3.20 MAX9375 and its peripheral circuits



3.5.2 Level Adjustment

The single-ended electrical signal is converted to a pair of differential signals through the AD8138, but the level of the converted signal does not meet the requirements of the laser driver, namely the MAX3738. Therefore, the level of the obtained differential signal must be adjusted to meet the input signal level required by the MAX3738. Thus, MAX9375 is used to complete such a process. The MAX9375 and its peripheral circuits [35] are shown in Fig. 3.20.

The chip accepts multiple types of input level and converts to low voltage positive emitter coupled logic (LVPECL) level signal output. With the operating frequency of up to 2 GHz, the clock jitter is very small, and it has a temperature compensation network. There are two levels of operational amplifiers with the same working principle inside the chip, both of which have one in-phase input and output; the other input and output are inverted. After the adjustment of the two operational amplifiers, the LVPECL level signal is output.

3.5.3 Laser Driver

The laser driver (MAX3738) is the core of the entire drive circuit. It converts the input voltage signal into a current signal and drives the laser to emit light, thereby transmitting the signal. A typical application circuit [36] with MAX3738 is shown in Fig. 3.21.

As shown in Fig. 3.22, the MAX3738 comprises three parts: high-speed modulation driver, bias current unit with extinction ratio control, and protection circuit. The MAX3738 adopts automatic power control working mode. Data is input from IN⁻ and IN⁺ terminals. After processing by input buffer circuit and data channel, the output of differential pair modulator is controlled to realize the modulation. For input from IN⁺, Q2 is enabled and opened so that current flows through the LD and light is emitted. The transistor is then driven by the input signal after data buffer and switch. The modulated signal is output from OUT⁻ and OUT⁺ to drive an external

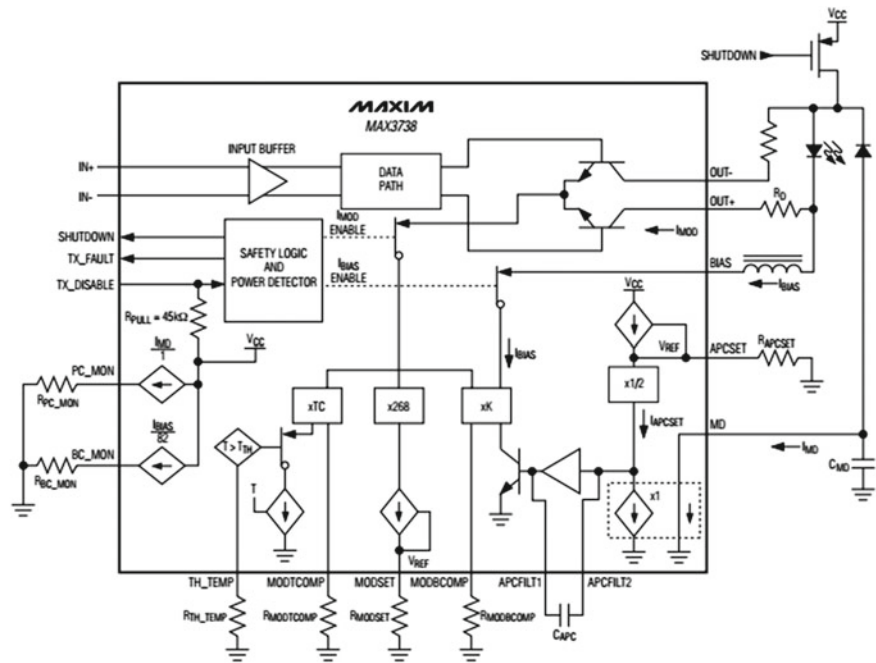


Fig. 3.21 MAX3738 and its peripheral circuits

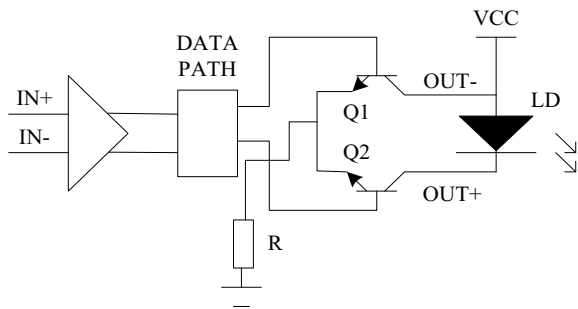


Fig. 3.22 High-speed modulation drive circuit inside a chip

laser tube. In this way, the required information can be modulated onto the laser beam for transmission. In contrast, for input from IN₋, Q1 is enabled and opened, and the LD does not emit light or emits weak light [29]. When the output power changes, the feedback signal is input from the MD terminal. The extinction ratio control circuit automatically controls the stability of the output optical power by adjusting the modulation current and the bias current change. When the temperature change exceeds the threshold, the temperature compensation circuit takes effect. The current is modulated to maintain power stability. When the circuit breaks down or

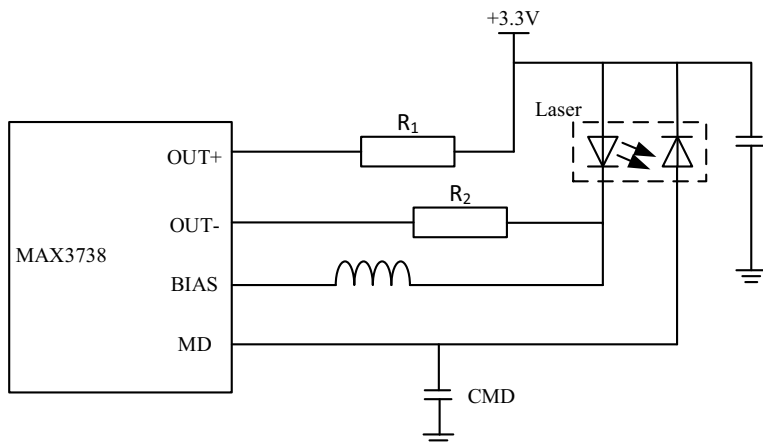


Fig. 3.23 DC coupling mode

other unexpected situations occur, the safety circuit is activated, the SHUTDOWN terminal outputs a control signal to turn off the laser tube output, and the TX_FAULT terminal outputs an alarm signal.

The MAX3738 receives differential input signals, provides a wide modulation current range (up to when AC coupling) and a high bias current range, making it suitable for driving Fabry–Perot/distributed feedback in optical module laser. According to the different connection methods of the laser and MAX3738, the laser drive interface can be divided into three modes: DC coupling, AC coupling, and differential drive.

(1) DC coupling

The basic connection of the DC coupling output network is shown in Fig. 3.23. It has fewer peripheral components and can provide a maximum modulation current of 60 mA.

(2) AC coupling

The basic connection of the AC coupling output network is shown in Fig. 3.24. The circuit will filter out the DC component and change the average value to zero [37]. The AC coupled output configuration can make the MAX3738 output a maximum modulation current of 85 mA.

For AC coupling, R_1 is used to divert the modulation current from the laser to reduce the total AC load impedance. The resistance value after connecting in series with the laser is connected in parallel. The total resistance should be approximately 15 ohms (MAX3738 is optimized for driving 15- Ω loads). AC coupling is suitable for driving vertical cavity surface emitting (VCSEL) lasers.

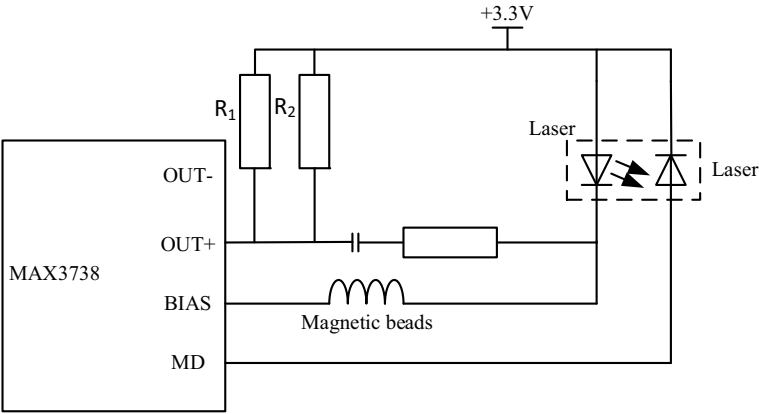


Fig. 3.24 AC coupling mode

(3) Differential drive mode

The differential drive mode must use AC coupled output. The configuration is divided into inductive pull-up and capacitive pull-up. The basic connections are shown in Figs. 3.25 and 3.26. The inductive pull-up type is very similar to the capacitive pull-up type. The main difference is whether the pull-up element on the OUT pin of the MAX3738 chip is resistive or capacitive. Inductive pull-up can provide greater modulation current than capacitive pull-up, whereas resistive pull-up requires fewer inductive components than capacitive pull-up and provides back matching [38].

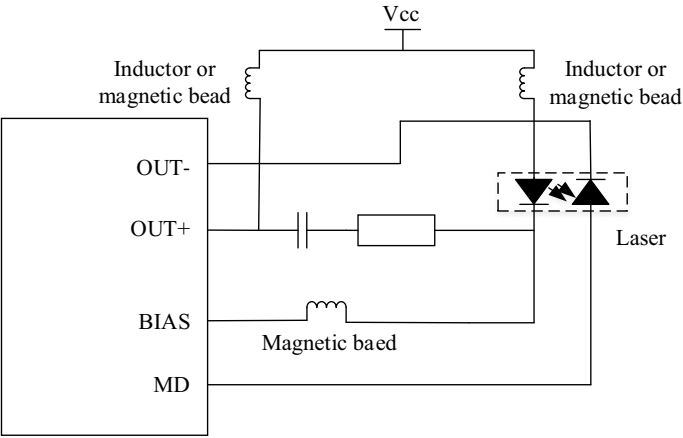


Fig. 3.25 Inductance pull-up

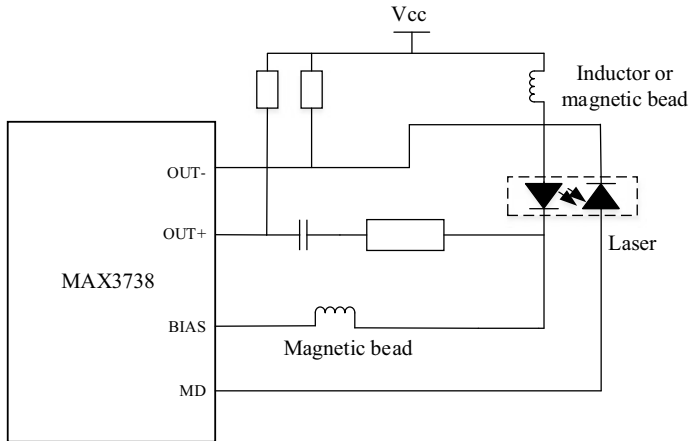


Fig. 3.26 Capacitance pull-up

3.5.4 Principle of Optical Feedback

Since the threshold current of the laser will change with the influence of temperature and components, the output power of the laser working under a given bias current will drop. To maintain constant output power of the optical transmitter, the automatic power control circuit of the laser can be used. The structure block diagram of the commonly used automatic power control principle [31] is shown in Fig. 3.27.

An important part of the structure is the optical feedback. The component for optical power measurement in Fig. 3.27 is a PIN photodetector encapsulated in a laser. It detects the change of the output optical power from the back light of the laser and turns it into an electrical signal through photoelectric conversion. The function of the power deviation detection circuit is to amplify the weak electrical signal output by the PIN, as the equivalent signal of the laser output optical power. This is sent to the input of the comparison integrating amplifier and compared with the reference level to adjust the DC bias of the laser current.

The principle of negative feedback control is as follows [39]. When the output optical power of the laser decreases, the output current of the PIN detector decreases,

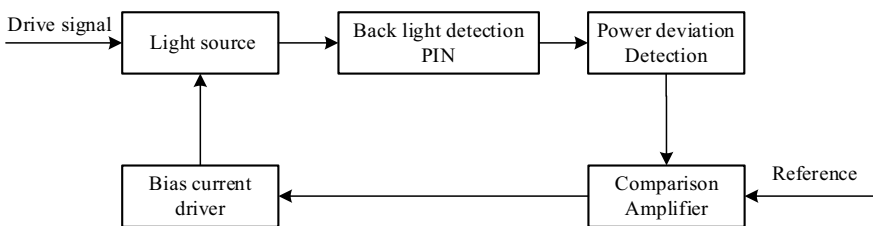


Fig. 3.27 Structure of an automatic power control

causing the level of the inverting input terminal of the comparison amplifier to drop. Because the reference level does not change, the output level of the comparison amplifier rises, and the base input current of the drive transistor increases, thereby increasing the DC bias current of the laser. Finally, the output optical power of the laser can be rebounded in time to achieve the purpose of stabilizing the output optical power of the LD. Conversely, when the output power of the laser increases, the PIN detection output current increases, and the level of the inverting input terminal of the comparison amplifier increases, resulting in a drop of the output level of the comparison amplifier. The output current of the base of the driving transistor decreases so that the DC bias current of the laser is reduced, and finally the output optical power of the laser is reduced.

The MAX3738 laser driver integrates an automatic power control module, which adjusts the average power to keep the laser coupled to the photodiode current constant and compensates for the modulation current to keep the peak power constant throughout the effective life and temperature range.

3.6 Subcarrier Intensity Modulation

The baseband signal is modulated onto an electric carrier, and the electric signal is used to modulate the intensity of the light source. The electric carrier is called a subcarrier. At the receiving end, the electric carrier containing the baseband signal is recovered through optical detection and then restored to the baseband signal. Since this modulation modulates the light intensity, it is called subcarrier intensity modulation. There are two types of carriers: an optical carrier and electric signal carrier (called subcarrier). A block diagram of the subcarrier intensity modulation optical wireless communication system is shown in Fig. 3.28.

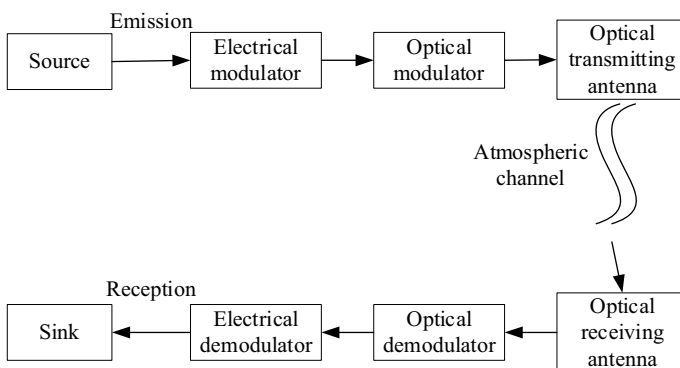


Fig. 3.28 Subcarrier modulation optical wireless communication system

3.6.1 Subcarrier Intensity Modulation

PSK modulation can be divided into binary PSK (BPSK) and multiple PSK (MPSK). In the binary modulation technology, the carrier phase has only two values of “0” and “ π ”, which correspond to the “0” and “1” of the modulation signal, respectively. When sending a “1” symbol, a carrier with a starting phase of π is transmitted. When sending a “0” symbol, a carrier with a starting phase of 0 is transmitted. After level conversion, the binary modulation signal represented by “0” and “1” becomes a bipolar non-return-to-zero signal represented by “-1” and “1”, which is then multiplied by the carrier to form a 2PSK signal.

In MPSK, the most commonly used modulation is quadrature phase-shift keying (QPSK), which can be regarded as a composition of two 2PSK modulators. The input serial binary information sequence is converted into two-way rate halved sequence after serial-parallel conversion. The level converter generates bipolar two-level signals $I(t)$ and $Q(t)$, respectively, and then the carriers $A \cos 2\pi f_c t$ and $A \sin 2\pi f_c t$ are modulated. After their addition, the QPSK signal is obtained.

In the subcarrier modulation system, it is assumed that the radio frequency subcarrier signal after pre-modulating the source is used to modulate the intensity of the laser light emitted by the laser. For MPSK subcarrier modulation, after serial-parallel conversion, one symbol is converted into in-phase branch data I , and quadrature branch data Q at a time and the amplitude is. According to the data of channels I and Q , it can be mapped to the corresponding phase; because the subcarrier signal is a sinusoidal signal with positive and negative signals, it must be injected into the laser as a drive current after adding a DC bias. In the N-channel subcarrier intensity modulated FSO system:

$$m(t) = \sum_{i=1}^N m_i(t). \quad (3.32)$$

The radio frequency subcarrier modulation signal can be expressed as

$$m_i(t) = g(t)a_{ic} \cos(\omega_{ci}t + \varphi_i) + g(t)a_{is} \sin(\omega_{ci}t + \varphi_i), \quad (3.33)$$

where $g(t)$ is the pulse shaping function, and carrier frequency and phase are $\{\omega_{ci}, \varphi_i\}_{i=1}^N$. When the receiver adopts direct detection, the light intensity signal is photoelectrically converted into a current signal $I(t)$:

$$I(t) = RA(t)[1 + \xi m(t)] + n(t), \quad (3.34)$$

where R is the photoelectric conversion constant, and $\xi = |m(t)/(i_B - i_{Th})|$ is the optical modulation index.

3.6.2 BPSK Subcarrier Modulation

Binary phase-shift keying (BPSK) is a phase modulation method that switches the carrier phase between two different values based on the two levels of the digital baseband signal. Usually, the phase difference between the two carriers is π radians, so it is sometimes called inverted keying PSK. Take binary phase modulation as an example. When the symbol is “0”, the modulated carrier is in phase with the unmodulated carrier; when the symbol is “1”, the modulated carrier is in phase with the unmodulated carrier. After modulation, the carrier phase difference for symbols “1” and “0” is π . For the light intensity modulation/direct detection (IM/DD) communication system, the light intensity $P(t)$ received by the receiver can be expressed as

$$P(t) = A(t)P_s(t) + n(t), \quad (3.35)$$

where $P_s(t)$ represents the light intensity emitted by the transmitter, and $n(t)$ represents the noise of the receiver. For the subcarrier BPSK modulation system, the light intensity emitted by the optical transmitter is

$$s(t) = 1 + \xi [s_i(t) \cos \omega_c t - s_q(t) \sin \omega_c t], \quad (3.36)$$

where $s_i(t) = \sum_j g(t - jT_s) \cos \Phi_j$ is the in-phase signal, and $s_q(t) = \sum_j g(t - jT_s) \sin \Phi_j$ is the orthogonal signal; ξ is the modulation index, $0 < \xi \leq 1$, Φ_j is the j -th phase, $g(t)$ is the gate pulse, and T_s is the symbol time. The light intensity received by the receiver is

$$P(t) = \frac{P_{\max}}{2} A(t) \{1 + \xi [s_i(t) \cos \omega_c t - s_q(t) \sin \omega_c t]\}. \quad (3.37)$$

After passing through the photodetector, the output electrical signal is

$$\begin{aligned} I(t) = & \frac{P_{\max} R}{2} A(t) \{1 + \xi [s_i(t) \cos \omega_c t - s_q(t) \sin \omega_c t]\} \\ & + n_i(t) \cos \omega_c t - n_q(t) \sin \omega_c t, \end{aligned} \quad (3.38)$$

where R is the photoelectric conversion constant, and $n_i(t)$ and $n_q(t)$ are Gaussian white noise with variance σ_g^2 .

In the subcarrier modulation system, the same optical power is required for transmission as in the OOK system discussed in the previous section, and the power spectral density of the signal received by the receiver is [40]:

$$I(f) = A(f) + \frac{B(f - f_c) + B(f + f_c)}{2} + \frac{N(f - f_c) + N(f + f_c)}{2}, \quad (3.39)$$

where $B(f) = A(f) * Z(f)$. The slow fading channel depends on the DC component $A(f)$. If the carrier frequency f_c is high enough, assuming that $f_c > B_A + B_B$ and $o f_c$ is the intermediate frequency, B_A is the single-sideband bandwidth of $A(f)$, and B_B is the single-sideband bandwidth of $B(f)$. The first term in Eq. (3.39) can be filtered out by the band-pass filter in the receiver. The filtered signal passes through coherent demodulation for carrier phase recovery, and then through a low-pass filter to filter out high-frequency components to obtain the in-phase signal of the output signal:

$$r_i(t) = \frac{P_{\max} R}{2} \xi A(t) s_i(t) + n_i(t), \quad (3.40)$$

and quadrature signal:

$$r_i(t) = \frac{P_{\max} R}{2} \xi A(t) s_q(t) + n_q(t). \quad (3.41)$$

If the subcarrier modulation scheme is BPSK, the atmospheric fading effect is not considered, and the channel is Gaussian distribution, then the bit error rate of the system can be expressed as

$$P_e = Q(\sqrt{2\text{SNR}}), \quad (3.42)$$

where $\text{SNR} = \frac{(P_{\max}/2)^2 (R)^2 \xi^2}{2\sigma_g^2}$. If the atmospheric fading effect is considered, the demodulated signal is

$$r(t) = \frac{P_{\max} R}{2} [\xi A(t) s(t) + n(t)]/2. \quad (3.43)$$

Assuming equal probability of transmitting “0” code and “1” code, namely, $p(1) = p(0) = 0.5$, the bit error rate of BPSK optical wireless communication system is

$$P_e = p(1)p(r|1) + p(0)p(r|0). \quad (3.44)$$

- (1) In the case of weak turbulence, the light intensity fluctuation $A(t)$ obeys the log-normal distribution. For BPSK subcarrier modulation, the conditional probability density function $p(r|x)$ of the received signal is:

$$p(r|x) = \begin{cases} \frac{\exp(-\sigma_l^2/2)}{2\pi\sigma_l\sigma_g} \int_0^\infty \frac{1}{t^2} \exp\left\{-\left[\frac{\ln^2 x}{2\sigma_l^2} + \frac{(\xi r - t)^2}{2\sigma_g^2}\right]\right\} dt & x = +1 \\ \frac{\exp(-\sigma_l^2/2)}{2\pi\sigma_l\sigma_g} \int_{-\infty}^0 \frac{1}{t^2} \exp\left\{-\left[\frac{\ln^2 x}{2\sigma_l^2} + \frac{(\xi r + t)^2}{2\sigma_g^2}\right]\right\} dt & x = 0, \end{cases} \quad (3.45)$$

For BPSK modulation, the decision threshold value is 0. Bringing Eq. (3.44) into Eq. (3.45), the following results can be obtained:

$$P_e = \frac{\exp(-\sigma_l^2/2)}{\sqrt{2\pi}\sigma_l} \int_0^\infty \frac{1}{x^2} \exp\left(-\frac{\ln^2 x}{2\sigma_g^2}\right) Q\left(\frac{x}{\sigma_g}\right) dx. \quad (3.46)$$

- (2) When the light intensity fluctuation $A(t)$ obeys the gamma-gamma distribution, the conditional probability density function $p(r|x)$ of the received signal for BPSK subcarrier modulation is [41]:

$$p(r|x) = \begin{cases} \frac{2}{\sqrt{2\pi}\sigma_g\Gamma(\alpha)\Gamma(\beta)} \left(\frac{\alpha\beta}{\xi}\right) \int_0^\infty t^{\frac{\alpha+\beta}{2}} K_{\alpha-\beta}\left(2\sqrt{\frac{\alpha\beta t}{\xi}}\right) \exp\left\{-\left[\frac{(r-t)^2}{2\sigma_g^2}\right]\right\} dt & x = +1 \\ \frac{2}{\sqrt{2\pi}\sigma_g\Gamma(\alpha)\Gamma(\beta)} \left(\frac{\alpha\beta}{\xi}\right) \int_0^\infty t^{\frac{\alpha+\beta}{2}} K_{\alpha-\beta}\left(2\sqrt{\frac{\alpha\beta t}{\xi}}\right) \exp\left\{-\left[\frac{(r+t)^2}{2\sigma_g^2}\right]\right\} dt & x = 0, \end{cases} \quad (3.47)$$

The bit error rate can be obtained by introducing Eq. (3.47) into Eq. (3.44):

$$P_e = \frac{(\alpha\beta)^{\frac{\alpha+\beta}{2}}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty x^{\frac{\alpha+\beta}{2}-1} K_{\alpha-\beta}\left(2\sqrt{\alpha\beta x}\right) \operatorname{erfc}\left(\frac{\xi x}{\sqrt{2}\sigma_g}\right) dx. \quad (3.48)$$

3.6.3 FSK Subcarrier Modulation

FSK modulation can be divided into two types according to whether the phase is continuous: the first is the discontinuous phase FSK modulation according to the input data bits (0 and 1) and switch between two independent oscillators. The phase of the waveform generated by this method is discontinuous at the moment of switching. The second is the continuous phase FSK signal, whose power spectral density function fades according to the negative fourth power of the frequency offset.

If the phase is discontinuous, the power spectral density function decays according to the negative quadratic power of frequency offset. The time domain expression of binary FSK signal is [42]

$$e_{2FSK}(t) = b(t) \cos(\omega_1 t + \varphi_1) + \overline{b(t)} \cos(\omega_2 t + \varphi_2), \quad (3.49)$$

where $b(t)$ is the baseband signal with the expression:

$$b(t) = \sum_{n=-\infty}^{\infty} a_n g(t - nT_s) \quad a_n = \begin{cases} 0, & \text{probability } P \\ 1, & \text{probability } 1 - P. \end{cases} \quad (3.50)$$

The intensity of the laser emission is

$$s(t) = 1 + \sum_{n=-\infty}^{\infty} a_n g(t - nT_s) \cos(\omega_1 t + \varphi_1) + \sum_{n=-\infty}^{\infty} \bar{a}_n g(t - nT_s) \cos(\omega_2 t + \varphi_2). \quad (3.51)$$

Without loss of generality, let

$$s(t) = 1 + \sum_{n=-\infty}^{\infty} a_n g(t - nT_s) \cos(\omega_1 t) + \sum_{n=-\infty}^{\infty} \bar{a}_n g(t - nT_s) \cos(\omega_2 t); \quad (3.52)$$

then, the received signal is

$$\begin{aligned} r(t) = & A(t) + \sum_{n=-\infty}^{\infty} a_n g(t - nT_s) A(u, t) \cos(\omega_1 t) \\ & + \sum_{n=-\infty}^{\infty} \bar{a}_n g(t - nT_s) A(u, t) \cos(\omega_2 t) + n(t). \end{aligned} \quad (3.53)$$

The first term of Eq. (3.53) can be filtered by a band-pass filter, and the received signal is

$$\begin{aligned} r(t) = & \sum_{n=-\infty}^{\infty} a_n g(t - nT_s) A(u, t) \cos(\omega_1 t) \\ & + \sum_{n=-\infty}^{\infty} \bar{a}_n g(t - nT_s) A(u, t) \cos(\omega_2 t) + n(t) \end{aligned} \quad (3.54)$$

If the synchronous detection method is used and the symbol sent at time $(0, T_s)$ is “1”, the waveform of the two signals sent to the sampling arbiter for comparison at this time is

$$\begin{cases} x_1(t) = A(t) + n_1(t) \\ x_2(t) = n_2(t), \end{cases} \quad (3.55)$$

where $n_1(t)$ and $n_2(t)$ are normal random variables with variance σ_g^2 , sampling value $x_1(t) = A(t) + n_1(t)$ is a normal random variable with mean value $A(t)$ and variance σ_g^2 , and sampling value $x_2(t) = n_2(t)$ is a normal random variable with mean value of 0 and variance of σ_g^2 . Since $x_1 < x_2$, the “1” code will be wrongly determined as the “0” code at this time, and the error probability P_{e1} at this time is (here $A(t)$ is replaced by a)

$$P_{e1} = p(x_1 < x_2) = p[(a + n_1) < n_2] = p(a + n_1 - n_2 < 0). \quad (3.56)$$

Let $z = a + n_1 + n_2$, where z is a normal random variable with a mean value of a and variance of σ_z^2 ; $\sigma_z^2 = 2\sigma_g^2$, so the probability density function $p(z)$ of z is

$$p(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(z-a)^2}{2\sigma_z^2}\right] = \frac{1}{2\sqrt{\pi}\sigma_g} \exp\left[-\frac{(z-a)^2}{4\sigma_g^2}\right]. \quad (3.57)$$

The probability density function of $A(t)$ is

$$p(A) = \frac{1}{\sqrt{2\pi}\sigma_l A} e^{-\frac{(\ln A + \sigma_l^2/2)^2}{2\sigma_l^2}}. \quad (3.58)$$

According to Eqs. (3.57) and (3.58), the joint probability density function is

$$p(r|s_1) = \frac{\exp(-\sigma_l^2/2)}{4\pi\sigma_l\sigma_g} \int_0^\infty \frac{1}{x^2} \exp\left\{-\left[\frac{\ln^2 x}{2\sigma_l^2} + \frac{(r-x)^2}{4\sigma_g^2}\right]\right\} dx. \quad (3.59)$$

Since the probabilities of transmitted “0” being judged as “1” and transmitted “1” being judged as “0” are equal, the bit error rates in these two cases are the same. Let $P(0) = 0.5$ to obtain the total bit error rate:

$$p_e = \frac{1}{\sqrt{2\pi}\sigma_l} \exp\left(-\frac{\sigma_l^2}{2}\right) \int_0^\infty \frac{1}{x^2} \exp\left(-\frac{\ln^2 x}{2\sigma_l^2}\right) Q\left(\frac{x}{\sqrt{2}\sigma_g}\right) dx. \quad (3.60)$$

3.6.4 Intermodulation Distortion and Carrier-to-Noise Ratio

The nonlinearity of laser is the nonlinearity of the modulation response. In LD modulation system, we analyze the rate equation using the Bessel function method and solve for the intermodulation distortion. Since the transmission bandwidth is one octave, the second-order intermodulation distortion can be ignored. Here, we only consider the third-order intermodulation distortion.

The variance of the third-order intermodulation distortion [43] is

$$\sigma_{IMD3}^2 = \frac{1}{32} (\eta P_x)^2 m^6 (N_{21} + N_{111}), \quad (3.61)$$

where η is the responsivity, P_x is the average received optical power, N_{21} and N_{111} is the number of 3rd-order intermodulation products, and $\omega_x + \omega_y - \omega_z$ and $2\omega_x - \omega_y$ meet the bandwidth requirements in one octave.

The injection current of the rate equation is taken as the sum of subcarriers with equal amplitude and bandwidth [44]

$$P(t) = P_T \exp\left(m \sum_{n=1}^N \cos(\omega_n t + \phi_n(t))\right), \quad (3.62)$$

where P_T is the average transmitted optical power, and m is the optical modulation index. The output light intensity can be expressed as:

$$I(t) = I[1 + mx(t) + a_2 m^2 x^2(t) + \cdots a_i m^i x^i(t) + \cdots], \quad (3.63)$$

where $\{\alpha_n\}_{n=2}^{\infty}$ is the nonlinear coefficient of the laser, which is a constant. As the multichannel subcarrier PSK modulation is considered, we set $x(t) = \sum_{i=1}^N \cos(\omega_i t + \phi_i)$.

Since the light intensity fluctuates very slowly and only AC signals are considered, the received current signal is [4]:

$$i(t) = \eta I m \left[\sum_{n=1}^N \cos(w_n t + \phi_n) \right] + n(t). \quad (3.64)$$

Therefore, the carrier power and noise power are obtained as follows:

$$S_P = \frac{\eta^2 I^2}{2} \left[m + \frac{3}{4} a_3 m^3 (2N - 1) \right]^2, \quad (3.65)$$

$$\sigma^2 = \sigma_{Sh}^2 + \sigma_{Th}^2. \quad (3.66)$$

The carrier-to-noise ratio is thus obtained as

$$\text{CINR} = \frac{S_P}{\sigma^2 + \sigma_{IMD}^2} = \frac{16R^2 I^2 \left[m + \frac{3}{4} a_3 m^3 (2N - 1) \right]^2}{32\sigma^2 + (\eta P_r)^2 a_3^2 I^2 m^6 (N_{21} + N_{111})}. \quad (3.67)$$

Let $\frac{d(\text{CINR})}{dm} = 0$; then, we have $m = \left[\frac{16\sigma^2}{(\eta P_r)^2 a_3^2 I^2 (N_{21} + N_{111})} \right]^{1/6}$.

Because the fluctuation of light intensity caused by atmospheric turbulence is not considered in Eq. (3.67), the modulation index obtained is not the optimal modulation index. Under general atmospheric turbulence conditions, the carrier-to-noise ratio is obtained as [45]

$$\text{CINR}_{\text{ave}} = \int_0^{\infty} \text{CINR} \cdot p(I) dI. \quad (3.68)$$

Under the conditions of weak turbulence, if we use the log-normal distribution model, Eq. (3.68) becomes

$$\text{CINR}_{\text{ave}} = \frac{16R^2 \left[m + \frac{3}{4}a_3m^3(2N-1) \right]^2}{\sqrt{2\pi}\sigma_1} \int_0^\infty \frac{I}{32\sigma^2 + Ka_3^2I^2m^6} \exp\left(-\frac{[\ln(I/I_0) + \sigma_1^2/2]^2}{2\sigma_1^2}\right) dI \quad (3.69)$$

where $K = (\eta P_r)^2(N_{111} + N_{21})$.

For intermodulation distortion produced by BPSK modulation, the system unconditioned bit error rate can be obtained by the following formula [46]:

$$\begin{aligned} P_e(m) &= \int_0^\infty Q(\sqrt{\text{CINR}}) p(I) dI \\ &= \int_0^\infty Q(\sqrt{\text{CINR}}) \frac{1}{\sqrt{2\pi}\sigma_1 I} e^{-\frac{[\ln(I/I_0) + \sigma_1^2/2]^2}{2\sigma_1^2}} dI \end{aligned} \quad (3.70)$$

Under the conditions of moderately strong turbulence, the log-normal distribution model is not applicable, and we use the gamma-gamma model. Then, Eq. (3.70) becomes:

$$\begin{aligned} \text{CINR}_{\text{ave}} &= \frac{32R^2 \left[m + \frac{3}{4}a_3m^3(2N-1) \right]^2 (\alpha\beta)^{\frac{\alpha+\beta}{2}}}{\Gamma(\alpha)\Gamma(\beta)} \\ &\int_0^\infty \frac{I^{\frac{\alpha+\beta}{2}+1}}{32\sigma^2 + Km^6I^2a_3^2} K_{\alpha-\beta}(2\sqrt{\alpha\beta}I) dI \end{aligned} \quad (3.71)$$

where $\Gamma(\cdot)$ is the gamma function, $K_n(\cdot)$ is a Bessel function of the second kind of order n , α and β are the outer and inner scale parameters, respectively. SI is the scintillation index, defined as

$$\text{SI} = \frac{1}{\alpha} + \frac{1}{\beta} + \frac{1}{\alpha\beta}. \quad (3.72)$$

When gamma-gamma model is used for the variance of light intensity fluctuation under moderate and strong turbulence conditions, the bit error rate of the FSO system changes to

$$\begin{aligned}
P_e(m) &= \int_0^{\infty} Q(\sqrt{\text{CINR}}) p(I) dI \\
&= \int_0^{\infty} Q(\sqrt{\text{CINR}}) \frac{2(\alpha\beta)^{\frac{\alpha+\beta}{2}}}{\Gamma(\alpha)\Gamma(\beta)} I^{\frac{\alpha+\beta}{2}-1} K_{\alpha-\beta}(2\sqrt{\alpha\beta}I) dI \quad (3.73)
\end{aligned}$$

3.7 Orthogonal Frequency-Division Multiplexing

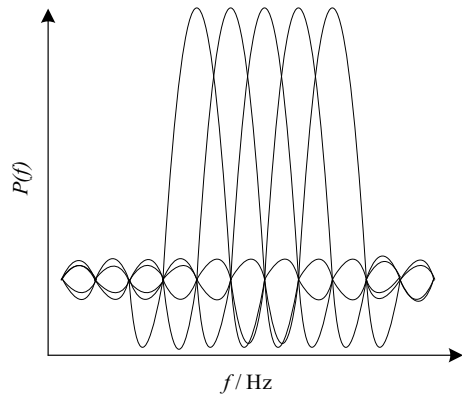
The concept of orthogonal frequency-division multiplexing (OFDM) originated from frequency-division multiplexing (FDM) and multicarrier modulation (MCM) technology [47]. The OFDM system converts broadband channels into many parallel orthogonal sub-channels, thereby converting frequency-selective channels into a series of frequency-flat fading channels. Figure 3.29 shows the OFDM subcarrier spectrum distribution diagram. OFDM selects subcarriers that are orthogonal to each other in the time domain. Although they are aliased in the frequency domain, they can still be separated at the receiving end.

3.7.1 Basic Principles

OFDM modulates the transmitting end data signal into multiple parallel subcarriers for transmission, and these subcarriers share the system bandwidth. The basic block diagram of the OFDM system is shown in Fig. 3.30.

Figure 3.31 shows the schematic diagram of the OFDM transmitter. At the transmitter of the OFDM system, the serial data stream is first converted into M parallel

Fig. 3.29 Spectrum distributions of orthogonal frequency-division multiplexing subcarriers



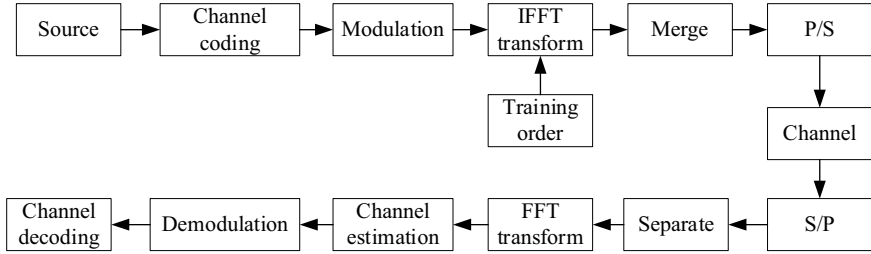


Fig. 3.30 Basic principle of an orthogonal frequency-division multiplexing system

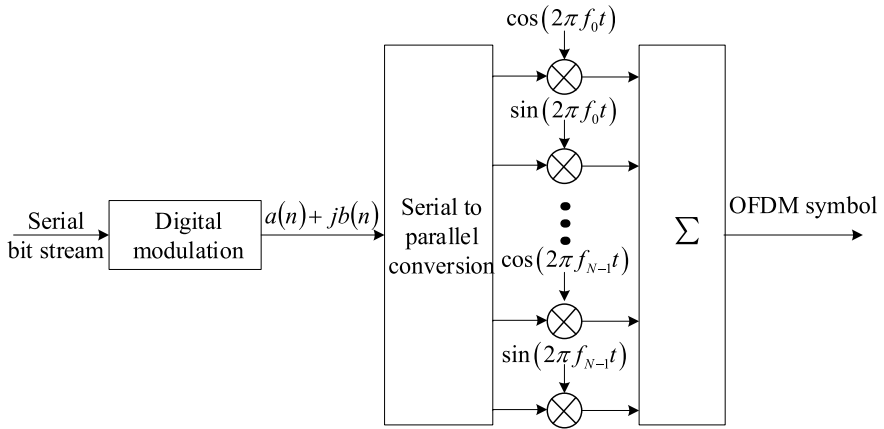


Fig. 3.31 Orthogonal frequency-division multiplexing (OFDM) modulation

sub-data streams through serial-parallel conversion (a_m and b_m represent the m -th in-phase and quadrature components of the data symbol, respectively), each sub-data stream is modulated to the corresponding subcarrier, and these modulated signals are synthesized into the transmitter signal $s(t)$ [47].

$$s(t) = \sum_{m=0}^{M-1} \text{Re}\{(a_m + jb_m) \exp(j2\pi f_m t)\} = \sum_{m=0}^{M-1} [a_m \cos(2\pi f_m t) - b_m \sin(2\pi f_m t)]. \quad (3.74)$$

An OFDM signal is composed of M subcarrier signals of different frequencies, which are orthogonal to each other; that is, the subcarrier signals satisfy the following relationship:

$$\int_0^{T_s} (a_m + jb_m) \exp(j2\pi f_m t) \times (a_n + jb_n)^* \exp(-j2\pi f_n t) dt = 0, \quad (3.75)$$

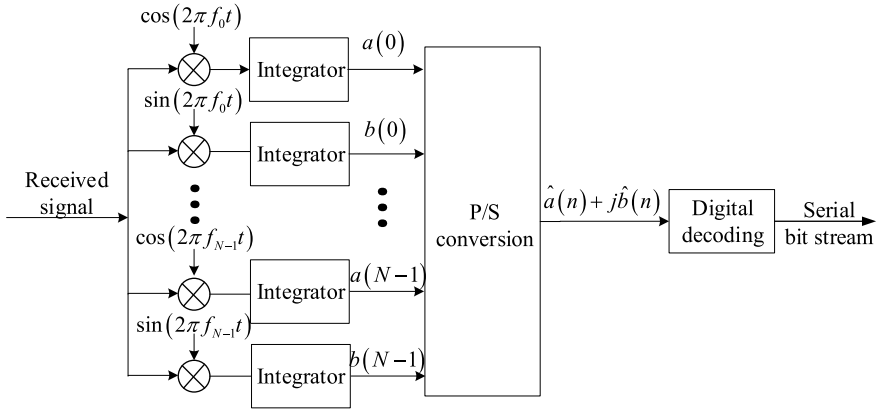


Fig. 3.32 Orthogonal frequency-division multiplexing (OFDM) demodulation

where $m \neq n$ and $*$ represents conjugation. Therefore, m frequencies satisfy the following relationship

$$f_m - f_n = \frac{i}{T_s}, \quad (3.76)$$

where i is an integer satisfying $i \geq 1$.

As shown in Fig. 3.32, the demodulation of OFDM can be expressed as

$$\begin{aligned} \frac{1}{T_s} \int_0^{T_s} s(t) 2 \cos(2\pi f_0 t) dt &= \frac{1}{T_s} \int_0^{T_s} \sum_{m=0}^{M-1} 2 \cos(2\pi f_0 t) dt \\ &= \sum_{m=0}^{M-1} a_m \frac{1}{T_s} \int_0^{T_s} (\cos[2\pi(f_m + f_0)t] + \cos[2\pi(f_m - f_0)t]) dt \\ &\quad - \sum_{m=0}^{M-1} b_m \frac{1}{T_s} \int_0^{T_s} (\sin[2\pi(f_m + f_0)t] + \sin[2\pi(f_m - f_0)t]) dt, \end{aligned} \quad (3.77)$$

Suppose that the frequencies of subcarriers are i/T_s in an OFDM system; then, from Eq. (3.77), we obtain

$$\frac{1}{T_s} \int_0^{T_s} \cos[2\pi(f_m + f_0)t] dt = 0$$

$$\begin{aligned}\frac{1}{T_s} \int_0^{T_s} \cos[2\pi(f_m - f_0)t] dt &= \begin{cases} 1 & f_m = f_0 \\ 0 & f_m \neq f_0 \end{cases} \\ \frac{1}{T_s} \int_0^{T_s} \sin[2\pi(f_m + f_0)t] dt &= 0 \\ \frac{1}{T_s} \int_0^{T_s} \sin[2\pi(f_m - f_0)t] dt &= 0.\end{aligned}$$

Substituting the above formula into Eq. (3.77), we obtain

$$\frac{1}{T_s} \int_0^{T_s} s(t) 2 \cos(2\pi f_0 t) dt = a_0. \quad (3.78)$$

The orthogonal component b_0 transmitted on the subcarrier f_0 can be obtained according to the same process:

$$\frac{1}{T_s} \int_0^{T_s} s(t) [-2 \sin(2\pi f_0 t)] dt = b_0. \quad (3.79)$$

The data symbols transmitted on other subcarriers can be demodulated similarly. After all subcarrier signals are demodulated, they are restored to serial data by parallel–serial conversion.

3.7.2 Implementation of Discrete Fourier Transform in OFDM

In an OFDM system, the signal of Eq. (3.79) can be rewritten as

$$\begin{aligned}s(t) &= \sum_{m=0}^{M-1} \operatorname{Re} \left\{ (a_m + jb_m) \exp \left(j \frac{2\pi mt}{T_s} \right) \right\} \\ &= \sum_{m=0}^{M-1} \left[a_m \cos \left(\frac{2\pi mt}{T_s} \right) - b_m \sin \left(\frac{2\pi mt}{T_s} \right) \right].\end{aligned} \quad (3.80)$$

At time $0, \Delta t, 2\Delta t, \dots, (M-1)\Delta t$ for $\Delta t = T_s/M$, the OFDM signal $s(t)$ in Eq. (3.73) is sampled to obtain a vector $\mathbf{s} = [s_0, s_1, \dots, s_{M-1}]^T$ (superscript T stands for vector transposition). Then, the n th element of vector \mathbf{s} can be expressed as

$$s_n = s(n\Delta t) = \sum_{m=0}^{M-1} \operatorname{Re} \left\{ (a_m + jb_m) \exp \left(j \frac{2\pi mn}{M} \right) \right\} \quad (3.81)$$

$$= \sum_{m=0}^{M-1} \left[a_m \cos \left(\frac{2\pi mn}{M} \right) - b_m \sin \left(\frac{2\pi mn}{M} \right) \right], \quad (3.82)$$

where $n = 0, 1, \dots, M-1$. It can be seen from Eqs. (3.81) and (3.82) that, except for a constant difference $1/\sqrt{M}$, the vector that refers to $\{S_k = a_k + jb_k\}_{k=0}^{M-1}$ is used as the real part of inverse discrete Fourier transform (IDFT) for M transmission symbols. Therefore, this constant coefficient (related to transmission power) is ignored. The elements in \mathbf{s} can be expressed as

$$\begin{aligned} s_n &= \operatorname{Re}\{\operatorname{IDFT}\{a_0 + jb_0, a_1 + jb_1, \dots, a_{M-1} + jb_{M-1}\}\} \\ &= \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} \left[a_m \cos \left(\frac{2\pi mn}{M} \right) - b_m \sin \left(\frac{2\pi mn}{M} \right) \right], \quad n = 0, 1, \dots, M-1. \end{aligned} \quad (3.83)$$

Therefore, the OFDM transmission signal $s(t)$ can be obtained by passing the elements in \mathbf{s} through a low-pass filter at intervals of Δt [47]. Therefore, the multi-carrier modulation in the OFDM system can be realized by IDFT. The transmitted signal corresponding to Eq. (3.83) can be expressed as

$$s(t) = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} \left[a_m \cos \left(\frac{2\pi mt}{T_s} \right) - b_m \sin \left(\frac{2\pi mt}{T_s} \right) \right]. \quad (3.84)$$

Considering a distortion-free channel, the input signal of the OFDM demodulator is the same as Eq. (3.84). Since the transmitter only transmits the real part of the IDFT, the receiver must sample the signal at a time interval of $\Delta t/2 = T_s/2M$ and sampling rate of twice $1/\Delta t$. The sampled signal is [47]

$$s_n = s\left(\frac{n\Delta t}{2}\right) = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} \left[a_m \cos \left(\frac{2\pi mn}{2M} \right) - b_m \sin \left(\frac{2\pi mn}{2M} \right) \right] \quad (3.85)$$

for $n = 0, 1, \dots, 2M-1$. Since $\cos \alpha = (e^{j\alpha} + e^{-j\alpha})/2$ and $\sin \alpha = (e^{j\alpha} - e^{-j\alpha})/2j$, Eq. (3.85) can be expressed as

$$s_n = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} \left[\left(\frac{a_m}{2} - \frac{b_m}{2j} \right) \exp\left(j \frac{2\pi mn}{2M}\right) + \left(\frac{a_m}{2} + \frac{b_m}{2j} \right) \exp\left(-j \frac{2\pi mn}{2M}\right) \right] \quad (3.86)$$

for $n = 0, 1, \dots, 2M - 1$. Performing discrete Fourier transform (DFT) on the above sequence $\{S_n\}_{n=0}^{2M-1}$, we obtain

$$\begin{aligned} S_k &= \frac{1}{\sqrt{2M}} \sum_{n=0}^{2M-1} s_n \exp\left(-j \frac{2\pi nk}{2M}\right) \\ &= \frac{1}{\sqrt{2}} \frac{1}{M} \sum_{n=0}^{2M-1} \sum_{m=0}^{M-1} \left[\left(\frac{a_m}{2} - \frac{b_m}{2j} \right) \exp\left(j \frac{2\pi mn}{2M}\right) + \left(\frac{a_m}{2} + \frac{b_m}{2j} \right) \exp\left(-j \frac{2\pi mn}{2M}\right) \right] \exp\left(-j \frac{2\pi nk}{2M}\right) \\ &= \frac{1}{\sqrt{2}} \frac{1}{M} \sum_{n=0}^{2M-1} \sum_{m=0}^{M-1} \left[\left(\frac{a_m}{2} - \frac{b_m}{2j} \right) \exp\left(j \frac{2\pi n(m-k)}{2M}\right) + \left(\frac{a_m}{2} + \frac{b_m}{2j} \right) \exp\left(-j \frac{2\pi n(m+k)}{2M}\right) \right] \end{aligned} \quad (3.87)$$

for $k = 0, 1, 2, \dots, M - 1$. Using the identity

$$\frac{1}{\sqrt{2M}} \sum_{m=0}^{2M-1} \exp\left(j \frac{2\pi mn}{2M}\right) = \begin{cases} 1, & m = 0, \pm 2M, \pm 4M, \dots \\ 0, & \text{other} \end{cases}, \quad (3.88)$$

we obtain

$$\begin{aligned} S_k &= \sum_{m=0}^{2M-1} s_n \exp\left(-j \frac{2\pi nk}{2M}\right) \\ &= \begin{cases} \sqrt{2}a_0, & k = 0 \\ \frac{1}{\sqrt{2}}(a_k - jb_k), & k = 1, 2, \dots, M - 1 \\ \text{irrelevant}, & k \geq M. \end{cases} \end{aligned} \quad (3.89)$$

This shows that after ignoring the constant $1/\sqrt{2}$, when $k = 1, 2, \dots, M - 1$, the real and imaginary parts of the received signal after the DFT represent the in-phase and quadrature components of the transmitted symbol, respectively. Assuming that $a_0 = b_0$ at the transmitting end, the OFDM transmission signal can be demodulated using DFT technology at the receiving end.

3.7.3 Protection Interval and Cyclic Prefix

OFDM converts the input high-speed data stream into N parallel sub-channels, so that the data symbol period of each modulation subcarrier is N times of the original data symbol period, and the ratio of the delay spread to symbol period is thus reduced N times. To eliminate inter-symbol interference (ISI) to the maximum extent, guard interval can be inserted between each OFDM symbol, and the length of the protection interval is generally larger than the maximum delay extension of a wireless channel; thus, the multipath component of one symbol will not cause interference with the next symbol. In this protection interval, no signal can be inserted: it is a transmission free period. However, in this case, due to the influence of multipath propagation, inter-channel interference (ICI) occurs; that is, orthogonality between subcarriers is destroyed and interference between different subcarriers appears [47]. This effect is shown in Fig. 3.33.

Each OFDM symbol includes all non-zero subcarriers, and the delay signals of the OFDM symbol also appear; the delay signals of the first and second subcarriers are shown in Fig. 3.33: the two subcarriers will cause mutual interference.

To eliminate ICI caused by multipath fading, the OFDM symbol must be filled with a cyclic prefix (CP) in its guard interval, as shown in Fig. 3.34. In this way, it can be ensured that in the FFT period, the number of cycles of the waveform contained in the time delay copy of the OFDM symbol is also an integer, and the delay signal with a delay less than the guard interval T_g will not generate ICI during the demodulation process.

The introduction of CP is one of the key technologies in an OFDM system. Under certain conditions, CP can completely eliminate ISI caused by multipath propagation and suppress the influence of ICI without damaging the orthogonality of subcarriers. CP is a copy of a part of the tail of the OFDM symbol added to the front end of the OFDM symbol at the OFDM transmitter and removed at the receiver. The length of CP should be equal to the length of channel unit impulse response.

CP can act as a protection interval to eliminate ISI. With CP, the multipath copies of the former symbol fall within the CP range of the latter symbol, and the interference

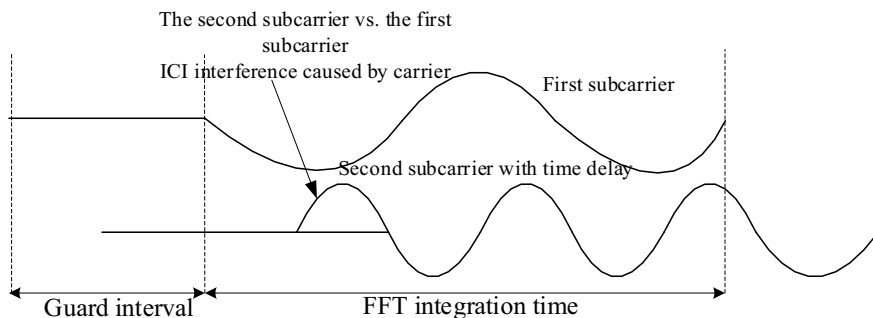


Fig. 3.33 Interference between subcarriers caused by idle protection interval due to multipath effect

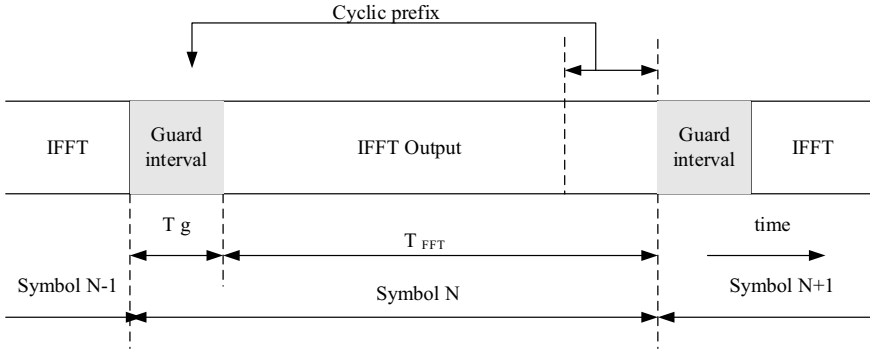


Fig. 3.34 Cyclic expansion of orthogonal frequency-division multiplexing symbols

between the first two symbols is eliminated. Due to the insertion of CP, a part of each OFDM symbol has certain periodicity, and the linear convolution of signal and channel impulse response is converted into cyclic convolution; each subcarrier will maintain orthogonality, thus preventing ICI.

3.7.4 Peak-to-Average Power Ratio and Its Reduction Method

The output of a multicarrier system is the superposition of multichannel signals. If the phases of multiple signals are consistent, the instantaneous power of the superimposed signals will be much higher than the average power of signals, resulting in a large peak-to-average power ratio (PAPR). Therefore, linearity of the amplifier in the transmitter is required. Otherwise, signal distortion may occur, the orthogonality between sub-channels will be destroyed, and the performance of the system will deteriorate.

The amplitude limiting technology adopts nonlinear process, which directly uses nonlinear operation to reduce the peak-to-average power ratio of the OFDM signal at or near the peak amplitude. The disadvantage of nonlinear process is that it causes signal distortion, which leads to the degradation of bit error rate performance of the entire system. Some of the amplitude limiting techniques deal with the signal after inverse Fourier transform and before interpolation. The processed signal must be interpolated before the digital to analog (D/A) conversion, which leads to peak regeneration. To avoid peak regeneration, the amplitude of the interpolated signal must be limited.

Coding technology can be used for the set of signal codewords for transmission. Only codewords with amplitude peaks lower than a certain threshold can be selected for transmission, thereby completely avoiding signal peaks. This type of technology is a linear process, which does not introduce limiting noise that is introduced in other amplitude limiting technology. The starting point of this type of technology is

not to reduce the maximum signal amplitude, but to lower the probability of peak occurrence.

3.8 Space–Time Coding

A multiple-input multiple-output (MIMO) system uses multiple antennas at the transmitter to transmit signals independently and simultaneously uses multiple antennas at the receiver to receive and restore the original information. Space–time coding is a signal coding technology that can achieve higher data transmission rates. It is a method that combines space and time in two-dimensional processing and uses multiple antennas to suppress channel fading. Figure 3.35 shows a block diagram of a MIMO system. There are several types of space–time coding methods, such as orthogonal space–time block coding, Bell layered space–time architecture, and space–time trellis coding.

3.8.1 Evolution of Space–Time Coding

The transmit diversity-space–time coding scheme proposed by Alamouti is shown in Fig. 3.36. In Alamouti space–time coding, the binary information bits sent by the source are first modulated (using constellation mapping), and then the modulated symbols x_1 and x_2 are respectively sent to the encoder. Then, they are encoded in the following way

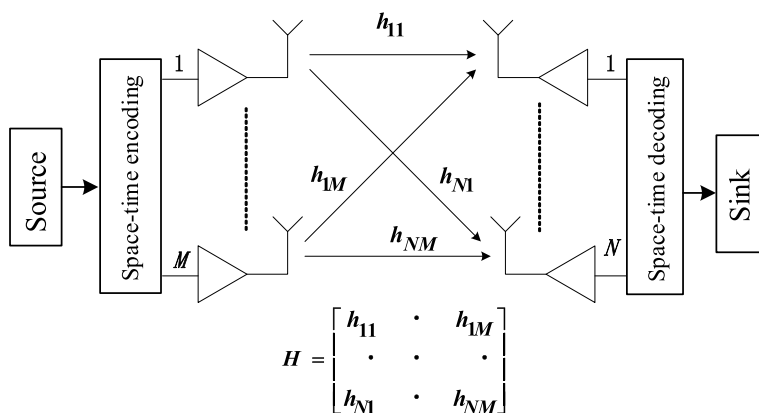


Fig. 3.35 Multiple-input multiple-output system

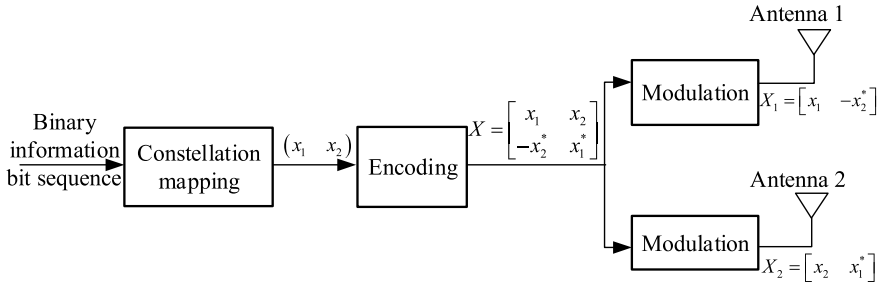


Fig. 3.36 Alamouti transmission diversity-space–time coding scheme

$$X = \begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix}. \quad (3.90)$$

The encoded symbols are transmitted from two antennas: in the first symbol transmission period, symbols x_1 and x_2 are simultaneously transmitted from antennas 1 and 2, respectively; in the second symbol transmission period, the symbols $-x_2^*$ and x_1^* are simultaneously transmitted from antennas 1 and 2. Here, x_1^* is the complex conjugate of x_1 . This method encodes information in both space and time domains. Let X_1 and X_2 denote the symbols sent from the transmitting antenna 1 and 2, respectively:

$$X_1 = (x_1 - x_2^*), \quad (3.91)$$

$$X_2 = (x_2 \ x_1^*), \quad (3.92)$$

$$X_1 X_2^T = x_1 x_2^* - x_1 x_2^* = 0. \quad (3.93)$$

The coding matrix has the following characteristics:

$$X X^T = \begin{bmatrix} |x_1|^2 + |x_2|^2 & 0 \\ 0 & |x_1|^2 + |x_2|^2 \end{bmatrix} = (|x_1|^2 + |x_2|^2) \mathbf{I}_2, \quad (3.94)$$

where \mathbf{I}_2 is an identity matrix with the dimension of 2×2 .

Figure 3.37 shows the receiver of the Alamouti space–time coding. Assume that the channel is a fast-fading channel; that is, the fading coefficient remains unchanged during two consecutive symbol transmission periods. Let $h_1(t)$ represent the channel fading coefficient from transmitting antenna 1 to the receiving antennas at time t , and $h_2(t)$ represent the channel fading coefficient from transmitting antenna 2 to the receiving antennas at time t ; they are Rayleigh distributed random variables. The received signal in two consecutive symbol periods at the receiving end can be expressed as

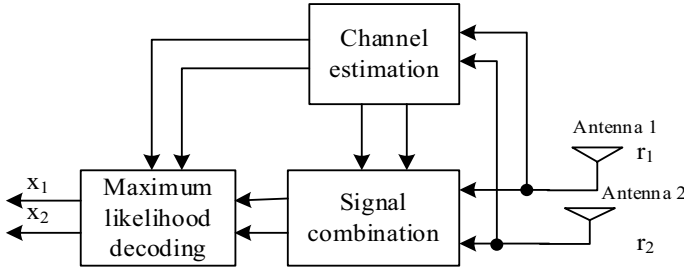


Fig. 3.37 The receiver of Alamouti space-time block code

$$r_1 = h_1 x_1 + h_2 x_2 + n_{01}, \quad (3.95)$$

$$r_2 = -h_1 x_2^* + h_2 x_1^* + n_{02}, \quad (3.96)$$

where r_1 and r_2 represent the received signal of the receiving antenna at time t and $t + T$, respectively, and n_{01} and n_{02} represent the independent complex Gaussian white noise of the receiving antenna at time t and $t + T$, respectively. Suppose that the receiver has the knowledge of the fading channel coefficients h_1 and h_2 and uses the maximum likelihood decoding method, that is, finds a pair of symbols (\hat{x}_1, \hat{x}_2) from the constellations so that the Euclidean distance of the following formula is the smallest

$$\begin{aligned} & d^2(r_1, h_1 \hat{x}_1 + h_2 \hat{x}_2) + d^2(r_2, -h_1 \hat{x}_2^* + h_2 \hat{x}_1^*) \\ &= |r_1 - h_1 \hat{x}_1 - h_2 \hat{x}_2|^2 + |r_2 + h_1 \hat{x}_2^* - h_2 \hat{x}_1^*|^2 \rightarrow \min \end{aligned} \quad (3.97)$$

Substituting Eqs. (3.95) and (3.96) into Eq. (3.97), can be changed into.

$$(\hat{x}_1, \hat{x}_2) = \arg \min (|h_1|^2 + |h_2|^2 - 1) (|\hat{x}_1|^2 + |\hat{x}_2|^2) + d^2(\tilde{x}_1, \hat{x}_1) + d^2(\tilde{x}_2, \hat{x}_2), \quad (3.98)$$

where $(\hat{x}_1, \hat{x}_2) \in C$ and C is the set of all possible transmitted symbols, and \tilde{x}_1, \tilde{x}_2 are the signals obtained by combining the channel fading coefficient and the received signal:

$$\tilde{x}_1 = h_1^* r_1 + h_2 r_2^* = (|h_1|^2 + |h_2|^2) x_1 + h_1^* n_{01} + h_2 n_{02}^*, \quad (3.99)$$

$$\tilde{x}_2 = h_2^* r_1 - h_1 r_2^* = (|h_1|^2 + |h_2|^2) x_2 - h_1 n_{02}^* + h_2 n_{01}. \quad (3.100)$$

In the case where the receiver has obtained the channel fading coefficients h_1 and h_2 , the combined signals \tilde{x}_1 and \tilde{x}_2 are functions of x_1 and x_2 , respectively. Therefore, Eq. (3.98) can be decomposed into two independent decoding algorithms:

$$\hat{x}_1 = \arg \min (|h_1|^2 + |h_2|^2 - 1) |\hat{x}_1|^2 + d^2(\tilde{x}_1, \hat{x}_1), \quad \hat{x}_1 \in S, \quad (3.101)$$

$$\hat{x}_2 = \arg \min (|h_1|^2 + |h_2|^2 - 1) |\hat{x}_2|^2 + d^2(\tilde{x}_2, \hat{x}_2), \quad \hat{x}_2 \in S, \quad (3.102)$$

where S is the modulation mapping constellation. If QAM or MPSK constellation is used, given the signal fading coefficient, $(|h_1|^2 + |h_2|^2 - 1)|\hat{x}_i|^2 (i = 1, 2)$ is constant for all signals. Therefore, the decision Eq. (3.101) and formula (3.102) can be simplified as

$$\hat{x}_1 = \arg \min d^2(\tilde{x}_1, \hat{x}_1), \quad \hat{x}_1 \in S, \quad (3.103)$$

$$\hat{x}_2 = \arg \min d^2(\tilde{x}_2, \hat{x}_2), \quad \hat{x}_2 \in S. \quad (3.104)$$

It can be seen from Eqs. (3.103) and (3.104) that the maximum likelihood decoding is completed if the codeword with the smallest measurement is found.

3.8.2 Space–Time Coding in Optical Wireless Communication

Simon and Vilnrotter proposed an improved Alamouti code. The coding method operates in the real number field and avoids negative and complex signals by complementing the symbols. The improved coding matrix is

$$X = \begin{bmatrix} x_1 & \bar{x}_2 \\ x_2 & x_1 \end{bmatrix}, \quad (3.105)$$

where \bar{x}_2 represents the complement of the symbol x_2 . In this coding method, the key is to obtain the complement of \bar{x}_2 ; for different modulation methods, the complement form of the symbol x , namely \bar{x} , will be given below. The complement is defined as

$$\bar{x}_i = A - x_i, \quad (3.106)$$

where A represents the amplitude of the pulse and is a constant related to the intensity of the emitted light. It can be seen from Eq. (3.106) that complementary code means the state inversion of the transmitted optical signals “on” and “off”. For OOK

modulation, the signal is assumed to be

$$\begin{cases} s_1(t) = 0 & 0 < t < T \\ s_2(t) = A & 0 < t \leq T, \end{cases} \quad (3.107)$$

where T represents the character period. Its complement code is

$$\begin{cases} \bar{s}_1(t) = A & 0 < t < T \\ \bar{s}_2(t) = 0 & 0 < t \leq T. \end{cases} \quad (3.108)$$

For binary PPM, it is assumed that the transmitted signal form is

$$\begin{aligned} s_1(t) &= \begin{cases} 0 & 0 < t < \frac{T}{2} \\ A & \frac{T}{2} \leq t \leq T \end{cases} \\ s_2(t) &= \begin{cases} A & 0 < t < \frac{T}{2} \\ 0 & \frac{T}{2} \leq t \leq T, \end{cases} \end{aligned} \quad (3.109)$$

Then, the corresponding complement form is

$$\begin{aligned} \bar{s}_1(t) &= \begin{cases} A & 0 < t < \frac{T}{2} \\ 0 & \frac{T}{2} \leq t \leq T \end{cases} \\ \bar{s}_2(t) &= \begin{cases} 0 & 0 < t < \frac{T}{2} \\ A & \frac{T}{2} \leq t \leq T, \end{cases} \end{aligned} \quad (3.110)$$

3.8.3 *Space-Time Decoding in Optical Wireless Communication*

For a 2×1 atmospheric laser communication system, according to the characteristics of the improved Alamouti code and the channel model of optical MIMO, mark r_1 and r_2 represent the signals received by the detector at time t and $t + T_s$, respectively:

$$\begin{cases} r_1 = \eta(h_1 C_1 + h_2 C_2) + n_1 \\ r_2 = \eta(h_1 \bar{C}_2 + h_2 C_1) + n_2, \end{cases} \quad (3.111)$$

where n_1 and n_2 represent Gaussian white noise with a mean value of 0 and a variance of N_0 , respectively. Combine the signals in the same way as the traditional Alamouti coding scheme:

$$\begin{cases} \tilde{x}_1 = h_1 r_1 + h_2 r_2 = \eta(h_1^2 + h_2^2)C_1 + \eta h_1 h_2 A + h_1 n_1 + h_2 n_2 \\ \tilde{x}_2 = h_2 r_2 - h_1 r_1 = \eta(h_1^2 + h_2^2)C_2 - \eta h_1 h_2 A + h_2 n_1 + h_1 n_2, \end{cases} \quad (3.112)$$

It can be seen from Eq. (3.112) that \tilde{x}_1 and \tilde{x}_2 are related not only to the sent codeword but also to $\eta h_1 h_2 A$. When $\eta h_1 h_2 A$ is a constant in two adjacent character periods, then the signal can be judged directly using Eq. (3.112), or the signal can be judged using the deformed formula:

$$\begin{cases} \tilde{x}_1 - \eta h_1 h_2 A = h_1 r_1 + h_2 r_2 = \eta(h_1^2 + h_2^2)C_1 + h_1 n_1 + h_2 n_2 \\ \tilde{x}_2 + \eta h_1 h_2 A = h_2 r_2 - h_1 r_1 = \eta(h_1^2 + h_2^2)C_2 + h_2 n_1 + h_1 n_2, \end{cases} \quad (3.113)$$

To find a space-time coding scheme suitable for any Q-ary PPM (QPPM), we generalize the coding method to any Q-PPM. Assuming that any Q-PPM signal is

$$s_i(t) = \begin{cases} A, & \frac{iT}{Q} \leq t \leq \frac{(i+1)T}{Q} \\ 0, & \text{else,} \end{cases} \quad (3.114)$$

Similar to binary PPM modulation, the complement form of $\bar{s}_i(t)$ is

$$\bar{s}_i(t) = \begin{cases} 0, & \frac{iT}{Q} \leq t \leq \frac{(i+1)T}{Q} \\ A, & \text{else.} \end{cases} \quad (3.115)$$

It can be seen from the Eq. (3.115) that although the complement form $\bar{s}_i(t)$ of $s_i(t)$ can be obtained for any Q-PPM, the signals of $\bar{s}_i(t)$ in other time slots are all A. This means that the transmission “on” The number of time slots, that is, the number of “on” time slots is $Q-1$, then the total power on each symbol will be changed at this time.

Space-time coding technology belongs to the category of diversity. It uses two-dimensional information of time and space to construct codewords, which can effectively suppress fading and improve transmission efficiency. Furthermore, it uses multiple antennas at the transmitter and receiver to realize parallel multichannel transmission and improve the spectrum utilization.

3.9 Channel Coding

3.9.1 Channel Coding

Channel coding is an important technical means to solve the problem of error control in information transmission and storage systems. The basic idea is to add some redundant check bits to a given digital sequence m according to certain rules so that the non-regular digital sequence m (usually called information element) is transformed into a

digital sequence c with certain regular constraints, called the code word sequence, to realize error control and protection of the digital sequence m . There is a one-to-one correspondence between m and c .

According to the purpose, channel coding can be divided into error detection code and error correction code. According to the value of the symbol, it can be divided into binary and q -ary codes, usually $q = p^m$, where p is a prime number, and m is a positive integer. According to different processing methods of information elements, error correction codes can be divided into two categories: block and convolutional codes. According to the relationship between the check and information elements, it can be divided into linear and nonlinear codes. If the relationship between the check and information elements is linear (i.e., satisfying the principle of linear superposition), it is called a linear code; otherwise, it is a nonlinear code.

3.9.2 Linear Error Correction Code

Linear block codes are a very important type of error correction codes. The basic idea is to add $r = n - k$ checksums to each k -bit information sequence according to a certain rule to form a sequence of length n , $c = (c_{n-1}, c_{n-2}, \dots, c_1, c_0)$; such a sequence is called a codeword (code group or code vector). In the case of a q -ary system, there are a total of q^k information groups; thus, after passing through the encoder, the corresponding codewords also have q^k , and the q^k codeword set (allowable code group or legal codeword set) is called the q -ary system (n, k) block code; its code rate is $R = k/n$. For digital communication and computer systems, the situation where $q = 2$ or $q = 2^m$ is usually of interest. From a mathematical point of view, the q -ary linear block code $[n, k]$ is a k -dimensional linear subspace $V_{n,k}$ in the n -dimensional linear space V_n on $GF(q)$. Since the linear subspace forms an abelian group under the addition operation, the linear block code is also called a group code. For $[n, k]$ linear block codes, the generator matrix $G_{n \times k}$ and check matrix $H_{(n-k) \times n}$ are usually used to describe the linear block code. The information sequence m is encoded as the codeword $c = n \cdot G_{n \times k}$, and all codeword sequences c satisfy the equation constraint $c \cdot H_{(n-k) \times n} = 0$.

Between two n -fold sequences (vectors) x and y , the number of different values of the corresponding position elements is called their Hamming distance and is represented by $d(x, y)$. In the (n, k) block code, the minimum distance between any two codewords is called the minimum Hamming distance d_{\min} (or abbreviated as d) of the block code. The block code is usually represented by (n, k, d) . When the code satisfies the linear relationship, it is recorded as the linear block code $[n, k, d]$, where the minimum distance d indicates its anti-interference ability. For any (n, k) block code, to detect all e random errors in the codeword, the minimum distance of the code is $d \geq e + 1$. To correct all t random errors, it is required that $d \geq 2t + 1$. To correct t random errors and detect $e \geq t$ errors at the same time, $d \geq t + e + 1$ is required.

For a linear block code, if any cyclic shift of its codeword is still a legal codeword sequence, it is called a cyclic code. A one-to-one correspondence is established between the code word sequence of the q -ary cyclic code and the polynomial whose coefficients are taken from the finite field $\text{GF}(q)$. The cyclic code can be obtained by a multiple of the generator polynomial $g(x)$. The BCH code, as a type of cyclic code that can correct multiple random errors, is usually described by the root of the generator polynomial $g(x)$. Consider any finite field $\text{GF}(q)$ and its extension $\text{GF}(q^m)$, where q is a prime number or a power of a prime number, and m is a positive integer. If the symbol is taken from a cyclic code on $\text{GF}(q)$, and the root set of its generator polynomial $g(x)$ contains the following $\delta-1$ continuous roots $\{\alpha^{m_0}, \alpha^{m_0+1}, \dots, \alpha^{m_0+\delta-2}\}$, then the cyclic code generated by $g(x)$ is called the q -ary BCH code, where α is the n -level element in $\text{GF}(q^m)$.

Reed–Solomon (RS) code is a type of multi-system BCH code with strong error correction ability, and it is a type of typical algebraic geometry code. The original BCH code with code length $n = q - 1$ on the finite field $\text{GF}(q)$, ($q \neq 2$), is called the RS code. One of the most important features of RS codes is that the symbols are taken from $\text{GF}(q)$, and the roots of the generator polynomial are also on $\text{GF}(q)$. Therefore, the RS code is the original BCH code whose symbol domain is consistent with its root domain. The RS code with code length $n = q-1$ and minimum distance δ , has a generator polynomial $g(x) = (x - \alpha^{m_0})(x - \alpha^{m_0+1}) \dots (x - \alpha^{m_0+\delta-2})$, usually with m_0 set to 1 or 0, where $m_0 = 1$ means $g(x) = (x - \alpha)(x - \alpha^2) \dots (x - \alpha^{\delta-1})$. This polynomial can generate a q -ary $[q-1, q-\delta, \delta]$ RS code with a minimum distance of δ . Since the maximum possible value of the minimum distance of the linear block code is the number of check elements plus 1, and the RS code meets this condition, the RS code is a type of maximum minimum distance separable code.

3.9.3 Convolutional Code

The convolutional code, denoted by (n_0, k_0, m) , uses certain rules for each piece of k_0 long information, adding $r_0 = n_0 - k_0$ check elements to form a code segment (or subcode) of length n_0 . Here, r_0 check elements are related not only to the information elements in the current code segment but also to m information elements in the previous code segment. The parameter m is called the convolutional code encoding storage and represents the unit of time that the input information group must be stored in the encoder. The parameter $N = m + 1$ is called the coding constraint degree and represents the number of code segments that are mutually constrained during the coding process. It corresponds to the code length n of the block code. In the convolutional code, $n_c = n_0 (m + 1)$ is called the coding constraint length, which means the number of mutually constrained symbols in the coding process, that is, k_0 information elements, are input from the encoder. The number of symbols affected in the code sequence when leaving. Thus, m and N are important parameters representing the complexity of the convolutional code encoder.

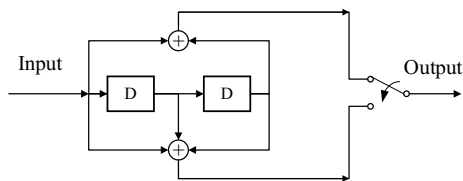


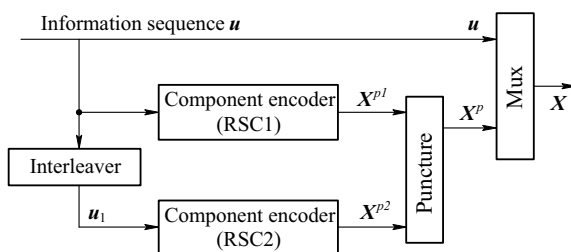
Fig. 3.38 (2, 1, 2) convolutional code encoder

Figure 3.38 is a (2, 1, 2) convolutional code encoder block diagram. The input information sequence is sent to the encoder with every $k_0 = 1$ as a section, and the convolutional code subcode sequence with $n_0 = 2$ is output; the code rate $R = 1/2$, code storage $m = 2$, code constraint degree $N = 3$, and coding constraint length $n_c = 6$. The process of convolutional code decoding is based on not only the subcode input to the decoder at the moment but also the subcode received over a long period of time, such as the m_d segment, per unit of time. The information element of the subcode is usually $m_d \geq m$. Here, $m_d + 1 = N_d$ is called the decoding constraint degree, and $n_0 N_d$ is the decoding constraint length, which represents the number of code segments or symbols that are mutually constrained during the decoding process.

Shannon's channel coding theorem states that random code that can reach the limit of channel capacity is a good code, but its decoding complexity is too high to be used for specific implementation. In 1993, Berrou et al. proposed a turbo code scheme in which two recursive systematic convolutional code encoders were randomly interleaved and cascaded, using soft output iterative decoding to approximate maximum likelihood decoding, and realized the idea of random coding. In 1996, MacKay et al. restudied the low-density parity-check (LDPC) code proposed by Gallager in 1963 and found that LDPC also has the characteristics of a random sparse graph of turbo codes and can achieve good performance approaching the Shannon capacity limit. Turbo and LDPC codes are collectively called sparse graph codes and are an important part of modern error correction code theory.

The turbo code encoder comprises two feedback system convolutional code encoders cascaded in parallel through a random interleaver, and the encoded check bits are punctured to obtain turbo code sequences with different code rates. Figure 3.39 shows a typical turbo code encoder structure block diagram. The information sequence $u = \{u_1, u_2, \dots, u_N\}$ passes through an N-bit interleaver to form

Fig. 3.39 Block diagram of a turbo code encoder



a reordered new sequence $u_1 = \{u'_1, u'_2, \dots, u'_N\}$; u and u_1 are respectively sent to the component code encoders (RSC 1 and RSC 2) of the two recursive systematic convolutional codes to obtain the corresponding check sequences X^{p1} and X^{p2} . To improve the code rate of the turbo code, some check bits are periodically deleted from the sequences X^{p1} and X^{p2} to obtain the turbo code check bit sequence X^p , which is multiplexed with the information sequence u to synthesize the turbo code sequence $X = (u, X^p)$. For example, if the code rates of the two component convolutional code encoders in Fig. 3.39 are both 1/2, then the (u, X^{p1}, X^{p2}) sequence corresponds to a turbo code sequence of 1/3 code rate. To obtain a turbo code with a rate of 1/2, the even-numbered bits of the RSC1 check sequence X^{p1} and the odd-numbered bits of the RSC2 check sequence X^{p2} can be deleted in sequence, and the obtained (u, X^p) is the 1/2 code rate turbo code sequence. The basic idea of turbo code decoding is to use the iterative decoding algorithm of soft input/soft output multiple times, through the transfer of external information between two component code decoders, to achieve the best decoding performance.

The LDPC code is a type of linear block error correction code with a sparse check matrix \mathbf{H} . The sparseness means that there are only a small number of non-zero elements in the check matrix \mathbf{H} and most of the elements are zero elements; that is, it has a low-density non-zero element distribution. Any type of linear block code can be represented with a two-way graph: the Tanner graph is a commonly used model description method. The Tanner graph of the LDPC code is composed of two types of node sets, variable nodes and check nodes, corresponding to n columns and $n - k$ rows of the check matrix $\mathbf{H}_{(n-k) \times n}$, respectively. There are no connections inside the same type of node set; only two points belonging to different sets may have a connection, and each connection corresponds to 1 in the check matrix.

Figure 3.40a shows a 4×8 check matrix \mathbf{H} of [4, 8] linear block code with row and column weights of 4 and 2; according to the check matrix, the Tanner graph can be drawn as shown in Fig. 3.40b. There are no connected edges in the variable node set (x_1, x_2, \dots, x_8) and check node set (z_1, z_2, z_3, z_4) in the graph, but there are lines between the two types of nodes. If there is a connection between a pair of variable nodes and a check node, it means that the variable bit participates in the check equation constraint of the check node; that is, it indicates the position of a “1” in a certain row of the check matrix. In the Tanner diagram of Fig. 3.40b, the four dashed lines form a directed closed cycle, which passes from the beginning to the

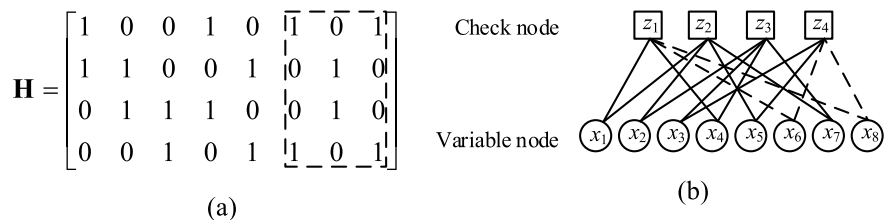


Fig. 3.40 a Check matrix and b Tanner graph representation of [4, 8] linear block codes with row and column weights of 4 and 2

end. In a Tanner graph of an LDPC code, each node has many closed cycles; the minimum length of the cycle is referred to as the shortest cycle of the node. Among the cycles formed by all nodes in the Tanner graph, the cycle with the smallest length is called the girth of the Tanner graph. For example, the girth of the Tanner graph in Fig. 3.40b is 4. The belief-propagation (BP) decoding algorithm of LDPC codes is based on the iterative update decision process of the symbol reliability information between the variable node and the check node in the Tanner graph, and the short cycle of length 4 considerably reduces the convergence performance of its decoding. Therefore, the design principle of the LDPC code must ensure that there are no cycles of length 4 in the Tanner graph corresponding to the \mathbf{H} matrix; that is, any two rows or two columns of the \mathbf{H} matrix have at most one position with a non-zero element at the same time.

3.10 Summary and Prospects

There are two modulation methods in optical wireless communication, the pulse position modulation and sinusoidal modulation. The pulse position modulation is simple but requires a large bandwidth. After the signal passes through the channel, the signal is distorted, and the channel capacity is limited. The sinusoidal modulation, a part of continuous wave modulation, occupies a narrow bandwidth, but advanced modulation methods can be used, and the channel consumption is large. However, nonlinear phenomena have considerable influence on continuous wave modulation. The internal modulation rate is low, and the external modulation rate is relatively high. The coding technique can suppress the influence of channel characteristics such as atmospheric turbulence. New coding methods, such as space–time coding and error correction coding, are the trend in optical wireless communication coding.

3.11 Questions

- 3.1 What is analog modulation? What is digital modulation?
- 3.2 What is the function of modulation?
- 3.3 What is passive modulation?
- 3.4 What are the types of subcarrier modulation? What is the basic principle of modulation?
- 3.5 What are the pulse-like position modulations? Which pulse-like position modulations require time synchronization? Which dose not?
- 3.6 Try to compare the required bandwidth between pulse like position modulation and subcarrier modulation.
- 3.7 What are the advantages of OFDM?
- 3.8 What are the space–time coding schemes? What are their characteristics?

- 3.9 Describe the working principle of direct drive of light source.
- 3.10 Compare the similarities and differences between electro-optic, acousto-optic and magneto-optic modulations.
- 3.11 What is the function of cyclic prefix in OFDM?
- 3.12 What is channel coding? How many types of channel coding are there?
- 3.13 What is error detection code? What is error correcting code?
- 3.14 Describe the coding principle of RS code briefly.
- 3.15 Describe the principle of LDPC code briefly.
- 3.16 Please explain that QAM is a combination of ASK and PSK modulations.

3.12 Exercises

- 3.1 Consider a pulse laser with a repetition rate of 200 kHz, PPM with pulse width of 3 ns and protection time slot width of 5 ms. Try to design the structure of the laser with 16-PPM and analyze the modulation rate of 16-PPM and 16-DPPM.
- 3.2 Try to analyze the channel capacity of PPM without background noise.
- 3.3 Try to analyze the detection problem of PPM under the background of white noise, and deduce the maximum likelihood detection algorithm.
- 3.4 Considering only the shot noise, try to analyze the relationships between the bit error rate of the optical wireless communication system, transmitted optical power, radius of the transmitting antenna, radius of the receiving aperture, and sensitivity of the detector.
- 3.5 For Alamouti's coding, calculate the coding gain, paired error probability and code word distance.
- 3.6 Taking the lateral modulation of a Lithium Niobate crystal for example, find the phase difference of two beams of light emitted from the crystal.
- 3.7 Try to analyze the optimal working conditions of acousto-optic modulation.
- 3.8 Try to analyze the signal-to-noise ratio of the wireless optical subcarrier modulation system under the conditions of amplitude, frequency and phase modulation.
- 3.9 Try to compare the bit error rate of pulse position modulation and subcarrier modulation.
- 3.10 Prove by calculation that, due to the addition of the cyclic prefix, a part of each OFDM symbol is periodic, and the linear convolution of the signal and the channel impulse response is converted into a cyclic convolution, and each subcarrier's orthogonality is thus maintained.
- 3.11 Prove that the ICI and ISI can be eliminated by adding cyclic prefix in OFDM.
- 3.12 Supposing a visible light communication system with M transmitters and N receivers, calculate the bit error rate under zero-forcing detection criterion.

- 3.13 Supposing a receiver with a working wavelength of $1\text{ }\mu\text{m}$ and bandwidth of 10 MHz, what is the average detection power for signal-to-noise ratio of 50 dB?
- 3.14 Try to prove that the complex orthogonal design of size n only exists when $n = 2$ or $n = 4$.
- 3.15 Suppose that a RF subcarrier, modulated by a baseband signal with unit power and bandwidth of 1 MHz, modulates an optical carrier at frequency 10^{14} Hz. This system is working in quantum limit state with required SNR threshold for demodulating the subcarrier of 20 dB.
- (1) How much received optical power is required for the subcarrier modulation system to work properly?
 - (2) If AM/IM is used, what is the baseband SNR after subcarrier demodulation?
 - (3) If FM/IM is used and the subcarrier frequency is offset by 10 MHz, what is the baseband SNR after subcarrier demodulation?
 - (4) What is the required bandwidth of the subcarrier mentioned in (2) and (3)?
 - (5) If the ratio of the background noise to the received signal power is 0.5 (no longer in quantum limit), how much optical power is required in (1)?
- 3.16 Consider an optical carrier, whose intensity is modulated by signal $m(t)$, transmitting through a fading channel. The fading effect is the transmission intensity multiplied by a multiplication term $q(t) = q_0(1 + r(t))$, where q_0 is a constant and the frequency spectrum of $r(t)$ is $S_r(\omega)$. Suppose that the frequency spectrum of $q(t)$ is covered by $S_m(\omega)$, the frequency spectrum of signal $m(t)$.
- (1) How to represent the fading by the frequency spectrum of the detected shot noise?
 - (2) Assuming that the signal $m(t)$ modulates the amplitude of the subcarrier before the carrier is modulated by the intensity of the subcarrier, try to determine the spectrums in this case.
- 3.17 Suppose an optical system with intensity modulation by signal $m(t)$ with power spectrum $S_m(\omega)$, where the Wiener filter is used after direct detection, omitting the background noises.
- (1) Assuming that $S_m(\omega) = S_0$ for $|\omega| \leq 2\pi B_m$, and when $S_0 B_m > 1$ and $S_0 B_m \leq 1$, plot the Wiener filter amplitude function.
 - (2) When $S_m(\omega) = \frac{C}{1+(\omega/2\pi B_m)^2}$ is reset, draw a sketch of the Wiener filter amplitude function and show the peak and half power frequency of the filter response.

- 3.18 Suppose an M -level pulse phase modulation system with a time error Δ . Prove that, if the terminal effects are considered (i.e., all possible pulse positions in two consecutive word intervals are considered), $PWE|\Delta$ is

$$PWE|\Delta = \left(\frac{M^2 - 2M + 2}{M^2}\right)P[K_1, K_2, (M - 2)K_3] \\ + \left(\frac{M - 1}{M^2}\right)P[K_1, 2K_2, (M - 1)K_3] + \left(\frac{M - 1}{M^2}\right)P[K_1, 0, (M - 1)K_3]$$

where $P[K_1, bK_2, CK_3]$ is the error probability of the correct interval count K_1 , the b -interval count is K_2 , and the C -interval count is K_3 . Here, K_1 , K_2 , and K_3 are Poisson counts with energy $[K_s(1 - \varepsilon) + K_s]$, $(\varepsilon K_s + K_b)$ and K_b .

- 3.19 The responsivity of a detector is the ratio of its output voltage to its input voltage. Noise equivalent power (NEP) is the amount of radio power per unit bandwidth required by the detector to generate output power equal to the noise power output by the detector. Detection degree (D) is defined as $1/\text{NEP}$. The parameter D^* is the detection degree of 1 cm^2 detector.

- (1) Try to derive the relationship between NEP and the responsivity mentioned above.
- (2) Try to prove $D^* = (A_r B_0)^{1/2} / \text{NEP}$, where B_0 is the optical frequency bandwidth and A_r is the receiving area.

- 3.20 Consider an optical system with working wavelength $0.6 \mu\text{m}$, photodetector's efficiency 50%, and load impedance 100Ω . Keeping the system temperature at 300 K,

- (1) Try to determine the range of signal count rate required to reach the limit of shot noise.
- (2) Convert (1) into power expressed in watts.
- (3) If the effective temperature of the black body background is 1000 K, try to determine the received signal power required to reach the quantum limit characteristics.
- (4) Determine the quantum limit SNR value obtained within 1 MHz bandwidth.

- 3.21 Consider the problem of estimating the intensity level of a monochrome field when the background noise count level is n_b . Calculate the MAP estimate of θ , the following equation must be satisfied for the prior density

- (1) Rayleigh distribution: $p(\theta) = \frac{\theta}{\sigma^2} e^{-\theta^2/2\sigma^2}$
- (2) x^2 square distribution: $p(\theta) = \frac{(2\sigma^2)^{-D}}{(D-1)!} \theta^{D-1} e^{-\theta/2\sigma^2}$, $D \geq 2$
- (3) Uniform distributed: $p(\theta) = \frac{1}{\theta_0}$, $0 \leq \theta \leq \theta_0$
- (4) In the case of the quantum limit condition $n_b = 0$, for the above three cases, determine the MAP estimate and the estimate variance.

- 3.22 Given the light intensity $n(t) = n_s[1 + s(t)]$, $|s(t)| \leq 1$, try to prove: if the peak power of the light field is limited, for all times t , $s(t) = \pm 1$, that is a square wave of any period, the largest pulse energy appears in the time interval $(0, T)$.

References

1. Fan C (2006) Principles of communication, 6th ed. National Defense Industry Press, Beijing
2. Ni X (2002) Electro-optical modulation and its application. *J Sun Yat-Sen Univ* 22(1):34–36
3. Wang B (2014) Photoelectric effect and its application. *China Extra-Sch Educ* (z1):991–992
4. Zhao J, Xue Z, Chen J et al (2005) The application of combined light modulation of Pockels effect and Faraday effect. *J Henan Univ (Natural Science Edition)* 35(3):23–27
5. Zhang F (2013) Talking about the optical Kerr effect. *Chin Times* 3:163
6. Han J, Liu J (2007) Engineering optics. Xidian University Press, Xi'an
7. Yang G, Song F (2008) Advanced physical optics. University of Science and Technology of China Press, Hefei
8. Zhang X, Wang P, Chang Z et al (2007) Development of acousto-optic modulation system driver. *Piezoelectric Acousto-Optic* 29(3):255–257
9. Xu S, Xiao S, Wang S et al (2014) Research on laser communication modulation system based on acousto-optic modulation. *Opt Commun Res* 5:59–62
10. Shi S, Wang X, Liu J et al (2008) Physical optics and applied optics, 2nd ed. Xidian University Press, Xi'an
11. Li Y, Li X, Li S et al (2007) Simulation and characteristic analysis of magneto-optical modulation. *J Northwest Univ (Natural Science Edition)* 37(5):719–723
12. Jin W, Liu X, Shi X et al (2013) Observation of the phenomenon of magneto-induced optical rotation. *Phys Bull* 8:76–78
13. Meng T, Fu Z, Liu H et al (2011) Simulation and experimental study of high-precision polarization angle measurement method based on magneto-optical modulation principle. *J Northwest Univ (Natural Science Edition)* 41(6):964–968
14. Shan N, Xiao S (2007) Application of faraday effect in optical communication and detection. *Infrared Laser Eng* 36(z2):624
15. Lewis PM (2007) Free-space optical communication link across 16 kilometers to a modulated Retro-Reflector array. University of Maryland, pp 22–24
16. Li Z (2008) Research on cat's eye optical inverse modulator with large field of view. University of Electronic Science and Technology of China
17. Zhao X, Gao Z, Zhang Y (2003) Laser reconnaissance technology based on "cat's eye" effect and its military application. *Opt Technol* 29(4):415–417
18. Bian X (2005) Analysis of the "cat's eye" effect of optical lens and its application in short-distance information exchange. University of Electronic Science and Technology of China, pp 12–13
19. Qing G, Wang X, Guo Y, Chen D, Zhang C (1995) Physical model and proof of "cat's eye effect." *Laser Technol* 04:244–248
20. Lu N (2009) Fourier optics. Mechanical Industry Press, Beijing
21. Wei B (2012) "Cat's Eye" reverse light modulation technology and application based on acoustic transducer. University of Electronic Science and Technology of China, pp 24–25
22. He T, Niu Y, Zhang P et al (2012) The influence of defocusing amount of photoelectric system on the echo power of "cat's eye" effect and its cause analysis. *Infrared Laser Eng* 41(11):2956–2960
23. Shay TM, MacCannell JA, Garrett CD et al (2004) First experimental demonstration of full-duplex communication on a single laser beam, vol 5160. SPIE, pp 265–271

24. Wang Q, Junique S, Almqvist S et al (2005) 1550 nm surface normal electro absorption modulators for free space optical communication, vol 5986. SPIE, p 598610
25. Ohgren J, Kullander F, Sjoqvist L et al (2007) A high-speed modulator communication link with a transmissive modulator in a cat's eye optics arrangement, vol 6736. SPIE, p 673619
26. Ke X, Yin Z (2008) Coding theory in wireless laser communication system. Science Press, Beijing
27. Sui MH, Yu XS, Zhou ZG (2009) The modified PPM modulation for underwater wireless optical communication. IEEE Comput Soc, 173–177
28. Xu Z, Shen L, Wang J, Liu Y (2009) Research on separated double pulse position modulation and its performance. J Commun 30(11):113–119
29. Kaluarachi ED, Ghassemlooy Z, Wilson B (2014) Digital pulse interval modulation for optical free space communication links. IEEE, pp 996–1000
30. Aldibbiat NM, Ghassemlooy Z, McLaughlin R (2002) Dual header pulse interval modulation for dispersive indoor optical wireless communication systems. IEEE ProcCircuits Deices Syst 149(3):187–192
31. Cheng G, Wang H, Sun X, Zhang T (2008) A new type of wireless optical communication modulation method. Chin Laser 35(12):1914–1918
32. Cheng G, Wang H, Sun X, Zhang T (2010) Dual pulse-pulse interval modulation for optical wireless communications. Chin Laser 37(07):1914–1918
33. Zhang K, Zhang H, Gong M et al (2003) Performance analysis of infrared dual-amplitude pulse interval modulation communication system. J Infrared Millim Waves 06:411–414
34. ADI (2006) AD8138 Data Sheet
35. MAXIM (2003) MAX9375 Data Sheet
36. MAXIM (2008) MAX3738 Data Sheet
37. Zeng Y, Xiong H, Zhu H et al (2012) Laser drive circuit design and optimization based on MAX3738. Opt Commun Technol 36(8):20–22
38. Tian G (2011) Research and analysis of optical transmitter APC and ATC circuit [J]. Electronic Design Engineering 19(3):18–20
39. Qi Z (2012) Research on the constant temperature control and driving method of low-power semiconductor lasers. Harbin University of Science and Technology, p 3
40. Jia L, Liu QJ, Taylor DP (2007) Optical communication using subcarrier PSK intensity modulation through atmospheric turbulence channels. IEEE Trans Comm 55(8)
41. Wu H, Li X, Yan H (2008) Analysis of bit error characteristics of atmospheric optical communication system in Gamma-Gamma turbulence channel. Acta Optica Sinica 12(28):99–104
42. Fan C, Zhang F, Xu B, Wu C (2001) Principles of communication. National Defense Industry Press, Beijing
43. Liu D (2004) The principle and application technology of Turbo code. Electronic Industry Press, Beijing, pp 126–129
44. Huang W, Nakagawa M (July 1994) Nonlinear effect of direct-sequence CDMA in optical transmission. In: IEEE 3rd international symposium on spread spectrum techniques & applications, (ISSSTA'94), Oulu, Finland, pp 1185–1192
45. Bekkali A, Naila CB, Kazaura K, Wakamori K, Matsumoto M (June 2010) Transmission analysis of OFDM-based wireless services over turbulent ratio-on-FSO links modeled by Gamma-Gamma distribution. IEEE Photonics J 2(3)
46. Ghassemlooy Z, Popoola WO, Leitgeb E (2007) Free-space optical communication using subcarrier modulation in Gamma-Gamma atmospheric turbulence. In: 2007 9th international conference on transparent optical networks, pp 156–160
47. Tong X, Luo T (2003) The principle and application of OFDM mobile communication technology. People's Posts and Telecommunications Press, Beijing, pp 36–37

Chapter 4

Atmospheric Channel, Channel Estimation, and Channel Equalization



Various natural phenomena in atmospheric channels, such as rain, snow, fog, haze, gas molecules, and aerosols, may cause energy attenuation of laser beams. When a light wave propagates in the atmosphere, the absorption and scattering of atmospheric gas molecules and aerosol particles will cause attenuation of the beam energy, and the uneven refractive index of air will cause the fluctuation of the light wave amplitude and phase. This chapter introduces the atmospheric channel characteristics, channel estimation, and channel equalization.

4.1 Atmospheric Attenuation

Atmospheric attenuation can be divided into atmospheric absorption attenuation and scattering attenuation; both are primarily caused by atmospheric molecules and aerosol particles. Both absorption and scattering weaken the intensity of the transmitted light radiation. The scattering of molecules has little effect on the attenuation of the light wave, but the absorption of molecules cannot be ignored in any light wave band. When light is transmitted in the atmospheric channel, it is absorbed and scattered by gas molecules (e.g., water vapor and carbon dioxide), water vapor condensates (e.g., ice crystals, snow, and fog), and suspended particles (e.g., dust, smoke, salt particles, and microorganisms) in the atmosphere, forming an absorption band in which the light radiation energy is attenuated [1].

4.1.1 Atmospheric Attenuation Coefficient and Transmittance

When a laser propagates in the atmosphere, a part of the radiation energy is absorbed and converted into other forms of energy, and another part of the energy is scattered

and deviates from the original propagation direction. Figure 4.1 shows a diagram of atmospheric attenuation.

As shown in Fig. 4.1, the monochromatic light wave with light intensity of $I(\lambda)$ passes through a thin atmospheric layer of a uniform medium. The light wave causes the charged particles in the thin layer to vibrate, and the energy provided for the particle vibration comes from the energy consumed by the light wave. A part of the energy of the light wave deviates from the original transmission direction due to scattering. Therefore, the radiation energy of the light wave decreases to $I'(\lambda)$ due to absorption and scattering after passing through the thin atmospheric layer. If the atmospheric attenuation coefficient is defined as β (with unit km^{-1}), the attenuation of light intensity is dI . Then, the relative change of light intensity is

$$\frac{dI}{I} = \frac{I' - I}{I} = -\beta dl. \quad (4.1)$$

The atmospheric transmittance is obtained by integrating Eq. (4.1). The expression of atmospheric transmittance τ after transmission distance L is

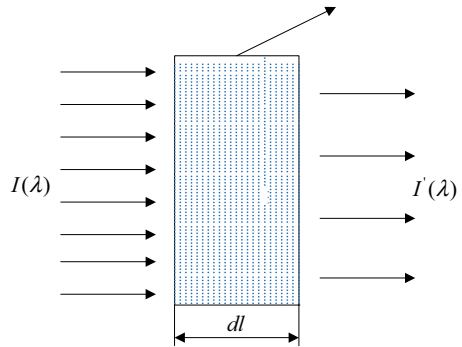
$$\tau = I/I_0 = \exp(-\beta L) \quad (4.2)$$

where I_0 and I represent the light intensity before and after passing through the atmosphere with width L , respectively. Equation (4.2) is Lambert's law, which describes the effects of attenuation and scattering on light propagation. Generally, the attenuation coefficient of atmospheric number is expressed as

$$\beta = \alpha_m + \alpha_a + \sigma_m + \sigma_a,$$

where α_m is the absorption coefficient of the atmospheric molecules, α_a is the absorption coefficient of the aerosol particles, σ_m is the scattering coefficient of the atmospheric molecules (Rayleigh scattering coefficient), and σ_a is the scattering coefficient of the aerosol particles.

Fig. 4.1 Model of atmospheric attenuation. $I(\lambda)$ is the radiation intensity of monochromatic light, dl is the atmospheric thin layer with uniform thickness, and $I'(\lambda)$ is the light radiation intensity after passing through the atmospheric thin layer



4.1.2 Absorption and Scattering of Atmospheric Molecules

The absorption of atmospheric molecules refers to the energy converted into molecular motion after a part of the light radiation energy interacts with them. The internal motion of polar molecules is generally composed of the atomic vibration of the constituent molecules, rotation of the molecules around their mass centers, and motion of electrons within the molecules. The natural frequency of a molecule is determined by its internal motion. The corresponding resonance absorption frequencies correspond to the ultraviolet, visible, near-infrared, mid-infrared, and far-infrared regions of the light wave, respectively. Therefore, the absorption characteristics of molecules depend on the frequency of the light waves.

The atmosphere is composed of many types of gases. If the normal air composition is calculated by volume fraction, nitrogen accounts for the most, approximately 78%. Oxygen accounts for approximately 21%, and rare gases, such as helium, neon, krypton, and xenon, and carbon dioxide account for approximately 0.03%. Other gases and impurities, such as ozone, nitric oxide, nitrogen dioxide, and water vapor, account for approximately 0.03%. Nitrogen and oxygen molecules account for the largest proportion (approximately 99%) in the atmosphere, but they have almost no absorption effect on visible (wavelength range 0.4–0.76 μm) and infrared regions. However, a strong absorption characteristic is shown in far-infrared (wavelength range 6–1000 μm) and near-infrared (wavelength ranged 0.75–1.5 μm). In addition to nitrogen and oxygen, there are helium, neon, krypton, xenon, ozone, and other rare gases in the atmosphere. Their molecules have considerable absorption spectral lines in visible and near-infrared light, but their proportion in atmospheric molecules is very small, so their absorption is generally not considered. Their absorption is considered only when other attenuation factors are weak enough at high altitude. Additionally, the absorption of atmospheric molecules is related to altitude. Due to its pure vibration structure, water molecules have very strong absorption capacity of visible light and near-infrared region, and the closer to the ground (several thousand meters), the higher the concentration of water vapor. Therefore, water vapor absorption ability is very strong near the ground. The main absorption spectral lines in visible and near-infrared regions are shown in Table 4.1.

It can be seen from Table 4.1 that the absorption of light of certain wavelengths by the atmosphere is very strong, which impedes the ability of the light wave to pass through. According to the difference of the ability of the atmosphere to absorb light of different wavelengths, the near-infrared region is divided into eight sections,

Table 4.1 Main absorption spectral lines in visible and near-infrared regions [1]

Absorptive molecules	Central wavelength of the main absorption line (μm)
H ₂ O	0.72, 0.82, 0.93, 0.94, 1.13, 1.38, 1.46, 1.87, 2.66, 3.15, 6.26, 11.7, 12.6, 13.5, 14.3
CO ₂	1.4, 1.6, 2.05, 4.3, 5.2, 9.4, 10.4
O ₂	4.7, 9.6

and the band with highest transmittance is called the atmospheric window. Within this window, atmospheric molecules show weak absorption properties. The laser wavelengths commonly used for communication are located in these windows [2].

Moreover, the density fluctuation of atmospheric molecules leads to inhomogeneity of the optical properties of the atmosphere, so scattering of the light waves passing through the large inhomogeneous gas field will occur. Scattering changes the direction of the light wave radiation; thus, some light waves do not continue to propagate in the original radiation direction. The scattering intensity depends on the contrast between the radius of the atmospheric molecules and the wavelength of the scattered light waves. When the light wavelength is much smaller than the particle radius, Rayleigh scattering theory is used for scattering analysis, whereas Mie scattering theory is used when the light wavelength is comparable to the particle radius. At all altitudes below 100 km, the atmospheric molecular coefficients σ_m can be approximately expressed as

$$\sigma_m = 4.56 \times 10^{-22} Ng \left(\frac{0.55}{\lambda} \right)^4 \quad (\text{km}^{-1}), \quad (4.3)$$

where λ is the wavelength with unit μm , and Ng is the number of molecules per unit of volume: $Ng = 2.55 \times 10^{-9} \text{ cm}^{-3}$ under standard atmospheric conditions at sea level.

4.1.3 Absorption and Scattering of Atmospheric Aerosol Particles

There are a large number of solid and liquid particles in the atmosphere, such as dust and Organic matter. These particles are generally in a gelatinous state in the atmosphere, so they are also called atmospheric aerosols. The size distribution of aerosol particles is very complex, and it is considerably affected by weather. When the wavelength (λ) of light is less than or equal to the size of aerosol particles (r), Mie scattering will occur:

$$\sigma_a \propto \lambda^{-2} \quad (r \geq \lambda). \quad (4.4)$$

When $r \geq \lambda$, whereas, the intensity is independent of the wavelength for scattering occurs when $r < \lambda$. The Mie scattering depends on the size and density distribution of the scattering particles, and the dependence on the wavelength is not as strong as that of Rayleigh scattering. The following empirical models are often used to estimate the attenuation coefficient of aerosol particles in engineering:

$$\beta_a = \frac{3.912}{V_M} \left(\frac{0.55}{\lambda} \right)^q, \quad (4.5)$$

where V_M is the atmospheric visibility distance with km as its unit, and $q = 0.585 V_M/3$; when $V_M \leq 6$ km, $q = 1.3$.

4.1.4 Atmospheric Window

Atmospheric window refers to the band with very low transmission loss and high transmission rate of electromagnetic radiation in the atmosphere. In general, the band with higher near-infrared transmittance is called atmospheric window, in which atmospheric molecules present weak absorption characteristics. The commonly used laser wavelengths are within these windows, as shown in Table 4.2.

Generally, the wavelength of the laser used in the atmospheric laser communication system is selected in the atmospheric window; thus, it can ignore the light intensity attenuation caused by atmospheric absorption, and the atmospheric attenuation effect is mainly composed of scattering. According to the relationship between the laser wavelength and scattering particle linearity, scattering can be divided into three types:

- (1) When the wavelength is larger than the diameter of scattering particles, Rayleigh scattering occurs. The influence of nitrogen and carbon dioxide on the light wave causes primarily Rayleigh scattering.
- (2) When the wavelength of the light wave is equal to the size of the scattering particles in the atmosphere, Mie scattering occurs. The primary influence of haze, fog, and aerosol particles on the signal light wave cause Mie scattering.

Table 4.2 Different atmospheric windows and their usage [2]

0.15–0.2 μm	The ultraviolet window has not yet been utilized
0.3–1.3 μm	Visible light is the main part, including some ultraviolet and infrared bands. Currently, it is widely used
1.4–1.9 μm	Near-infrared window, transmittance of 60–95%, cannot be film sensitive, only for spectrometer and ray detector
2.05–3.0 μm	The transmittance of near-infrared window is more than 80%, which is not sensitive to film. The window of 2.08–2.35 μm is useful in remote sensing
3.5–5.5 μm	Mid-infrared window, with 60–70% transmittance, is used for remote sensing of high-temperature targets, forest fires, volcanic eruptions, etc.
8–14 μm	When the physical temperature is 27 °C, the maximum emission intensity can be measured
In millimeter wave band	The window located in millimeter wave band has not been used or cannot be used by remote sensing
Wavelength > 1.5 cm	Microwave window, whose electromagnetic waves are not interfered with by atmosphere; is called the transparent window, and microwave remote sensing is all-weather

- (3) When the wavelength of the light is smaller than the diameter of the scattering particles, geometric scattering occurs. The influence of raindrops, snow, and fog on the signal light wave cause geometric scattering.

4.1.5 Estimation of the Attenuation Coefficient

When light waves are transmitted horizontally, the main attenuation of the bottom atmosphere is Mie scattering. At this time, the atmospheric transmittance can be expressed by the empirical formula related to visibility. Visibility is an indicator of atmospheric transparency, which is defined as the maximum distance that a standard blackbody can be identified with the normal naked eye when the observer observes horizontally with the sky as the background. The visibility represents the degree of atmospheric turbidity. In meteorology, visibility is generally divided into ten grades according to meteorological conditions, as shown in Table 4.3 [2].

4.1.6 Transfer Equation

The transmission equation of wireless laser communication channel is as follows:

$$P_{\text{receive}} = P_0 \times \gamma(L, \theta) \times \tau_{\text{atmosphere}} \times \tau_{\text{transmit}} \times \tau_{\text{receive}}, \quad (4.6)$$

where P_{receive} is the optical power received by the receiving antenna, τ_{transmit} and τ_{receive} are the transmittance of the transmitting and receiving optical systems, respectively, and P_0 is the output power of the semiconductor laser; $\gamma(L, \theta)$ is the beam spreading loss calculated as the ratio of the received power to the total power (Eq. 4.7), where L is the transmission distance, θ is the divergence angle of the beam, and d is the aperture of the receiving lens. Finally, the atmospheric transmittance $\tau_{\text{atmosphere}}$ can be calculated using Eq. (4.8).

Table 4.3 International visibility level [2]

Grade	Weather conditions	Visibility (m)	Scattering coefficient	Grade	Weather conditions	Visibility (m)	Scattering coefficient
0	Thick fog	<50	>78.2	5	Haze	2000–4000	1.96–0.954
1	Moderate fog	500–200	78.8–19.6	6	Light haze	4000–10,000	0.954–0.391
2	Mist	200–500	19.6–7.82	7	Sunny	10,000–20,000	0.391–0.196
3	Mist	500–1000	7.82–3.91	8	Very sunny	20,000–50,000	0.196–0.078
4	Mist	1000–2000	3.91–1.96	9	Extremely sunny	>50,000	0.0141

$$\gamma(L, \theta) = \frac{e^{-2[d/(L\theta)]^2} - 1}{e^{-2} - 1} \quad (4.7)$$

$$\tau_{atmosphere} = e^{-(\alpha_m + \alpha_a + \beta_m + \beta_a)L} \quad (4.8)$$

4.2 Atmospheric Turbulence Model

The influence of various atmospheric effects on optical wireless communication is shown in Fig. 4.2 [1]. The atmospheric channel is a time-varying channel with memory. The atmospheric absorption effect leads to the decrease of the received optical power. The multiple scatterings of the laser signal by the atmosphere cause the laser transmission to produce multipath effect, which results in the broadening of the laser pulse in space and time, as well as inter-symbol interference in the receiver. The random fluctuation of the optical refractive index caused by atmospheric turbulence causes the laser signal to produce intensity fluctuation, beam drift, beam expansion, fluctuation of the angle of arrival, and other phenomena. Furthermore, it causes the received optical signal to be severely distorted, the communication error rate to increase, and even short-time communication interruption to occur. This seriously affects the stability and reliability of atmospheric optical communication [1].

The scattering, absorption, and turbulence effects on atmospheric light lead to the fluctuation of light intensity on the surface of photodetectors in the receiver; this reduces the signal-to-noise ratio and spatial resolution of the detector and results in a performance degradation of the photoelectric system in the actual atmospheric environment.

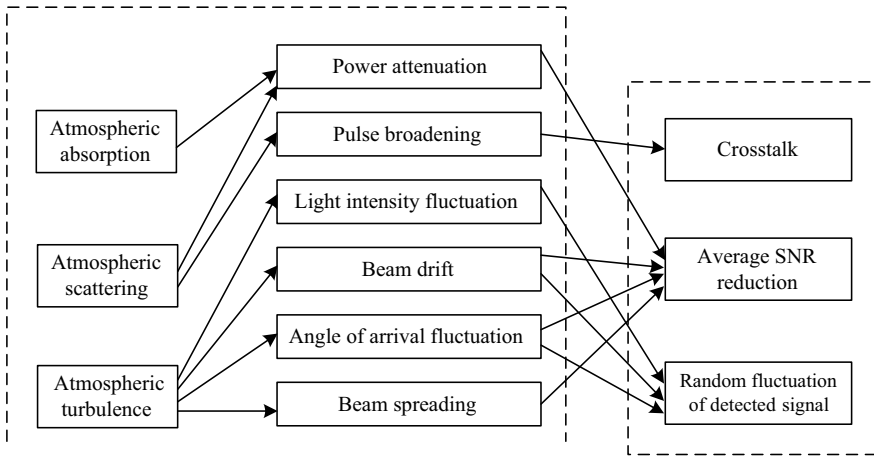


Fig. 4.2 Impact of various atmospheric effects on optical wireless communication

4.2.1 Atmospheric Turbulence

Atmospheric turbulence is caused by small random changes of atmospheric temperature and atmospheric wind speed caused by solar radiation and various meteorological factors. Random changes of atmospheric temperature cause random changes of atmospheric density; this leads to the change of atmospheric refractive index. The cumulative effect of these changes causes the atmospheric refractive index to be uneven. The fluctuation of the atmospheric refractive index caused by atmospheric turbulence is characterized by random fluctuation of laser light wave parameters (such as amplitude and phase), which eventually leads to beam scintillation, splitting, bending, expansion, and spatial coherence reduction. These are important factors that restrict the optical wireless communication system from reaching its maximum effectiveness.

Kolmogorov's theory indicates that a change of the average velocity of the turbulence increases its energy. The random fluctuation of atmospheric refractive index $n(r)$ is primarily caused by the random microstructure of temperature spatial distribution. This microstructure change is due to the large-scale non-uniformity of temperature caused by different regions of the earth's surface being heated differently by the sun. This large-scale temperature inhomogeneity leads to large-scale refractive index inhomogeneity, which is usually called turbulence vortex. Turbulence can be characterized by two scales. In the atmospheric boundary layer, the largest scale vortex, also known as the outer scale of turbulence, can be observed and analyzed. It is denoted by L_0 , and L_0 is usually in the range of tens to hundreds of meters. The minimum scale, represented by l_0 , is also known as the inner scale of turbulence, and l_0 is only a few millimeters [1], as shown in Fig. 4.3. When the beam propagates through these

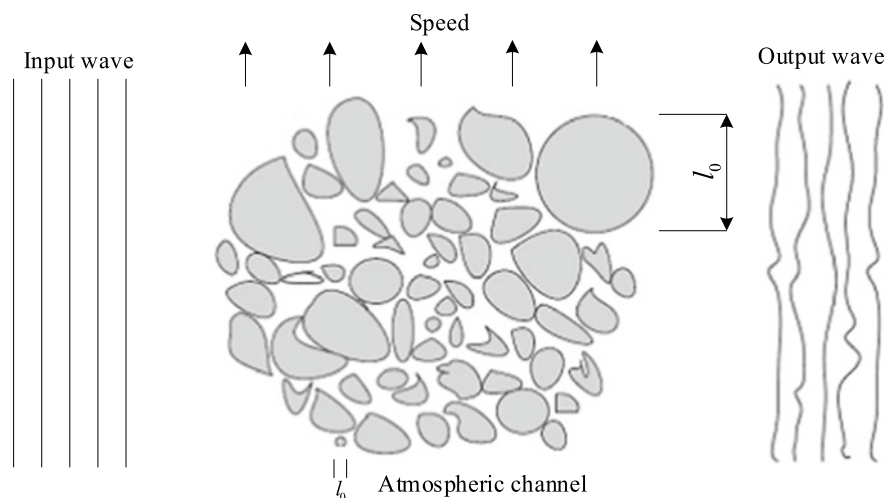


Fig. 4.3 Turbulent vortex in an atmospheric channel

vortices of different scales, the large-scale turbulence vortex primarily refracts the beam, whereas the small-scale vortex produces a diffraction effect on the beam.

Temperature, atmospheric refractive index, aerosol plasmid distribution, and other factors related to the formation of atmospheric turbulence will incur turbulence doping effect. The refractive index of light wave propagating in turbulent atmosphere is related only to the position of two points in space, which can be expressed as [1]

$$n(r, t) = n_0 + n_1(r, t), \quad (4.9)$$

where the refractive index $n(r, t)$ is related to time and position parameters, n_0 is the refractive index in free space (without turbulence), and $n_1(r, t)$ is the random fluctuation of refractive index around the average value n_0 , caused by atmospheric turbulence. The statistical characteristics of the random fluctuation of the refractive index in a turbulent flow field are very important for the study of laser propagation in atmospheric turbulence.

4.2.2 Atmospheric Turbulence Channel Mode

The atmospheric turbulence caused by the uneven atmospheric temperature and pressure leads to the random fluctuation of light intensity in time and space on the receiving surface; this is called “intensity scintillation effect”. The random fluctuation of light intensity at the receiving end is an important manifestation of atmospheric turbulence effect and a major factor affecting the performance of intensity modulation/direct detection optical communication system.

Taylor proposed that [1] under certain conditions, turbulence can be considered to be frozen. This means that the statistical characteristics of the temporal and spatial variation of the beam are caused by the component of the local wind perpendicular to the direction of the beam propagation. The coherent time of atmospheric turbulence t_0 is in milliseconds, which is very different from the typical data symbol time. Therefore, atmospheric turbulence channel can be used as a “slow fading channel” to describe the duration of data symbols statically [1]. In following subsections, the probability density functions of three different light intensity fluctuations and adaptations to different turbulence intensity link models are discussed in detail.

4.2.3 Log-Normal Turbulence Model

The Maxwell equation of light propagation in a medium is as follows:

$$\nabla^2 E + k^2 n^2 E = 0, \quad (4.10)$$

where k is the spatial wave number, n is the refractive index of a point r in space, E is the electric field vector at position r in space, and ∇ is the Hamiltonian operator. Different from the conventional wave equation, $n(r)$ in Eq. (4.10) is a function of position r . For atmospheric turbulence, the fluctuation of refractive index function $n(r)$ is a random process in time, so it must be described by statistical theory. Rytov approximation method [1] can be used to solve the equation of light wave propagation in atmospheric turbulence.

Since the three components of the electric field vector obey the same wave equation, the vector equation shown in Eq. (4.10) can be replaced by the scalar equation

$$\nabla^2 \tilde{u} + k^2 n(r)^2 \tilde{u} = 0, \quad (4.11)$$

where \tilde{u} is any field component of E_x , E_y , or E_z . Rytov transformation of Eq. (4.11) with respect to \tilde{u} yields

$$\psi = \ln(\tilde{u}). \quad (4.12)$$

Then, Eq. (4.11) is transformed into Riccati equation:

$$\nabla^2 \psi(r) + [\nabla \psi(r)]^2 + k^2 n^2(r) = 0. \quad (4.13)$$

For the earth's atmosphere, $n(r) = 1 + n_1(r)$, so Eq. (4.13) can be transformed into

$$\nabla^2 \psi + [\nabla \psi]^2 + k^2 [1 + n_1(r)]^2 = 0. \quad (4.14)$$

Let $\psi = \psi_0 + \psi_1 + \psi_2 + \dots$, and ψ_0 satisfies

$$\nabla^2 \psi_0 + [\nabla \psi_0]^2 + k^2 = 0. \quad (4.15)$$

Ignore all items higher than ψ_1 , and set $\psi = \psi_0 + \psi_1$ to obtain

$$\nabla^2 \psi_1 + \nabla \psi_1 (2\nabla \psi_0 + \nabla \psi_1) + 2k^2 n_1(r) + k^2 n_1^2(r) = 0. \quad (4.16)$$

For atmospheric turbulence, if $n_1(r) \ll 1$ and we assume $|\nabla \psi_1| \ll |\nabla \psi_0|$, then the second-order small quantities $(\nabla \psi_1)^2$ and $k^2 n_1^2(r)$ in Eq. (4.16) can be ignored. Finally, the following equation is obtained:

$$\nabla^2 \psi_1 + 2\nabla \psi_1 \nabla \psi_0 + 2k^2 n_1(r) = 0. \quad (4.17)$$

Since the order of magnitude $|\nabla \psi_0|$ is $k = 2\pi/\lambda$, $|\nabla \psi_1| \ll |\nabla \psi_0|$ can be written as

$$\lambda \nabla \psi_1 \ll 2\pi. \quad (4.18)$$

Equation (4.18) indicates that the change of ψ_1 is a small quantity at the distance of wavelength λ , which can be obtained from Eq. (4.13):

$$\tilde{u} = \exp(\psi_0 + \psi_1), \quad (4.19)$$

$$\tilde{u}_0 = \exp(\psi_0). \quad (4.20)$$

According to the central limit theorem, the solution of \tilde{u} obeys Gaussian distribution:

$$\frac{\tilde{u}}{\tilde{u}_0} = 1 + \frac{\tilde{u}_1}{\tilde{u}_0} = \exp(\psi_1), \quad (4.21)$$

and

$$\psi_1 = \ln\left(1 + \frac{\tilde{u}_1}{\tilde{u}_0}\right) \approx \frac{\tilde{u}_1}{\tilde{u}_0}. \quad (4.22)$$

Because $|\tilde{u}_1| \ll |\tilde{u}_0|$, Eq. (4.22) is approximately true. With $\psi_1 = \exp(-\psi_0)\tilde{u}_1$, Eq. (4.17) becomes

$$\nabla^2 \tilde{u}_1 + k^2 \tilde{u}_1 + 2k^2 n_1(r) \exp(\psi_0) = 0. \quad (4.23)$$

According to the scalar scattering theory, the solution is

$$\tilde{u}_1 = \frac{k^2}{2\pi} \iiint_V n_1(r') \tilde{u}_0(r') \frac{\exp(ik|r - r'|)}{|r - r'|} dV, \quad (4.24)$$

where V is the scattering volume. Since $\psi_1 \approx \tilde{u}_1/\tilde{u}_0$, we obtain

$$\psi_1(r) = \frac{k^2}{2\pi \tilde{u}_0(r')} \iiint_V n_1(r') \frac{\exp(ik|r - r'|)}{|r - r'|} dV. \quad (4.25)$$

Let the amplitude and phase of \tilde{u} be A and S , respectively, and the amplitude and phase of vacuum solution (undisturbed) \tilde{u}_0 be A_0 and S_0 , respectively. Then,

$$\tilde{u} = A \exp(iS), \quad (4.26)$$

$$\tilde{u}_0 = A_0 \exp(iS_0). \quad (4.27)$$

Therefore, we obtain

$$\psi_1(r) = \psi(r) - \psi_0(r) = \ln(A/A_0) + i(S - S_0). \quad (4.28)$$

The real and imaginary parts of record $\psi_1(r)$ are

$$\chi = \ln(A/A_0), \quad (4.29)$$

$$\delta = S - S_0, \quad (4.30)$$

where χ is the logarithmic amplitude fluctuation of light wave obeying Gaussian distribution, and δ represents the phase fluctuation of light wave obeying Gaussian distribution. The probability density function of log amplitude χ can be expressed as

$$P(\chi) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{-\frac{(\chi - E(\chi))^2}{2\sigma_x^2}\right\}, \quad (4.31)$$

where $E(\chi)$ is the mean value of χ , and σ_x^2 is the variance of logarithmic amplitude fluctuation.

Using the power spectrum of *Kolmogorov* refractive index fluctuation, the variance of logarithmic amplitude fluctuation σ_x^2 of plane wave propagation in atmospheric turbulence can be obtained [2] as follows.

For horizontal uniform path,

$$\sigma_x^2 = 0.307k^{7/6}L^{11/6}C_n^2. \quad (4.32)$$

For slant path,

$$\sigma_x^2 = 0.56k^{7/6}[\sec \varphi]^{11/6} \int_0^L C_n^2(x)(L-x)^{5/6}dx, \quad (4.33)$$

where φ is the zenith angle ($\varphi < 60^\circ$), and $\sec \varphi$ is the correction factor for the oblique path. Similarly, the variance of logarithmic amplitude fluctuation can be obtained by using *Kolmogorov* index fluctuation power spectrum.

For horizontal uniform path,

$$\sigma_x^2 = 0.124k^{7/6}L^{11/6}C_n^2. \quad (4.34)$$

For slant path,

$$\sigma_x^2 = 0.56k^{7/6}[\sec \varphi]^{11/6} \int_0^L C_n^2(x)(x/L)^{5/6}(L-x)^{5/6}dx. \quad (4.35)$$

If the amplitude of a light wave in atmospheric turbulence is known to be A , the light intensity of the light wave can be written as $I = A^2$. The variance of logarithmic intensity fluctuation σ_I^2 is defined as

$$\sigma_I^2 = \langle (\ln I - \langle \ln I \rangle)^2 \rangle. \quad (4.36)$$

For plane wave horizontal uniform path transmission, the variance of logarithmic intensity fluctuation can be written as

$$\sigma_I^2 = 1.23 k^{7/6} L^{11/6} C_n^2. \quad (4.37)$$

Equation (4.37) is also called Rytov variance. If the light intensity in free space (without turbulence) is $I_0 = A_0^2$, then the logarithmic light intensity is

$$l = \ln \left(\frac{A}{A_0} \right)^2 = 2\chi. \quad (4.38)$$

Therefore,

$$I = I_0 \exp(l). \quad (4.39)$$

To obtain the probability density function of the light wave intensity, variable substitution is used:

$$p(I) = p(\chi) \left| \frac{d\chi}{dI} \right|. \quad (4.40)$$

Substituting into Eq. (4.31), we obtain

$$P(I) = \frac{1}{\sqrt{2\pi} \sigma_I I} \exp \left\{ -\frac{(\ln(I/I_0) - E[l])^2}{2\sigma_I^2} \right\} I \geq 0, \quad (4.41)$$

where $\sigma_I^2 = 4\sigma_x^2$, $E[l] = 2E[\chi]$.

Generally, scintillation index (normalized variance of light intensity fluctuation) σ_I^2 is used to characterize the intensity of light intensity fluctuation caused by atmospheric turbulence, and the scintillation index is defined as

$$\sigma_I^2 = \langle (I - \langle I \rangle)^2 \rangle / \langle I \rangle^2, \quad (4.42)$$

where I is the light intensity. In the Rytov approximation, $\sigma_I^2 = \exp(\sigma_I^2) - 1$.

With different σ_I^2 , the log-normal distributed probability density function of the received light intensity fluctuation is shown in Fig. 4.4, where the average light intensity is $E[I] = 1$. It can be seen that as σ_I^2 increases, the log-normal distribution

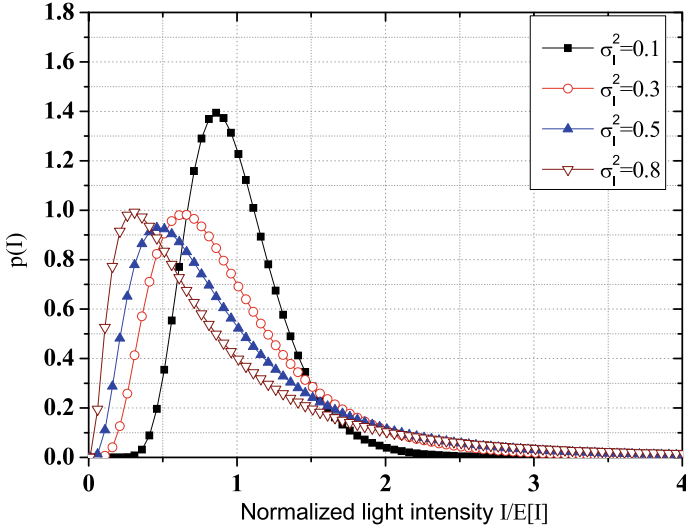


Fig. 4.4 Probability density function curve of intensity fluctuation of log-normal distribution

curve of light intensity fluctuation deviates from the average light intensity with a longer tail, and the approximation effect with log-normal distribution is less accurate. Simulation data and experiments show that the tail of the fluctuation distribution deviates from the log-normal distribution, so the statistical model of log-normal distribution is not suitable for describing the fluctuation behavior of light intensity in moderate to strong turbulence environments [1].

4.2.4 Gamma-Gamma Turbulence Model

Based on the assumption that the received light intensity fluctuation is modulated by small-scale turbulence fluctuation (diffraction effect) and large-scale turbulence fluctuation (refraction effect), Andrews proposed a probability distribution model of light intensity fluctuation [2]. In contrast to the log-normal distribution model, the probability distribution of light intensity fluctuation is a two-parameter model, and its parameters are closely related to the physical characteristics of the atmospheric turbulence.

The received light intensity is usually represented by a product [1] as $I = xy$, where x denotes the large-scale scattering coefficient and y represents the small-scale scattering coefficient. Assuming that x and y are independent random processes, the second moment of the received light intensity is

$$\langle I^2 \rangle = \langle x^2 \rangle \langle y^2 \rangle = (1 + \sigma_x^2)(1 + \sigma_y^2), \quad (4.43)$$

where σ_x^2 and σ_y^2 are the variances of x and y , respectively. To facilitate the calculation, the gamma-gamma distribution is used as the average light intensity $\langle I \rangle = 1$. The scintillation index is

$$\sigma_I^2 = (1 + \sigma_x^2)(1 + \sigma_y^2) - 1 = \sigma_x^2 + \sigma_y^2 + \sigma_x^2 \sigma_y^2; \quad (4.44)$$

x and y follow gamma distribution.

$$p_x(x) = \frac{\alpha(\alpha x)^{\alpha-1}}{\Gamma(\alpha)} \exp(-\alpha x), \quad x > 0, \alpha > 0 \quad (4.45)$$

$$p_y(y) = \frac{\beta(\beta y)^{\beta-1}}{\Gamma(\beta)} \exp(-\beta y), \quad y > 0, \beta > 0 \quad (4.46)$$

First, determine x and set $y = I/x$:

$$p_{I|x}(I|x) = \frac{\beta(\beta I/x)^{\beta-1}}{x\Gamma(\beta)} \exp(-\beta I/x), \quad I > 0. \quad (4.47)$$

According to the total probability formula,

$$p(I) = \int_0^{+\infty} p_y(I|x) p_x(x) dx = \frac{2(\alpha\beta)^{(\alpha+\beta)/2}}{\Gamma(\alpha)\Gamma(\beta)} I^{(\alpha+\beta)/2-1} K_{\alpha-\beta}[2(\alpha\beta I)^{1/2}], \quad I > 0. \quad (4.48)$$

Equation (4.48) is the gamma-gamma probability distribution function, also known as double gamma distribution. Here, α and β parameters represent the large- and small-scale scattering coefficients, respectively. $K_n(\cdot)$ is the second kind of modified *Bessel* function with order n , and $\Gamma(\cdot)$ is gamma function.

From the double gamma probability distribution function, we can obtain $\langle I^2 \rangle = (1 + 1/\alpha)(1 + 1/\beta)$. The parameters of large- and small-scale scattering are defined by Eq. (4.49):

$$\alpha = \frac{1}{\sigma_x^2}, \quad \beta = \frac{1}{\sigma_y^2}, \quad (4.49)$$

Since $\langle I \rangle = 1$,

$$\sigma_I^2 = \langle I^2 \rangle - \langle I \rangle^2, \quad (4.50)$$

and the relationship between the scattering index and the above parameters can be given by Eq. (4.42).

$$\sigma_I^2 = \frac{1}{\alpha} + \frac{1}{\beta} + \frac{1}{\alpha\beta}, \quad (4.51)$$

where α and β are related to the beam model. For plane waves, there is [1]

$$\alpha = \left[\exp \left(\frac{0.49\sigma_I^2}{(1 + 0.65d^2 + 1.11\sigma_I^{12/5})^{7/6}} \right) - 1 \right]^{-1}, \quad (4.52)$$

$$\beta = \left[\exp \left(\frac{0.51\sigma_I^2(1 + 0.69\sigma_I^{12/5})^{-5/6}}{(1 + 0.9d^2 + 0.62d^2\sigma_I^{12/5})^{5/6}} \right) - 1 \right]^{-1}, \quad (4.53)$$

where $\sigma_I^2 = 1.23C_n^2 k^{7/6} L^{11/6}$ is Rytov variance, $d = (kD^2/4L)^{1/2}$, $k = 2\pi/\lambda$ is the number of light waves, λ is the wavelength, D is the diameter of receiver aperture, and L is the transmission distance of the laser beam. C_n^2 is the structure constant of atmospheric refractive index. The light intensity scintillation index of the gamma-gamma distribution model is σ_I^2 .

$$\sigma_I^2 = \exp \left(\frac{0.49\sigma_I^2}{(1 + 0.65d^2 + 1.11\sigma_I^{12/5})^{7/6}} + \frac{0.51\sigma_I^2(1 + 0.69\sigma_I^{12/5})^{-5/6}}{(1 + 0.9d^2 + 0.62d^2\sigma_I^{12/5})^{5/6}} \right) - 1 \quad (4.54)$$

Compared with the log-normal distribution model, the *Gamma-Gamma* light intensity fluctuation probability distribution has a wider application range and can accurately describe the statistical characteristics of light intensity fluctuation in weak, medium, and strong fluctuation areas, and the tail part of the probability distribution is more consistent with numerical simulation and experimental results. The probability density function of the distribution is shown in Fig. 4.5, where weak, medium, and strong turbulence intensity is considered.

Figure 4.6 shows the curve of the variation of the scintillation index σ_I^2 of the *Gamma-Gamma* distribution model with the variance of *Rytov* according to Eq. (4.54). Figure 4.6 shows that with the increase of *Rytov* variance, the scintillation index gradually increases to the maximum value, which is greater than 1. Once the channel fading caused by turbulence reaches saturation, the scintillation index does not increase with the increase of *Rytov* variance. With the further increase of turbulence intensity, the logarithmic amplitude disturbance reaches saturation, and the turbulence disturbance caused by phase disturbance becomes the main part. The scintillation index no longer changes substantially with the variance of *Rytov*, which is consistent with the results of the classical atmospheric light wave scintillation theory.

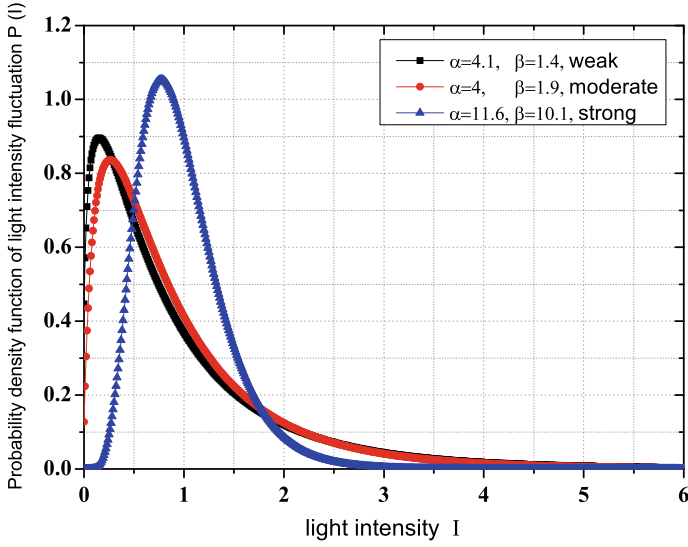


Fig. 4.5 Gamma-gamma distribution probability density function of light intensity

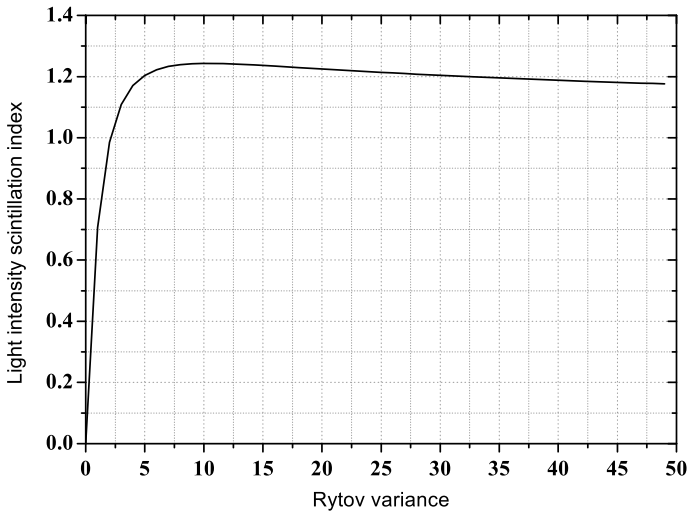


Fig. 4.6 Variation curve of the light intensity scintillation index with Rytov variance

The simulation of values α and β under different turbulence intensities according to Eqs. (4.52) and (4.53) is shown in Fig. 4.7. In the simulation, the diameter of the optical receiver is $d = 0$, which indicates a point receiver.

As can be seen from Fig. 4.7, in the case of very weak turbulence, $\alpha \gg 1$ and $\beta \gg 1$ show that the effective numbers of large-scale and small-scale scattering are large. When the variance of the logarithmic light intensity fluctuation increases gradually

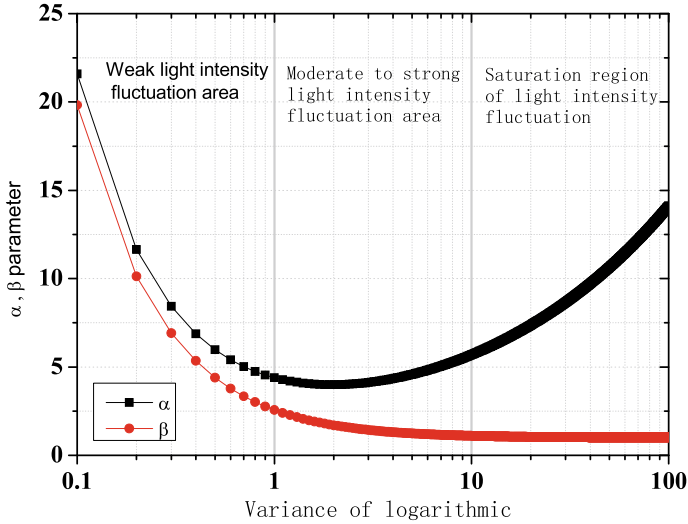


Fig. 4.7 Sum values at different turbulence intensities

(>0.2), the values of α and β decrease rapidly. When the turbulence intensity exceeds the moderate to strong turbulence area and reaches the turbulence intensity saturation area, $\beta \rightarrow 1$, it shows that the final value of effective number of small-scale scattering is from the horizontal space. However, the effective number of large-scale scattering increases again [3].

4.2.5 Negative Exponential Distributed Turbulence Model

As the turbulence intensity increases, the log-normal distribution model of light intensity fluctuation considerably deviates from the experimental data. In the case of strong turbulence, the radiation field of a light wave can be approximately Gaussian distribution with zero mean, so the distribution of light intensity is approximate to negative exponential distribution. Negative exponential distribution is considered as the limit distribution of light intensity distribution and is only suitable for the saturation region. In reference [4], experiments have shown that the probability distribution of light intensity fluctuation obeys the negative exponential distribution in case of strong turbulence.

$$p(I) = \frac{1}{I_0} \exp\left(-\frac{I}{I_0}\right), \quad I > 0, \quad (4.55)$$

where $I_0 = E[I]$ is the average light intensity. When the light intensity fluctuation reaches saturation state, the light intensity scintillation index approaches 1. The probability density function curve of light intensity fluctuation is shown in Fig. 4.8.

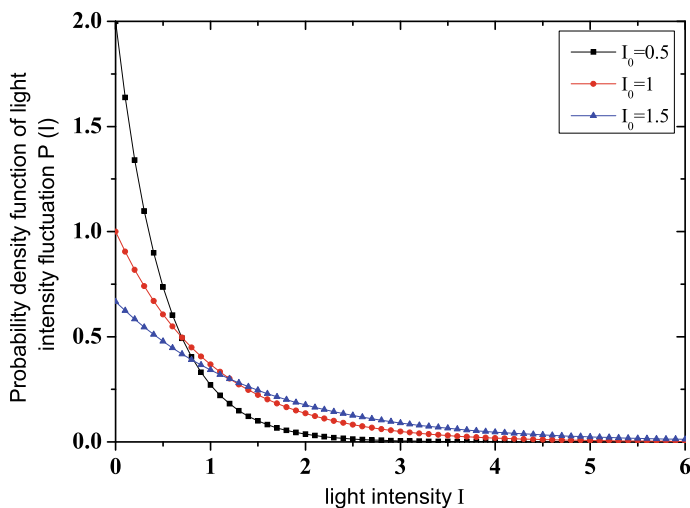


Fig. 4.8 Probability density function of light intensity fluctuation in negative exponential distribution

Figure 4.9 shows the probability distribution curve of light intensity fluctuation when Rytov variances are 0.2, 3.5, and 3.0 respectively, and the log-normal distribution and distribution results are given. It can be seen from Fig. 4.9 that in the weak turbulence region, the distribution is close to the log-normal distribution; in

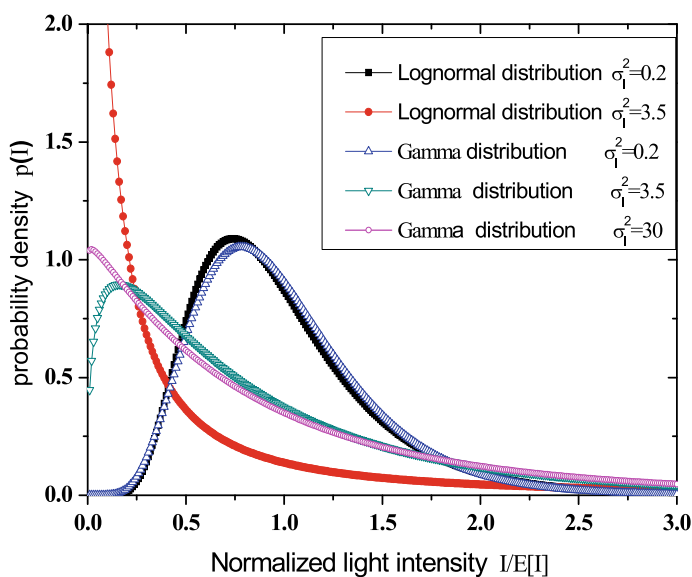


Fig. 4.9 Probability density function curve of light intensity distribution

the strong turbulence area, the gamma-gamma distribution gradually approaches the negative exponential distribution, but the log-normal distribution shows a serious deviation. Therefore, the gamma-gamma distribution is only suitable for the case of a wide range of normal turbulence intensity. In the weak turbulence region, the log-normal distribution model can fit the experimental data. However, in the strong turbulence area, the log-normal distribution model shows a large deviation from the experimental data, especially in the tail of the function curve, which directly affects the selection of the best decision threshold and the analysis of bit error rate. The distribution model is suitable for a wide range of turbulence intensities, whereas the negative exponential distribution is only suitable for the case of strong turbulence.

4.2.6 Atmospheric Structure Constant

For locally homogeneous and isotropic turbulence, the intensity of turbulence is usually expressed by the structure constant, and the atmospheric refractive index structure constant C_n^2 is an important parameter to represent the intensity of atmospheric optical turbulence [5]. Several models are described as follows.

4.2.6.1 Hufnagel-Valley Model

In Hufnagel-Valley model,

$$C_n^2(h) = 0.00594 \left(\frac{V}{27} \right)^2 \left(10^{-5} h \right)^{10} \exp \left(-\frac{h}{1000} \right) + 2.7 \times 10^{-16} \exp \left(-\frac{h}{1500} \right) + A \exp \left(-\frac{h}{100} \right), \quad (4.56)$$

where V is the wind speed, and H is the altitude in meters. In the modified Hufnagel-Valley model,

$$C_n^2(h) = 8.16 \times 10^{-54} h^{10} \exp \left(-\frac{h}{1000} \right) + 3.02 \times 10^{-17} \exp \left(-\frac{h}{1500} \right) + 1.90 \times 10^{-15} \exp \left(-\frac{h}{100} \right). \quad (4.57)$$

4.2.6.2 Model for Laser Communication with Submarines

The following model can be used to describe the communication between a satellite and submarine [6]:

$$C_n^2 = \begin{cases} 0 & 0 \leq h \leq 19 \text{ m} \\ 4.008 \times 10^{-13} & 10 \leq h \leq 230 \text{ m} \\ 1.300 \times 10^{-15} & 230 \text{ m} \leq h \leq 850 \text{ m} \\ 6.352 \times 10^{-7} h^{-2.966} & 850 \text{ m} \leq h \leq 7000 \text{ m} \\ 6.209 \times 10^{-16} h^{-6.229} & 7000 \text{ m} \leq h \leq 20000 \text{ m} \end{cases} \quad (4.58)$$

4.2.6.3 CLEAR I Model

The CLEAR I model is fitted by nighttime survey data of a desert in New Mexico (mid-latitude area, ground elevation of 1.23 km):

$$\log 10(c_n^2) = A + Bh + ch^2, \quad (4.59)$$

where

$$\begin{cases} 1.23 < h \leq 2.13 & A = -10.7025, B = -4.3507, C = 0.8141 \\ 2.13 < h \leq 10.14 & A = -16.2897, B = 0.0335, C = -0.0134 \end{cases}$$

$$\log 10(c_n^2) = A + Bh + ch^2 + D \exp(\{-0.5[(h - E)/F]^2\}) \quad (4.60)$$

where $10.34 < h \leq 30$, $A = -17.0577$, $B = -0.0449$, $C = -0.0005$, $D = 0.6181$, $E = 15.5617$, $F = 3.4666$.

4.2.7 Bit Error Rate Caused by Atmospheric Turbulence

The propagation equation of laser in atmospheric turbulence is

$$\nabla^2 \Psi(r, t) - \frac{n^2}{c^2} \cdot \frac{\partial^2 \Psi(r, t)}{\partial t^2} = 0. \quad (4.61)$$

The solution is

$$\begin{aligned} \Psi(r, t) &= A(r) \exp[i\varphi(r)] \exp(-i\omega t) \\ &= A_0(r) \exp[\chi + i\varphi(r)] \exp(-i\omega t), \end{aligned} \quad (4.62)$$

where $A_0(r)$ is the amplitude of the light wave when the beam is propagating in vacuum, $A(r)$ is the amplitude of the light wave when the beam is propagating in atmospheric turbulence, $\varphi(r)$ is the phase of the light wave in the presence of

atmospheric turbulence, and ω is the angular frequency. χ is the logarithmic amplitude fluctuation caused by turbulence, which has the following relationship with the fluctuation of logarithmic light intensity:

$$\ln \frac{I(r, t)}{I_0} = \ln \left[\frac{A(r)}{A_0(r)} \right]^2 = 2 \ln \frac{A(r)}{A_0(r)} = 2\chi. \quad (4.63)$$

Turbulent waves change the amplitude of the atmospheric light. Considering only the influence of atmospheric turbulence and no other noises on the bit error rate of a communication system, the amplitude change can be approximately regarded as the noise caused by the atmospheric turbulence. The variance expression of logarithmic intensity fluctuation is

$$\ln \frac{I(r, t)}{I_0} = 2 \ln \frac{A(r)}{A_0(r)} = 2 \frac{A_0(r) + A_i(r)}{A_0(r)} = 2 \ln(1 + \varepsilon), \quad (4.64)$$

where $A_i(r)$ is the noise amplitude, and $\varepsilon = A_i(r)/A_0(r)$ is the amplitude ratio of noise to signal. When ε is very small, $\chi = \ln(1 + \varepsilon) \approx \varepsilon$. If the signal intensity is I_0 and the noise intensity is $\langle I_n \rangle$, then the signal-to-noise ratio (SNR) caused by atmospheric turbulence can be expressed as [7]

$$\text{SNR} = \frac{I_0}{\langle I_n \rangle} = \frac{\langle A_0^2(r) \rangle}{\langle A_i^2(r) \rangle} = \frac{1}{\langle \varepsilon^2 \rangle} = \frac{1}{\langle \chi^2 \rangle}. \quad (4.65)$$

Under the condition of weak turbulence, the relationship between the difference of logarithmic intensity fluctuation and the variance of logarithmic amplitude fluctuation is

$$\sigma_I^2 = 4\chi^2. \quad (4.66)$$

Under the condition of strong turbulence, the simplified SNR can be approximately expressed as

$$\text{SNR} = \frac{1}{\langle \chi^2 + \chi^3 + \dots \rangle} \approx \frac{1}{\alpha \langle \chi^2 \rangle}, \quad 1 \leq \alpha \leq 2, \quad (4.67)$$

where α is the scintillation intensity factor. In a real laser communication system, the laser emitted from the transmitter can be regarded as a plane wave after collimation by optical lens; therefore, we analyze it as a plane wave. For plane waves, the logarithmic intensity fluctuation in weak turbulence is as follows:

$$\sigma_I^2 = 1.23 C_n^2 k^{7/6} L^{11/6}, \quad (4.68)$$

where C_n^2 is the atmospheric turbulence refractive index structure constant, $k = 2\pi/\lambda$ is the wave number, and L is the propagation distance. For digital laser communication system, the bit error rate of a receiver receiving optical signal is

$$\text{BER} = \frac{1}{2} \left[\text{erfc} \left(\frac{\text{SNR}}{\sqrt{2}} \right) \right]. \quad (4.69)$$

Combined with Eq. (4.67), the relationship between the bit error rate and variance of log intensity fluctuation is as follows:

$$\text{BER} = \frac{1}{2} \left[\text{erfc} \left(\frac{4}{\sqrt{2}\sigma_I^2} \right) \right] = \frac{1}{2} \left[\text{erfc} \left(\frac{4}{\sqrt{2} \times 1.23 C_n^2 k^{7/6} L^{11/6}} \right) \right]. \quad (4.70)$$

4.3 Diversity Reception

A general model of a spatial diversity system is shown in Fig. 4.10. M transmitters and N receivers ($M, N \geq 1$) constitute the $M \times N$ sub-channels. Diversity technology can suppress the fluctuation of light intensity caused by atmospheric turbulence by processing two or more uncorrelated signals. At the receiving end, the multichannel signal copies are combined by a certain combination rule, and the decision-maker decides on the merged signal. The common merging algorithms are maximum ratio combining, equal gain combining, and selection combining.

Suppose that the receiver receives N input signals. If the received signal of the first branch is $x_i(t)$ ($i = 1, 2, \dots, N$), and w_i is the weighting coefficient of the i th receiving branch, then the output signal $y(t)$ after merging can be expressed as

$$y(t) = w_1 x_1(t) + w_2 x_2(t) + \dots + w_M x_M(t) = \sum_{i=1}^M w_i x_i(t). \quad (4.71)$$

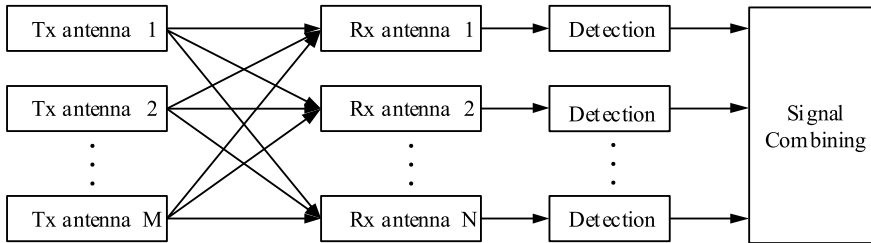


Fig. 4.10 Diagram of a space diversity system of an atmospheric laser communication system

To improve SNR, multiple signals must be combined. By selecting different weighting coefficients, different merger strategies are selected.

4.3.1 Maximum Ratio Combining

Maximum ratio combining (MRC) is the best way to combine multiple signals in the same phase. The weight is determined by the ratio of the signal power and noise power corresponding to each branch signal. The output SNR of maximum ratio combining is equal to the sum of all SNRs [8]. Therefore, even if the signal of any branch is very poor so that no single signal can be demodulated, maximum ratio combining can synthesize a signal that meets the SNR requirements and can be demodulated.

$$\text{SNR}_{\text{MRC}} = \frac{\eta^2}{2N^2\sigma_v^2} \sum_{n=1}^N I_n^2 = \sum_{n=1}^N \gamma_n, \quad (4.72)$$

where γ_n is the SNR of each branch, η is the photoelectric conversion efficiency, I_n is the received light intensity of each branch, N is the number of diversity branches, and σ_v^2 is the variance of additive white Gaussian noise.

4.3.2 Equal Gain Combining

Equal gain combining (EGC) does not require assigning a weight to the signal, and the signal of each branch is added with equal gain. This method simplifies the implementation of the merging method, and its performance is similar to that of maximum ratio merging [9]. In this case, the average SNR of the combined output is

$$\text{SNR}_{\text{EGC}} = \frac{\eta^2}{2N^2\sigma_v^2} \left(\sum_{n=1}^N I_n \right)^2, \quad (4.73)$$

where γ_n is the SNR of each branch, η is the photoelectric conversion efficiency, I_n is the received light intensity of each branch, N is the number of diversity branches, and σ_v^2 is the variance of additive white Gaussian noise.

4.3.3 *Selective Combining*

Selective combining (SC) detects signals of all diversity branches and selects the branch with the highest signal-to-noise ratio as the output of the combiner. In selection combining, one weighting factor is 1, and the others are 0. This method is simpler and easier to implement than the former two.

$$\gamma_{SC} = \max(\gamma_1, \gamma_2, \dots, \gamma_N) \quad (4.74)$$

Among the above three merging methods, equal gain merging is relatively simple and achieves good performance. Selecting the branch with the worst performance and discarding the rest results in a waste of resources, and maximum ratio merging achieves the best performance but is relatively complex. In engineering applications, the difficulty and performance of the implementation must be considered, and a reasonable compromise must be made.

4.4 Channel Estimation

4.4.1 *Concept of Channel Estimation*

Owing to the randomness of the atmospheric channel, the amplitude, phase, and frequency of the received signal will be distorted. Channel estimation is the process of estimating the model parameters (e.g., channel gain and signal-to-noise ratio) of a certain channel model from the received data. For linear channels, the impulse response of the channel is estimated.

Channel estimation [10] is a technology to obtain the channel state by processing the received signal according to certain criteria. There are three types of channel estimation algorithms: an algorithm based on a training sequence, blind estimation algorithm, and semi-blind estimation algorithm produced by the combination of the first two methods. Channel estimation based on a training sequence uses known information to estimate the channel. The transmitter periodically transmits a training sequence agreed upon by the sender and receiver. The training sequence can be a continuous block signal separated from the data information or can be evenly inserted in the data information [10].

The channel estimation based on training sequence can be used in general wireless communication systems. However, using the training sequence has some disadvantages: it reduces the effectiveness of channel transmission, wastes bandwidth, and reduces the spectrum efficiency. Because the entire frame signal can be extracted only after receiving the training sequence for channel estimation, a delay is inevitable, which limits the frame structure (e.g., the correlation time of the channel is less than the frame length). Blind channel estimation does not require a training sequence

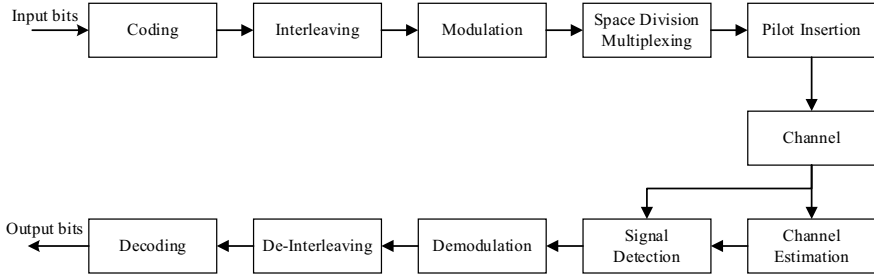


Fig. 4.11 Concept of channel estimation

but uses information inherent in the transmission data to realize channel estimation. Blind channel estimation can save bandwidth, but it requires a large amount of computation and has poor flexibility. Furthermore, the algorithm is difficult to implement. Therefore, its application in real-time systems is limited. The common channel estimation methods include the minimum mean square error (MMSE) and maximum likelihood (ML) [11, 12]. The process of channel estimation is shown in Fig. 4.11.

4.4.2 Least Squares Channel Estimation Algorithm

Suppose that the signal model of the system is represented by the following equation [13, 14]:

$$\mathbf{Y}_P = \mathbf{X}_P \mathbf{H} + \mathbf{W}_P, \quad (4.75)$$

where \mathbf{H} is the channel system function, \mathbf{X}_P is the known pilot signal, \mathbf{Y}_P is the received pilot signal, and \mathbf{W}_P is the additive white Gaussian noise vector. The least square channel estimation algorithm estimates the parameter \mathbf{H} in Eq. (4.68) to minimize the sum of squares error function e_P :

$$e_P = (\mathbf{Y}_P - \hat{\mathbf{Y}}_P)^H (\mathbf{Y}_P - \hat{\mathbf{Y}}_P) = (\mathbf{Y}_P - \mathbf{X}_P \hat{\mathbf{H}})^H (\mathbf{Y}_P - \mathbf{X}_P \hat{\mathbf{H}}), \quad (4.76)$$

where \mathbf{Y}_P is the vector composed of the received signals, $\hat{\mathbf{Y}}_P = \mathbf{X}_P \hat{\mathbf{H}}$ is the output signal obtained after channel estimation, and $\hat{\mathbf{H}}$ is the estimated value of the channel response. To minimize e_P , the following must be satisfied:

$$\frac{\partial e_P}{\partial \hat{\mathbf{H}}} = \frac{\partial \{(\mathbf{Y}_P - \mathbf{X}_P \hat{\mathbf{H}})^H (\mathbf{Y}_P - \mathbf{X}_P \hat{\mathbf{H}})\}}{\partial \hat{\mathbf{H}}} = 0. \quad (4.77)$$

The channel transfer function of the least squares algorithm is estimated as

$$\tilde{H}_{P,LS} = (X_P^H X_P)^{-1} X_P^H Y_P = X_P^{-1} Y_P. \quad (4.78)$$

The biggest advantage of the least square channel estimation algorithm is its simple structure and the small amount of calculation required. The channel characteristics of pilot position subcarriers can be obtained by only one division operation on each carrier.

4.4.3 MMSE Based Channel Estimation

Let $\hat{H}_{MMSE}(m)$ be the MMSE estimator of $H_p(m)$, and according to the MMSE estimation criterion, the value of $E \left| \hat{H}_{MMSE}(m) - H_p(m) \right|^2$ is expected to be the minimum.

$$\begin{aligned} \hat{H}_{MMSE}(m) &= R_{H_{p(m)} H_{LS(m)}} R^{-1} H_{LS}(m) \\ &= R_{H_{p(m)} H_{p(m)}} \left(R_{H_{p(m)} H_{p(m)}} + \delta_n^2 (X_p(m)^H X_p(m))^{-1} \right)^{-1} \hat{H}_{LS}(m), \end{aligned} \quad (4.79)$$

where δ_n^2 is the variance of Gaussian noise, the superscript H represents conjugate transposition, and

$$\begin{aligned} R_{H_{p(m)} H_{p(m)}} &= E \{ H_p(m) H_p(m)^H \}, \\ R_{H_{p(m)} \hat{H}_{LS}(m)} &= E \{ H_p(m) \hat{H}_{LS}(m)^H \}, \\ R_{H_{LS}(m) \hat{H}_{LS}(m)} &= E \{ H_{LS}(m) \hat{H}_{LS}(m)^H \}. \end{aligned}$$

The influence of noise is considered in MMSE estimation, so the mean square error of channel estimation is small. However, the MMSE algorithm requires a large amount of calculation, and its hardware implementation is difficult [15, 16].

4.5 Channel Equalization

Channel equalization technology aims to eliminate or suppress ISI caused by multi-path delay in broadband communication. According to the characteristics of constant parameter, variable parameter, and different data rate, the channel or the whole transmission system characteristics are compensated [13]. There are several ways to construct equilibrium. Generally, they can be divided into two categories: linear

equalization and nonlinear equalization. In practice, adaptive filters are used to achieve channel equalization.

4.5.1 ISI and Channel Equalization

In theory, the bandwidth occupied by the sharp square pulse waveform in the frequency domain is considered to be infinite. In optical wireless communication, the modulation of pulse system is preferable. When the pulse passes through a low-pass filter, it widens in the time domain and causes interference between adjacent pulses. When the channel bandwidth is much larger than the pulse bandwidth, the pulse broadening is small; when the channel bandwidth is close to the signal bandwidth, the broadening will exceed one symbol period, resulting in signal pulse overlap or ISI [17–19].

Crosstalk in baseband transmission communication system is inevitable. When the crosstalk is serious, the transfer function $H(\omega)$ of the system must be corrected so that it meets or approaches the characteristic of no ISI. If an adjustable or non-adjustable filter is inserted into the baseband system, the amplitude frequency and phase frequency characteristics of the entire system can be compensated to suppress the influence of ISI. This type of equalizer is called the equalization of the system.

Equalization is divided into frequency domain equalization and time domain equalization. Frequency domain equalization primarily considers the frequency response and ensures that the total transfer function of the entire system, including the equalizer, satisfies the condition of no distortion transmission. Time domain equalization considers the time response and ensures that the impulse response of the entire system, including the equalizer, satisfies the condition of no ISI.

Frequency domain equalization is applicable when the channel characteristics are unchanged and low-rate data is transmitted, whereas time domain equalization can be adjusted according to the changes of channel characteristics, effectively reducing the ISI; thus, it is widely used in high-speed data transmission.

4.5.2 Time Domain Equalization

Equalization technology produces the characteristics opposite to the channel characteristics by the equalizer at the receiving end to reduce, suppress, or eliminate the ISI caused by the time-varying multipath propagation characteristics of the channel.

It can be seen from Fig. 4.12 that the transmission function of the system without equalizer is

$$H(\omega) = G_T(\omega)H_T(\omega)G_R(\omega). \quad (4.80)$$

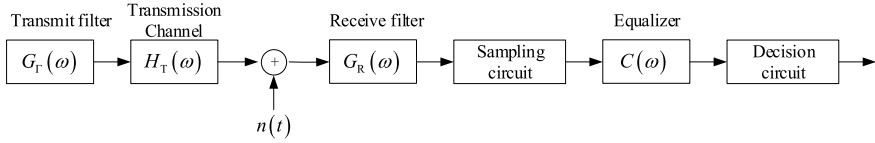


Fig. 4.12 Digital communication system model with equalizer

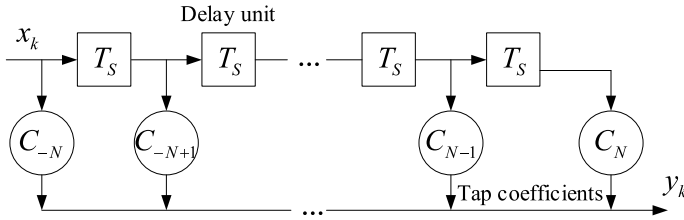


Fig. 4.13 Structure of a transverse filter

Generally, the transmitting and receiving filters are designed to match, and the equalizer is used to compensate for the channel distortion. After adding the equalizer, the transmission function of the entire digital communication system is as follows:

$$H'(\omega) = C(\omega)H(\omega). \quad (4.81)$$

Equalizers usually use filters to compensate for pulse distortion. The demodulated output sample obtained by the decision-maker is the sample that has been corrected by the equalizer or cleared of ISI. The basic structure of an equalizer is a transverse filter structure.

As shown in Fig. 4.13, the weighting coefficient of each tap is adjustable and set to a value that can eliminate ISI. Assuming that there are $(2n + 1)$ taps, the weighting coefficients are $C_{-N}, C_{-N+1}, \dots, C_N$. If the sample value sequence of an input waveform is $\{x_k\}$ and that of an output waveform is $\{y_k\}$, then

$$y_k = \sum_{i=-N}^N C_i x_{k-i} \quad (4.82)$$

for $k = -2N, \dots, 0, \dots, 2N$.

4.5.3 Linear Equalization

Time domain equalization can be divided into linear equalization and nonlinear equalization. If the result of the decision in the receiver is adjusted by the parameters

of the equalizer, it is a nonlinear equalizer; otherwise, it is a linear equalizer. There are two common algorithms: the zero forcing algorithm based on the minimum peak distortion and the least mean square algorithm based on the minimum mean square error [20–22].

4.5.3.1 Zero Forcing Algorithm

Zero forcing algorithm was proposed by Lucky in 1965. This method ignores the additive noise of the channel. Thus, in the presence of noise, the solution obtained by this algorithm is not necessarily optimal, but the algorithm is easy to implement. The peak distortion is defined as

$$D = \frac{1}{y_0} \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} |y_k|. \quad (4.83)$$

The physical meaning is that the distortion is the ratio of the sum of the absolute values of the system impulse response at all sampling times when $k \neq 0$ to the impulse response value at the sampling time of $k = 0$. From Eq. (4.83), the input peak error can be obtained as

$$D_0 = \frac{1}{x_0} \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} |x_k|, \quad (4.84)$$

and the output peak error as

$$D = \frac{1}{y_0} \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} |y_k|. \quad (4.85)$$

When the input peak error is $D_0 < 1$, the minimum value of the output peak error appears in $\begin{cases} y_0 = 1 \\ y_k = 0, 1 \leq |k| \leq N \end{cases}$. From Eq. (4.79), we can obtain the $2N + 1$ linear equations that the tap coefficient must satisfy; that is

$$\begin{cases} \sum_{i=-N}^N C_i x_{k-i} = 0, k = \pm 1, \pm 2, \dots, +N \\ \sum_{i=-N}^N C_i x_{-i}, k = 0 \end{cases}. \quad (4.86)$$

Written in matrix form,

$$\begin{bmatrix} x_0 & x_{-1} & \cdots & x_{-2N} \\ \vdots & \vdots & \dots & \vdots \\ x_N & x_{N-1} & \cdots & x_{-N} \\ \vdots & \vdots & \dots & \vdots \\ x_{2N} & x_{2N-1} & \cdots & x_0 \end{bmatrix} \begin{bmatrix} C_{-N} \\ C_{-N+1} \\ \vdots \\ C_0 \\ \vdots \\ C_{N-1} \\ C_N \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \quad (4.87)$$

It can be seen from the Eq. (4.87) that when the input sequence $\{x_k\}$ is given, adjusting or designing tap coefficient C_i according to the above equations can force each sample value y_k output by the equalizer to be zero.

4.5.3.2 Least Mean Square Algorithm

Another criterion to measure the equalization effect is the mean square distortion, which is defined as

$$e^2 = \frac{1}{y_0} \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} y_k^2. \quad (4.88)$$

Here, y_k is the sample value of impulse response after equalization. In adaptive equalization, the output waveform of the equalizer is no longer a single pulse impulse response but a data signal. Set the transmission sequence as $\{a_k\}$, and the output sample value sequence after equalization is $\{y_k\}$. Then, the error signal is

$$e_k = y_k - a_k. \quad (4.89)$$

In this case, the mean square error is defined as

$$\overline{e^2} = E(y_k - a_k)^2. \quad (4.90)$$

Substituting Eq. (4.88) into Eq. (4.90) we obtain

$$\overline{e^2} = E \left(\sum_{i=-N}^N C_i x_{k-i} - a_k \right)^2. \quad (4.91)$$

It can be seen that the mean square error $\overline{e^2}$ is a function of the gain of each tap. When the minimum mean square distortion is taken as the criterion, the equalizer should adjust its tap coefficients so that they meet the following requirements:

$$\frac{\partial \overline{e^2}}{\partial C_i} = 0, i = \pm 1, \pm 2, \dots, \pm N. \quad (4.92)$$

From Eqs. (4.84) and (4.85),

$$\frac{\partial \overline{e^2}}{\partial C_i} = 2E[e_k x_{k-i}] = 0, i = \pm 1, \pm 2, \dots, \pm N, \quad (4.93)$$

where

$$e_k = y_k - a_k = \sum_{i=-N}^N C_i x_{k-i} - a_k. \quad (4.94)$$

From Eq. (4.94), when the cross-correlation between the error signal and the input sample value is zero, the tap coefficient is optimal. Although the minimum mean square algorithm can balance the ISI caused by multipath transmission and the influence of additive noise, it cannot completely eliminate it.

4.6 Impacts of Atmospheric Turbulence on BER

For wireless optical communication systems, suppose that $s(t)$ is the modulated optical signal of the laser, $h(t)$ is the atmospheric transmission channel, $n_o(t)$ is the background light model, η is the photodetector conversion coefficient, $n_e(t)$ is the electrical noise model of the detector, and $y(t)$ is the output signal. The mathematical model of the system is

$$y(t) = [(s(t) + n_o(t)) * h(t)] \cdot \eta + n_e(t) = [(s(t) + n_o(t)) * f_\beta f_h f_A] \cdot \eta + n_e(t), \quad (4.95)$$

where $*$ represents convolution. From Eq. (4.95), the electrical noise $n_e(t)$ introduced by the photodetector, thermal movement of electrons in the resistance, and random generation and recombination of carriers in the semiconductor contributes the shot noise, resistance thermal noise, 1/f noise, and other noises in the photodetector, which are described as additive Gaussian noises with zero means. f_B is the light intensity attenuation term, f_h is the alignment error term, and f_A is the light intensity flicker term caused by turbulence. The multiplicative noise model fitted with Gaussian function corresponds to f_A in Eq. (4.95) is defined as

$$f_A = \frac{1}{\sqrt{2\pi} \times \sigma} \times e^{\left[-\frac{(I-\mu)^2}{2 \times \sigma^2}\right]}, \quad (4.96)$$

where I is the received light intensity, μ and σ^2 are the mean and variance of the multiplicative noise model. Figure 4.14 shows the Gaussian model of atmospheric turbulence multiplicative noise for optical wireless communication.

Therefore, for an optical wireless communication system, the impacts of atmospheric turbulence multiplicative noise f_A on the outage probability P_{out} and the bit error rate P_{BER} are derived as

$$P_{\text{out}} = F_A(\gamma_{th}) = \int_0^{\gamma_{th}} \frac{1}{\sqrt{2\pi} \times \sigma} \times e^{\left[-\frac{(x-\mu)^2}{2 \times \sigma^2}\right]} dx, \quad (4.97)$$

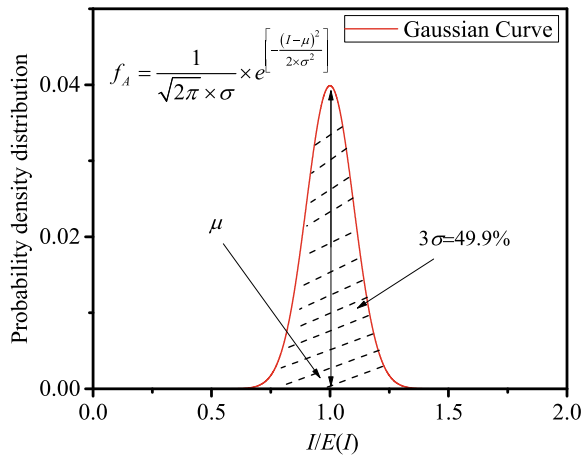
$$P_{\text{BER}} = \frac{q^p}{2\Gamma(p)} \int_0^\infty \exp(-qx) x^{p-1} F_A(x) dx. \quad (4.98)$$

In Eq. (4.98), different values of parameters p and q represent different modulation methods. For example, $p = 1$ and $q = 1$ represent the differential phase shift keying (DPSK).

Measurements of the impacts of atmospheric turbulence on the light intensity scintillation under different wavelengths and time periods are performed. Figure 4.15 shows the probability density distributions of light intensity scintillation after transmission with atmospheric turbulence in different time periods.

It can be seen that the impacts of atmospheric turbulence on an optical wireless communication system (i.e., the atmospheric turbulence multiplicative noise) are different in different time periods, and can be modeled by a Gaussian function.

Fig. 4.14 Gaussian model of atmospheric turbulence multiplicative noise



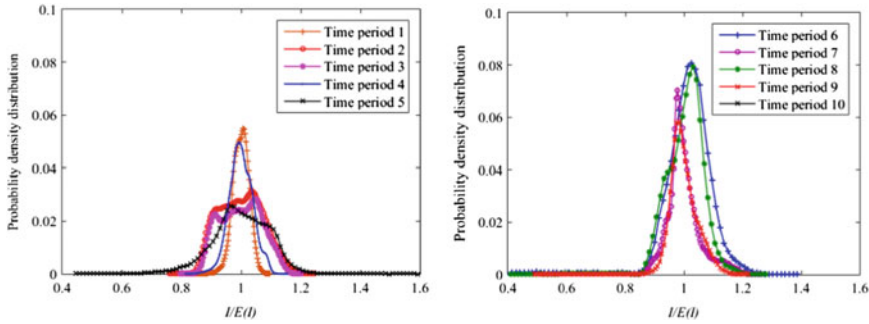


Fig. 4.15 Probability density distribution curve of light intensity scintillation transmitted by atmospheric turbulence in different time periods

4.7 Summary and Prospects

Channel is an indispensable part of a communication system. The channel causes beam drift and light intensity flicker in atmospheric laser communication, which has adverse effects on optical wireless communication. The purpose of analyzing the channel characteristics, such as atmospheric turbulence, is to suppress their influence on optical wireless communication. The key to improving the performance of optical wireless communication is to estimate the atmospheric channel and adopt appropriate channel coding, information equalization, and other measures to reduce the inter-symbol interference.

4.8 Questions

- 4.1 What is a coded channel? What is modulation channel? The similarities and differences are briefly described.
- 4.2 What is Mie scattering? What is Rayleigh scattering? What is geometric scattering? Try to compare the mechanism.
- 4.3 What is atmospheric window? What is visibility?
- 4.4 What is atmospheric turbulence? Three kinds of atmospheric turbulence models are briefly introduced.
- 4.5 Channel estimation and its common methods are briefly described.
- 4.6 Channel equalization and its function are briefly described.

4.9 Exercises

- 4.1 This paper analyzes the influence of atmospheric attenuation on wireless optical coherence detection, focusing on its influence on heterodyne efficiency.
- 4.2 It is assumed that under the condition of satellite to ground wireless laser, the influence of atmospheric turbulence on the bit error rate (BER) of the system under different zenith angles is analyzed.
- 4.3 Both rain and fog attenuate the wireless laser signal, and try to construct a calculation formula that can adapt to the coexistence of rain and fog, rainy and foggy days, and analyze its mechanism.
- 4.4 The logarithmic variance of laser propagation at 635, 780 and 1550 nm in different atmospheric turbulence is analyzed.
- 4.5 It is assumed that the laser transmitting power is 20 MW, the laser transmitting angle is 1mrad, the transmitting optical transmittance is 0.9, the receiving optical transmittance is 0.6, and the receiving aperture is 120 mm. Try to calculate the optical power that the receiving optical system can receive.
- 4.6 Try to calculate the laser emission power of optical wireless communication system under different conditions: (1) atmospheric attenuation coefficient 5 dB/km, communication distance 2 km, photo-detector sensitivity 1uW. Calculate the relationship between laser divergence angle and laser emission power; (2) beam divergence angle 0.2 mrad, communication distance 2 km, photo-detector sensitivity 1 uW. The relationship between laser emission power and atmospheric attenuation is analyzed. (3) The atmospheric attenuation coefficient is 5 dB/km, the beam divergence angle is 0.2 mrad, and the sensitivity of photo-detector is 1 uW. The relationship between laser transmitting power and transmission distance is analyzed.
- 4.7 A wireless laser communication system adopts 1550 nm laser wavelength, 500 MW transmitting power, 1mrad beam divergence angle, 130 mm receiving aperture and -36 dbm receiving sensitivity. The transmissivity of both transmitting and receiving optical systems is 0.5. (1) When the visibility is 20 km, what is the maximum communication distance of 1.25 gbps? (2) When visibility is 2 km, what is the maximum communication distance of 1.25 gbps?
- 4.8 Considering that atmospheric turbulence follows Gaussian distribution, the receiver model is shown in Fig. 4.16. P_s is the signal optical power, P_b is the background light power, $m(t)$ is the modulation signal, β is the modulation coefficient, h is the Planck constant, f is the optical carrier frequency, g is the APD gain, k is the Boltzmann constant, T is the load temperature, e is the electronic quantity, RL is the local load, try to detect the receiver directly under the atmospheric turbulence conditions Bit error rate.
- 4.9 The maximum zenith angle of satellite to ground laser downlink is not only related to communication laser power, beam divergence angle, background

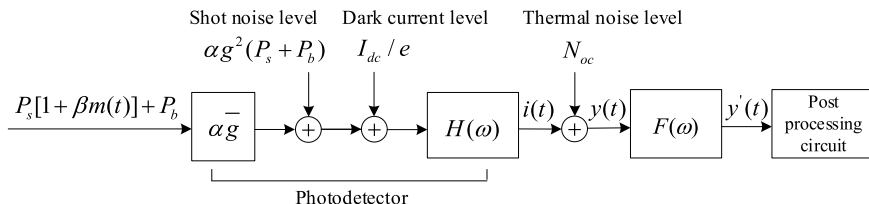


Fig. 4.16 Receiver model

light, bit error rate and other factors, but also affected by atmospheric turbulence caused by light intensity fluctuation and beam drift. The largest zenith angle of the satellite to ground laser downlink depends on the upper limit of bit error rate, and the influence of atmospheric turbulence on bit error rate is mainly caused by light intensity fluctuation, and the intensity of light intensity fluctuation depends on atmospheric turbulence intensity. Therefore, the influence of atmospheric turbulence on the maximum zenith angle is mainly caused by turbulence intensity. The relationship between BER and zenith angle is analyzed.

- 4.10 Try to analyze the attenuation of the aircraft to the satellite laser communication uplink. The specific parameters are shown in the table (Table 4.4).
- 4.11 Suppose there are two distinguishable paths, which delay one symbol time, we can write its transfer function

$$H(\omega) = 1 - e^{-j\omega\tau}$$

Table 4.4 Main link and environment parameters used in numerical simulation

Parameter	Value
Communication distance Z	4000 km
Aircraft altitude	8 km
Laser wavelength	810 nm
Extinction coefficient	0.001
Divergence angle of laser beam	25 urad
Receiver aperture area	415.48 cm ²
Zenith angle	30°
Radius of curvature of transmitting end wave front	∞
Beam radius of transmitter	2.15 cm
Satellite angular velocity	300 urad/s
Ground wind speed	5 m/s
Values near the ground	$3 \times 10^{-13} \text{ m}^{-2/3}$

In the discrete environment, we know from DSP that ω is usually replaced by operator $2\pi k/N$, where $k = \{0, 1, 2, \dots, N-1\}$. Try to calculate the channel amplitude and phase at different sampling values ($k = \{0, 1, 2, \dots, N-1\}$).

- 4.12 An optical wireless communication system is supposed to transmit a collimated beam with a diameter of 4 cm, the wavelength is 1.55 μm , and the receiving aperture is 10 cm, $C_n^2 = 7 \times 10^{-14} \text{ m}^{-2/3}$, try to use Kolmogorov spectrum to calculate:
- (1) The scintillation index on the receiving plane;
 - (2) The scintillation index on the detection plane.
- 4.13 The pulse broadening caused by atmospheric turbulence is analyzed. The wavelength of laser signal is 1.5 μm , the transmission distance is 100 km, and the zenith angle is 0.05.
- 4.14 Because the atmospheric refractive index is related to the laser wavelength, the pulse broadening at the receiving end will be different for different wavelength pulse signals. When the wavelength of laser signal is 0.85 μm , the broadening of pulse signal is calculated.
- 4.15 When the noise is very small, the BER and SNR depend on the logarithmic amplitude fluctuation of atmospheric turbulence. Try to deduce the relationship between SNR and BER in optical wireless communication.
- 4.16 This paper analyzes the bit error rate (BER) caused by light intensity flicker in OOK modulation and PPM modulation in optical wireless communication.
- 4.17 This paper analyzes the maximum likelihood detection of PPM modulation in atmospheric turbulence environment.
- 4.18 Considering the influence of atmospheric turbulence, the bit error rate of multi beam transmitting and receiving system is analyzed.
- 4.19 The bit error rate (BER) of optical wireless communication system is analyzed when the joint effect of atmospheric turbulence and quantum noise is analyzed.
- 4.20 A minimum mean square adaptive filter is designed, and the parameters of the filter and its working principle are analyzed.
- 4.21 In optical wireless communication, the BER of equal gain combining and optimal combining is discussed when N receiving antennas are used.
- 4.22 An optical wireless communication system with horizontal transmission, link distance of 2 km, transmitting lens diameter of 4 mm, receiving lens diameter of 100 mm, and working wavelength of 1.55 μm are investigated. The Kolmogorov spectral model is used to calculate the atmospheric structure constant $C_n^2 = 7 \times 10^{-14} \text{ m}^{-2/3}$.
- (1) The scintillation index of the power spectrum on the receiving plane;
 - (2) The scintillation index of the power spectrum on the transmitting plane.
 - (3) Calculate the attenuation probability below the critical value of 6 dB (assuming Γ - Γ distribution).

- 4.23 Spherical wave transmission horizontal link, communication distance 1 km, atmospheric structure constant $C_n^2 = 7 \times 10^{-14} \text{ m}^{-2/3}$, Kolmogorov spectrum and logarithmic spectrum were used to calculate the decay probability.
- (1) When the noise threshold is 7 and the average signal-to-noise ratio is 25 dB, the receiving aperture diameter is 150 mm.
 - (2) At that time $C_n^2 = 5 \times 10^{-13} \text{ m}^{-2/3}$, the calculation in (1) was repeated.
- 4.24 Spherical wave transmission horizontal link, communication distance 1 km, atmospheric structure constant $C_n^2 = 7 \times 10^{-14} \text{ m}^{-2/3}$, the BER is calculated by Kolmogorov spectrum and logarithmic spectrum respectively.
- (1) When the average SNR is 25 dB, the diameter of receiving aperture is 150 mm.
 - (2) When the average SNR is 25 dB, the diameter of receiving aperture is 300 mm.
- 4.25 When the probability of false alarm is 10^{-6} , the probability of detection is 0.999. Find the corresponding shot noise (or photo noise) Limited.
- 4.26 When the short-term signal-to-noise ratio (or photo noise) Limited corresponds to probability 1, the bit error rate of the detection system is calculated.
- 4.27 When the ratio of the probability of false alarm to bandwidth is 10^{-11} , the direct detection system is tested.
- (1) The corresponding threshold to noise ratio is calculated.
 - (2) When the probability of detection is 0.999, the corresponding shot noise (or photo noise) is limited.
 - (3) When the ratio of probability of false alarm to bandwidth is 10^{-9} , repeat the above calculation.
- 4.28 The condition is the same as 4.22. Considering the probability density function, the repeated calculation of 4.22 is carried out.
- 4.29 The laser transmission link is 3 km, other conditions are the same as 4.22.
- (1) The probability of fade is calculated by probability density function.
 - (2) The average fading time is calculated.
- 4.30 The Rytov approximation is used to replace the Helmholtz equation. It is proved that the Riccati equation is satisfied.
- 4.31 The transmission distance is 1 km, the diameter of Gaussian beam is 60 mm, and the wavelength is 1.55 m. The Kolmogorov spectrum is assumed to be the atmospheric structure constant $C_n^2 = 10^{-14} \text{ m}^{-2/3}$ for horizontal transmission. We know
- (1) Effective beam diameter

The Gaussian beam definition light field after the transmission distance L is defined as

$$U(r, L) = \frac{1}{p(L)} \exp(ikL) \exp\left(-\frac{r^2}{W^2} - i\frac{kr^2}{2F}\right)$$

F and W are defined as plane wavefront diameter and beam diameter

$$p(L) = 1 - \frac{L}{F_0} + i\frac{2L}{kW_0^2} = \Theta_0 + i\Lambda_0$$

The spot size can be expressed as

$$W_{LT}^2 = W^2(1 + T) = W^2(1 + 1.33\sigma_R^2\Lambda^{5/6})$$

The size of light spot can be divided into long-term term and short-term term.
Effective size of spot

$$W_{LT} = W\sqrt{1 + 1.33\sigma_R^2\Lambda^{5/4}} = 197 \text{ mm}$$

(2) Axial average light intensity

$$\langle I(0.L) \rangle = \frac{W_0^2}{W_{LT}^2} = 2.32 \text{ W/m}^2$$

(3) Horizontal scintillation index

$$\sigma_{I,L}(L) = 3.86\sigma_R^2 \text{Re}\left[i_1^{5/6} F_1\left(-\frac{5}{6}, \frac{11}{6}; \frac{17}{6}; \Theta + i\Lambda\right) - \frac{11}{16}\Lambda^{5/6}\right] = 0.015$$

For collimated beams, repeat the above calculation.

- 4.32 The diameter of collimated beam is 40 mm, the horizontal transmission is 1 km, and the atmospheric refractive index structure constant $C_n^2 = 0.75 \times 10^{-13} \text{ m}^{-2/3}$ is assumed to be Kolmogorov spectrum model. The minimum wavelength which can satisfy the scintillation index $\sigma_I^2(0.L) < 1$ is obtained.

References

1. Ke X, Yin Z (2008) Coding theory in wireless laser communication system. Science Press, Beijing
2. Han L, Wang Q, Xin T (2011) Performance of free space optical communication under gamma gamma gamma turbulence. Infrared and Low Light Level Eng 40(7):1318–1322
3. Luo T, Le G (2005) Principle and application of multi antenna wireless communication. Beijing Langdian University Press, Beijing

4. Simon MK, Vlnrotter V (2005) Alamouti-type space-time coding for free-space optical communication with direct detection. *IEEE Trans Wirel Commun* 4(1):35–39
5. Tarokh V, Seshadri N, Calderbank AR (1998) Space-time codes for high data rate wireless communications: performance criterion and code construction. *IEEE Trans Inform Theory* 3(44):744–765
6. Wilson SG, Brandt PM, Cao QL (2004) Optical MIMO transmission with Q-ary PPM for atmosphere channel. *IEEE Trans Commun* 53(8):1090–1094
7. Rrandt PM, Wilson S, Cao QL (2004) Code design for optical MIMO systems over fading channels. In: *Proceedings of the 38th Asilomar conference on signals, systems & computers, monterey*, pp 871–875
8. Has SM (2001) Pace-time codes for wireless optical channels. *IEEE Trans Inf Theory* 244
9. Alqudaha YA, Kavehrad M (2003) Orthogonal spatial coding in indoor wireless optical link reducing power and bandwidth requirements. *SPIE* 237–245
10. Ali S, Zaidi R, Afees M (2008) Cross layer design for orthogonal space time block coded optical MIMO systems. *Wirel Opt Commun Netw* 1–5
11. Wang H, Ke X (2009) Hybrid space time coding in free space optical communication. *Acta optica Sinica* 29(1):132–137
12. Wang H, Ke X (2010) Serial interference cancellation algorithm for layered space-time coding in FSO. *Optoelectron Eng* 37(2):1–6
13. Wang H, Ke X, Zhao L (2009) MIMO free space optical communication based on orthogonal space-time block coding. *Sci China* 39(8):896–902
14. Djordjevic IB, Vasic B, Neifeld AM (2007) LDPC coded orthogonal frequency division multiplex over the atmospheric turbulence channel. *Opt Express* 15(10):6332–6346
15. Stephanie S, Anne JV, Jean PC (2009) Soft decision LDPC decoding over chi-square based optical channels. *J Lightwave Technol* 27(16):3540–3545
16. Stephanie S, Damien F, Anne JV (2009) LDPC code design and performance analysis on 00K chi-square-based optical channels. *IEEE Photonics Technol Lett* 21(17):1190–1192
17. Djordjevic IB, Denic S, Anguita J (2008) LDPC-coded MIMO optical communication over the atmospheric turbulence channel. *J Lightwave Technol* 26(5):478–487
18. Tseng SM, Tsai S (1999) Performance of parallel concatenated convolutional coded on-off keying communication systems. *IEEE Photonics Technol Lett* 11(6):721–723
19. Wang H, Cao M, Jia K (2010) Cascaded space time block codes for atmospheric laser communication. *Optoelectron Eng* 37(12):116–121
20. Forney GD (1966) *Concatenated codes*. MIT Press, Cambridge
21. Clark GC Jr, Cain JB (1981) *Error-correction coding for digital communications*. Plenum Press, New York
22. Yuan D, Zhang L (2000) A theoretical method to calculate the performance of concatenated codes in Rayleigh channel. *Acta Commun Sinica* 21(10):86–89

Chapter 5

White LED Communication



The white light-emitting diode (LED) has the advantages of low power consumption, long service life, small size, easiness of driving, and environmental protection. LEDs are regarded as the fourth generation of energy-saving and environmentally friendly illumination products [1–3]. The white-light LED has a very high response sensitivity and good modulation characteristics, which make it suitable for high-speed wireless communication [4–9]. Therefore, white LEDs can be used as access points for wireless communication networks while illuminating. Combining the white LED lighting equipment with the base stations in indoor communication and connecting them to other communication networks comprise the visible light communication technology.

5.1 Light-Emitting Principle of LED

5.1.1 White LEDs

The wavelength range of visible light spectrum is 380–760 nm. The seven colors of light (red, orange, yellow, green, cyan, blue, and purple) that can be perceived by the human eye are all monochromatic. In contrast, the white light is a composite light composed of multiple monochromatic lights. For the LED to emit white light (white LED), its spectral characteristics must include the entire visible spectral range. The white light that human eye can see requires a mixture of at least two types of light. Generally, two-wavelength (blue + yellow light) or three-wavelength (blue + green + red) light emission mode is adopted.

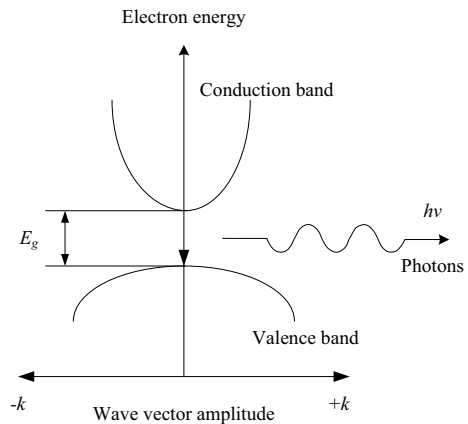
5.1.2 Light-Emitting Principle of LED

As shown in Fig. 5.1, when the diode PN junction is forward biased, a large number of majority carrier electrons on the N-type side cross the lowered barrier and are injected into the quasi-neutral region on the P-type side. It emits light with intensity roughly proportional to the driving current. The reason solid-state LEDs emit light is that a large number of minority carriers recombine and release a large number of photons [10].

To ensure that the LED has a higher recombination rate to release more photons, LEDs are generally made of direct band gap semiconductor materials. Direct band gap materials are compounds of III-V elements. This type of crystal materials primarily include GaAs, InP, InGaAsP, and AlGaAs [11, 12]. In these crystal materials, the minimum value of conduction band energy and the maximum value of valence band energy have the same wave vector value, and the crystal momenta of electrons and holes are almost equal, which causes most of the injected carriers to be eliminated due to the inter-band compound. Therefore, energy is released in the form of photons during the inter-band recombination process, and these photons become the light emitted by the LED.

Figure 5.2 is a schematic diagram of photon emission from a double heterojunction LED. Both the P and N regions of the junction are wide band gap materials, and this structure forms an electron potential well. Excess electrons diffuse from the N region through the depletion region when positively biased, and holes diffuse through the depletion region in the other direction. Carriers are easily trapped and confined in the potential well, thereby increasing the probability of recombination. Concurrently, the photons generated by recombination are also confined in the potential well region, which improves the photon emission rate. When the applied current density increases (e.g., by increasing the drive current), the optical power output by the device shows more nonlinear characteristics.

Fig. 5.1 Schematic diagram of light-emitting diode spontaneous emission photons



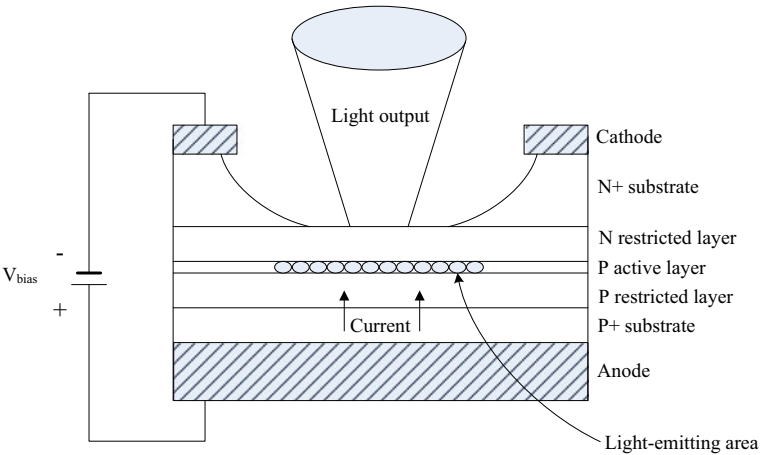
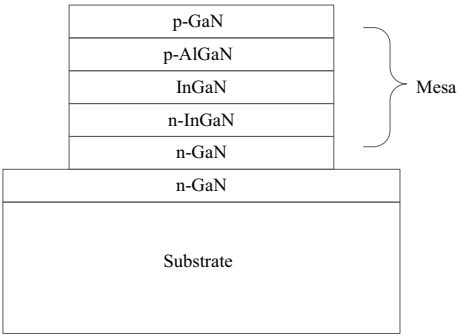


Fig. 5.2 Schematic diagram of a double heterojunction structure light-emitting diode

5.1.3 Light-Emitting Principle of White LED

Figure 5.3 shows a typical single quantum well white LED structure. The N-type GaN buffer layer is grown on a sapphire substrate. On the buffer layer between the N-type InGaN and P-type AlGaIn epitaxial layers is an undoped InGaIn quantum well, and the top layer is the P-type GaN layer. The structure must be partially etched to expose the N-type GaN layer. Ti/Al contacts and semi-transparent Ni/Au contacts are deposited on the N-type GaN and P-type GaN layers, respectively. Ni and Au are used as current spreading layers to solve the problem of low conductivity of P-type GaN materials. The wafer is usually cut into a rectangle and formed on the lead frame. Light can be emitted from the top P-type GaN layer in an upward direction. Other emission directions are realized by flip-chip packaging technology. The device impedance of the LED chip is primarily determined by the N-type GaN layer. The

Fig. 5.3 Single quantum hydrazine white light-emitting diode structure



typical unit area sheet resistance of this layer is 10–20 Ω , which also limits the lateral size of the LED device to approximately 300 μm .

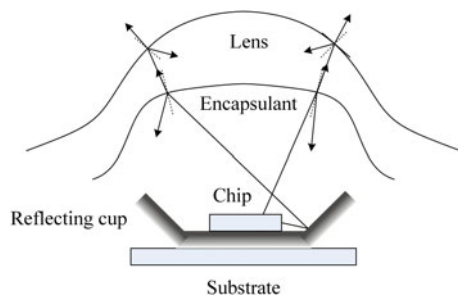
There are currently three main ways to make high-brightness white LEDs [10].

- (1) A blue light chip is used to excite yellow phosphor material to generate yellow light. The mixture of the blue and yellow light yields white light. This method is simple to implement, with high efficiency and low cost. At present, most white LEDs are constructed in this way. Since this generation of white light needs to excite phosphors, the modulation bandwidth is low, only tens of MHz, which cannot meet the needs of high-speed visible light communication systems.
- (2) An ultraviolet light chip is used to excite phosphors of three primary colors (i.e., blue, green, and red), and the mixed light becomes white light. The light generated in this way does not contain color cast, but the luminous efficiency is low, and the modulation bandwidth has difficulties meeting the requirements of high-speed visible light communication systems.
- (3) A combination of red, green, and blue light-emitting chips emits light that is white after mixing. Although this method has a high production cost, it has higher diffractive properties, does not need to excite phosphors to emit light, and has a high modulation response speed, making it very suitable for high-speed visible light communication system applications.

5.1.4 Lighting Model of White LED

The LED luminous mode is used to describe the light intensity distribution of the LED in different directions. Figure 5.4 shows a schematic diagram of photon emission from a high-power white LED. The white LED chip itself is a Lambert light source, which spontaneously radiates photons into the surrounding space. Part of the generated photons cannot be transmitted out of the LED due to factors such as material absorption and total internal reflection. The rest of the photons are transmitted out of the LED through a silicone resin sealing layer and lens, going through multiple reflections and refractions. Taking the center of the chip as the origin, the number of photons transmitted in different directions is different, making the light intensity in the different directions also different.

Fig. 5.4 Schematic diagram of white light-emitting diode photon emission



From a radiation point of view, the light-emitting area is a Lambert light source. The light emitted by the chip passes through different materials and types of surfaces before exiting the LED; this changes the initial light-emitting mode. The final light emitted by the LED is primarily composed of three parts: the light directly refracted from the sealed lens, light reflected inside the lens, and light reflected from the reflecting cup. From the perspective of light-emitting characteristics, the light in the LED is spontaneously radiated from the light-emitting chip, so the generated light is discontinuous and linearly superimposed in intensity. The final light-emitting mode depends on the geometry of the reflector cup and the sealed lens shape.

According to the path of the photon emission, the emitted light can be divided into two parts. One is the light refracted directly from the sealing layer and lens, and the other is the light refracted after multiple reflections on the surface of the sealing layer, lens, or reflecting cup. The final emission pattern is affected by the materials and geometric characteristics of the reflector cup, sealing layer, and lens. The light emission pattern of the diffusely reflected light from the rough surface of the reflecting cup and the diffusely refracted light passing through the sealing layer and the lens can be considered as a linear superposition of a cosine function or a Gaussian function [11–13]. The parameters that determine the final light-emitting mode are primarily the parameters of the light-emitting chip, reflector cup, sealing layer, and lens.

Considering all the above mentioned phenomena, the final light-emitting mode of the LED can be obtained by linear superposition of these three light-emitting modes. The surface of the LED chip is usually rough, and its light-emitting mode is generally Lambert mode. The radiation pattern of the diffusely reflected light from the rough surface of the reflector cup and the diffusely refracted light passing through the sealing layer and the lens can be considered a linear superposition of Gaussian power functions. From the point of view of light propagation, each light ray propagates through diffuse reflection or refraction and follows the Gaussian power function distribution. Therefore, the final radiation pattern is a linear superposition of several types of functions, with an angular offset along with the incident angle of each traced ray.

5.2 Background Noise Model for Internet of Vehicle

Since Guglielmo Marconi invented radio communication, atmospheric turbulence models based on Kolmogorov spectra have been widely known and intensively studied [14, 15]. Visible light communication belongs to one of the optical wireless communications [16, 17]. Indoor visible light communication is primarily used for wireless broadband networking and indoor positioning [18], with low background light noise interference, small communication range and low implementation complexity [19, 20]. Outdoor visible light communication is used to realize vehicle-to-road-based equipment and vehicle-to-vehicle traffic information sharing. Complex optical noises such as sunlight, high-brightness electronic billboards, and

a large number of artificial light sources at night make the photodetector prone to anomalies, such as saturation and nonlinear output [21].

Most of the current studies on outdoor optical channel properties are based on atmospheric turbulence theory, and they are commonly modeled using Log-normal distribution for weak and moderately turbulent channels [22]. From moderately to strongly turbulent states, the K distribution, Negative Exponential distribution, and Gamma distribution are used for modeling [23]. The Malaga atmospheric turbulence channel model describes a variety of general models of turbulence distributions, and setting the corresponding parameters for the Malaga distribution is extended to obtain the Log-normal model, the gamma-gamma model and the K-distribution model [24].

The visible light communication near the ground is significantly influenced by the underlying atmosphere and background light. Under this circumstance, the traditional Kolmogorov spectrum or Non-Kolmogorov spectrum is inapplicable to model the vehicular visible light communication system. In this section, a double Gaussian noise model in vehicular visible light communication is introduced through practical measurements on the background light noise of outdoor visible light communication in different cities at night, which provides a theoretical basis for system modeling as well as noise removal.

For the vehicular visible light communication system, we denote $s(t)$ the modulated light signal of LED, $h(t)$ the atmospheric transmission channel model, $n_o(t)$ the background noise, η is the photodetector conversion coefficient, $n_e(t)$ be the detector electrical noise, and $y(t)$ the output signal. The mathematical model of the system is expressed as Eq. (5.1), where $*$ denotes the convolution [25].

$$y(t) = [s(t) * h(t) + n_o(t)] \cdot \eta + n_e(t). \quad (5.1)$$

The thermal motion of electrons in the resistor and the random generation and compound of the carriers in the semiconductor cause a series of noises in the photodetector, such as shot noise, resistance thermal noise, and $1/f$ noise. The electrical noise $n_e(t)$ introduced by the photodetector is described as additive Gaussian noise with zero mean [24, 25]. The impact of the background noise $n_o(t)$ on the system is the photogenerated radiation current induced by various stray optical radiation in the background, i.e.,

$$V_{out}(t) = \left[\eta \cdot \int_s [E(t)]^2 ds \right]^2 R. \quad (5.2)$$

In Eq. (5.2), $E(t)$ is the cumulative optical field of the stray background light, s is the effective detector area, and R is the output impedance. $V_{out}(t)$ is the noise at the detector output. Since the output signal of the optical field square law detector has a mean value greater than 0, the Gaussian white noise with zero mean does not accurately characterize the channel in optical wireless communications.

In the context of vehicular visible light communication, Xizheng Ke proposes a noise model (Ke's Model) fitted by a double Gaussian function. Corresponding to $n_o(t)$ in Eq. (5.1), the probability distribution is

$$p(I) = A_1 \times \frac{1}{\sqrt{2\pi} \times \sigma_1} \times e^{\left[-\frac{(I-\mu_1)^2}{2 \times \sigma_1^2}\right]} + A_2 \times \frac{1}{\sqrt{2\pi} \times \sigma_2} \times e^{\left[-\frac{(I-\mu_2)^2}{2 \times \sigma_2^2}\right]}, \quad (5.3)$$

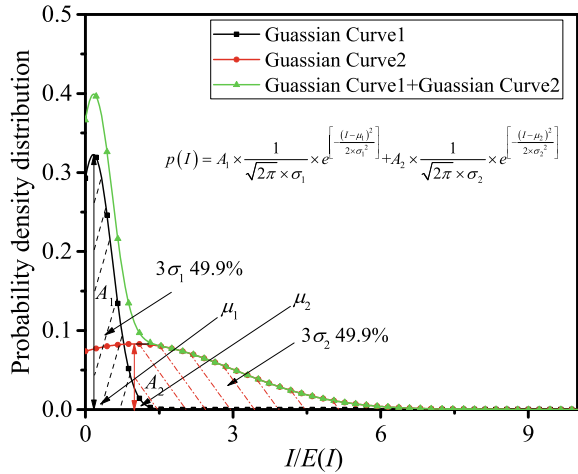
where μ_1 and σ_1 are the mean and variance values of light intensity flicker of the current street light in the vehicle's location, respectively. μ_2 and σ_2 are the mean and variance values of light intensity flicker of other light sources, respectively. A_1 and A_2 are normalized correction coefficients, ranging from 0 to 1, such that $\int_0^{+\infty} p(I) = 1$.

Figure 5.5 shows the double Gaussian distribution model (Ke's Model) for the vehicular visible light communication. Gaussian curve 1 indicates that the background noise mainly contains the background light component of the street light where the vehicle is currently located. Its light intensity is weak and less fluctuated with small variance, which causes a sharp peak. Gaussian curve 2 indicates that the outdoor visible light background noises are primarily induced by the head- and rear-lights, high brightness billboards, and traffic lights with high intensity and large variance.

As shown in Fig. 5.6, the light intensity acquisitions were carried out in different regions and under different weather conditions in Chinese cities, such as Xi'an, Beijing, Harbin, Changchun and Shijiazhuang, to verify the model compliance, where the detector model is OPHIR-PD300-UV with a sampling frequency of 15 Hz.

Figure 5.7 shows the background noise probability density distribution of visible light communication at night (a) in different cities, (b) in different locations in Xi'an, (c) in rainy days, and (d) in Qinling Terminal Mountain road tunnel. Table 5.1

Fig. 5.5 Dual Gaussian distribution model for vehicular visible light communication



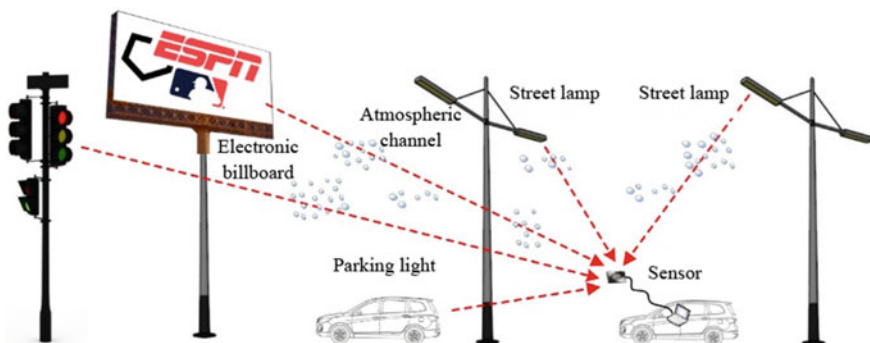


Fig. 5.6 Schematic diagram of visible light communication noise acquisition

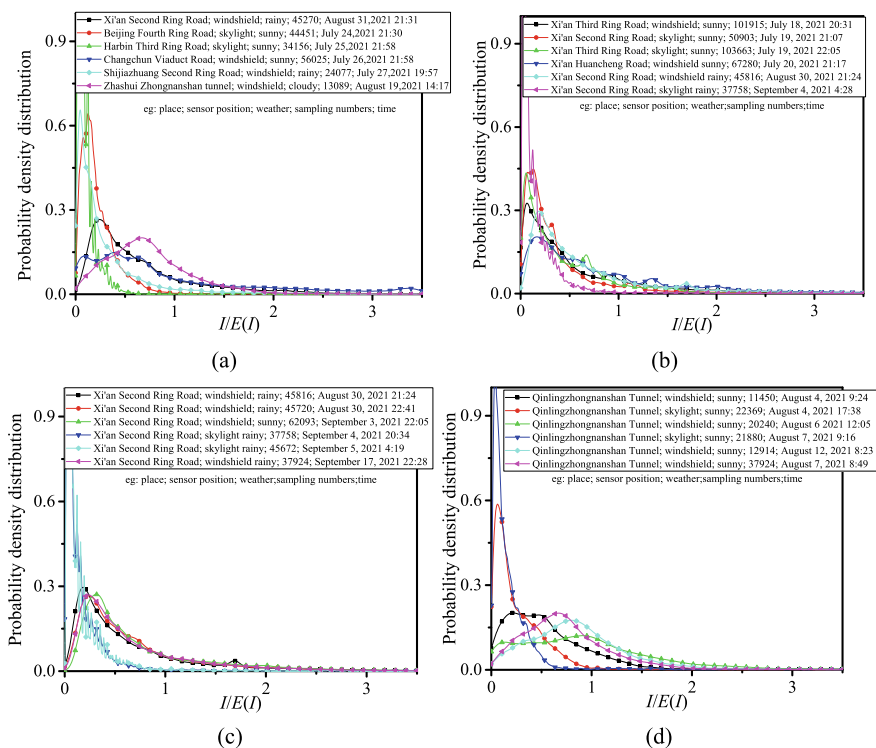


Fig. 5.7 Probability density distribution of background noise in vehicular visible light communication in **a** different cities, **b** different locations of Xi'an, **c** rainy days, and **d** tunnel

Table 5.1 Model parameters of background noise in vehicular visible light communication in different cities

Area/Weather	μ_1	σ_1	μ_2	σ_2	$P(\mu_1)$
Xi'an/Sunny	0.3	0.68	1.8	3.7	0.43
Xi'an/Rainy	0.24	0.49	1.7	3.7	0.43
Beijing/Sunny	0.7	0.68	1.9	3.9	0.31
Harbin/Sunny	0.4	0.5	1.8	3.7	0.42
Changchun/Sunny	0.8	0.69	2.4	4.8	0.15
Shijiazhuang/Rainy	0.3	0.68	1.8	3.7	0.65
Tunnel/Sunny	0.3	0.68	2.4	2.5	0.26

shows the parameters of the background noise model for vehicular visible light communication in different cities.

Several conclusions can be drawn as follows. (1) The background noise for vehicular visible light communication can be represented by a superposition of two Gaussian functions. (2) The probability density distribution in rainy days is similar as that in sunny days. (3) The light intensity fluctuation of the current streetlight is smaller than that generated by other light sources.

5.3 Multiplicative Noise Model

Multiplicative noise is also called convolutional noise. It refers to the process in which the signal's own amplitude changes randomly due to the physical properties of some media in the propagation process, and the information contained in the signal is lost. The characteristics of multiplicative noise indicate that it appears with the appearance of useful signals. When the signal disappears, the multiplicative noise will also disappear. Suppose that $s(t)$ is the modulated light signal of the light-emitting diode, $n_0(t)$ is the background light model, η is the photodetector conversion coefficient, $n_e(t)$ is the detector electrical noise model, and $y(t)$ is the output signal. The mathematical model of the system is shown in Eq. (5.4)

$$y(t) = [(s(t) + n_o(t)) * h(t)] \cdot \eta + n_e(t) = [(s(t) + n_o(t)) * f_\beta f_h f_A] \cdot \eta + n_e(t), \quad (5.4)$$

where $*$ is convolution operation, f_h is the light intensity attenuation term, f_h is the aiming error term, and f_A is the light intensity flicker term caused by atmospheric turbulence. The multiplicative noise model fitted with Gaussian function, corresponding to f_A in Eq. (5.4), is defined as

$$f_A = \frac{1}{\sqrt{2\pi} \times \sigma} \times e^{\left[-\frac{(I-\mu)^2}{2 \times \sigma^2}\right]}, \quad (5.5)$$

where I is the received light intensity, and μ and σ are the mean and variance of the multiplicative noise model, respectively. The Gaussian model of multiplicative noise for visible light communication is shown in Fig. 5.8.

Similarly, the impacts of multiplicative noise f_A on the system outage probability P_{out} and bit error rate P_{BER} are described as

$$P_{\text{out}} = F_A(\gamma_{th}) = \int_0^{\gamma_{th}} \frac{1}{\sqrt{2\pi} \times \sigma} \times e^{\left[-\frac{(x-\mu)^2}{2 \times \sigma^2}\right]} dx, \tag{5.6}$$

$$P_{\text{BER}} = \frac{q^p}{2\Gamma(p)} \int_0^\infty \exp(-qx) x^{p-1} F_A(x) dx. \tag{5.7}$$

The rainfall hinders the optical transmission of the transmitter and receiver of the outdoor visible light communication system, and causes a loss of optical carrier power in the transmission process. In meteorology, the rainfall intensity is usually divided by the amount of precipitation in a period of time. According to the standard specified by Chinese department of meteorological, the rainfall can be divided into six levels: light rain, moderate rain, heavy rain I, heavy rain II, heavy rain III, and extremely heavy rain, as shown in Table 5.2.

Fig. 5.8 Gaussian model of multiplicative noise for visible light communication

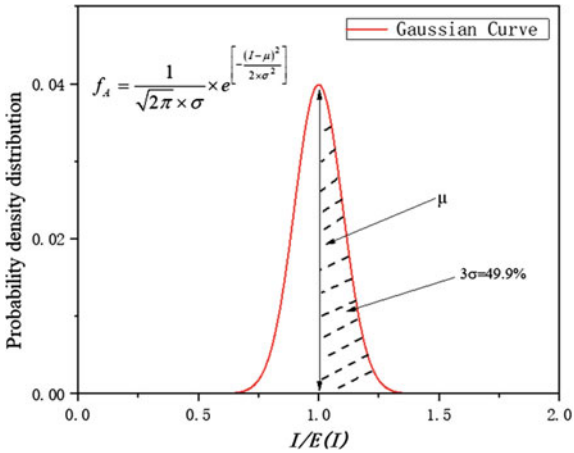


Table 5.2 Rainfall intensity standard by Chinese department of meteorological

Rainfall intensity level	Amount of precipitation in 12 h (mm)
Light rain	<5
Moderate rain	5–10
Heavy rain I	10–30
Heavy rain II	3–70
Heavy rain III	7–140
Extremely heavy rain	>140

According to the test, the attenuation coefficient of light beam transmission in rainy weather has a close relationship with rainfall intensity (i.e., the amount of precipitation). Since raindrops are spheres, and their molecular diameter is equivalent to the wavelength of visible light, so the attenuation of rain is suitable to be characterized by Mie theory. If raindrops conform to the Marshall-Palmer distribution, the rain attenuation prediction model widely used in existing optical wireless communication systems is an exponential function of the rainfall intensity. The channel attenuation coefficient γ_R in this model is

$$\gamma_R = \kappa(\kappa_h, \kappa_v, \tau_1, \tau_2) R^{\alpha(\kappa, \kappa_h, \kappa_v, \alpha_h, \alpha_v, \tau_1, \tau_2)}, \quad (5.8)$$

where R is the rainfall intensity, τ_1 and τ_2 are the path inclination angle and polarization angle of the rainy channel, respectively. κ and α are the regression coefficients of the rainy channel of the communication system.

Figure 5.9 shows the probability density distribution of Gaussian multiplicative noise in visible light communication. The Gaussian curve represents the multiplicative noise model under different rainfall intensities. The mean and variance of the Gaussian model are different under different rainfall intensities. It can be seen that, in the case of sunny, moderate rain and heavy rain, the mean and variance of the Gaussian function are similar. In light rain, the mean is slightly shifted to the left with a relatively large variance, which indicates significant changes on light intensity. That is, compared with other weather conditions, the light rain causes severer attenuation to the communication link and has a greater impact on changes in light intensity.

It can be seen from Fig. 5.9 and Table 5.3 that the Gaussian multiplicative noise model is quite consistent with the actual situation, and the following conclusions can be drawn: (1) The multiplicative noise model of visible light communication can be represented by a Gaussian function; (2) The probability density function distribution of rainy days is consistent with the actual situation. There is no obvious difference in the probability density distribution of sunny days; (3) The mean and variance of the probability density under light rain change slightly compared with other weather conditions, the mean shifts to the left, and the variance becomes larger.

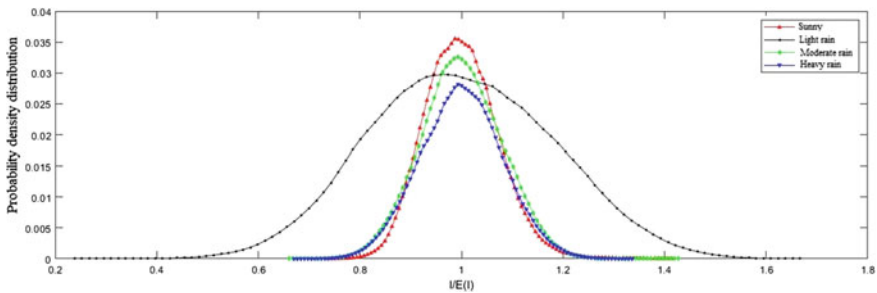


Fig. 5.9 Measurement of multiplicative noise under different rainfall intensities

Table 5.3 Multiplicative noise model parameters under different rainfall

Weather condition	μ	σ
Sunny	0.99	0.15
Light rain	0.97	0.19
Moderate rain	0.99	0.15
Heavy rain	0.99	0.15

5.4 Optimal Layout of Light Source

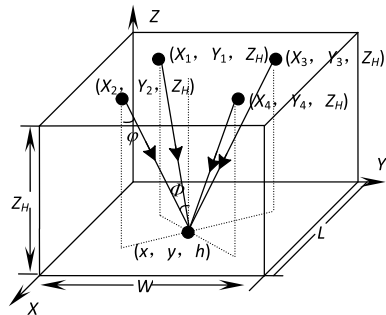
Since the sizes of rooms and indoor facilities are not the same, to optimize the communication effect in an indoor visible light communication system, it is necessary to minimize the change in the optical power distributed on the same level in a room. To achieve this goal, the layout of LED lights must be reasonably arranged according to different rooms [26].

Consider a room with a size of $L \times W \times Z_H$, suppose that the terminal equipment is placed on a plane with a height of h , and a coordinate system as shown in Fig. 5.10 is established. Suppose a total of four LED lights are symmetrically distributed on the ceiling. Analysis of the received power distribution on the plane $z = h$ is as follows. When the received power distribution changes the least, it can be concluded that the LED lights have the best layout. Suppose the received power of any point (x, y, h) on the plane $z = h$ is.

$$P_r = \sum_{i=1}^N P_{ti} H_i(0), \quad (5.9)$$

where N is the number of LED lights. Suppose the coordinates of the four LED lights are (X_1, Y_1, Z_H) , (X_2, Y_2, Z_H) , (X_3, Y_3, Z_H) , and (X_4, Y_4, Z_H) . Then,

$$H_i(0) = \frac{m+1}{2\pi} \cdot \frac{(Z_H - h)^2}{[(Z_H - h)^2 + (x - X_i)^2 + (y - Y_i)^2]^{\frac{m+3}{2}}}. \quad (5.10)$$

Fig. 5.10 Schematic diagram of a light source layout

Since the LED lights are symmetrically distributed, the coordinate relationships satisfy the following: $X_1 = X_3$, $X_2 = X_4$, $Y_1 = Y_2$, $Y_3 = Y_4$, $X_2 = L - X_1$, and $Y_3 = W - Y_1$.

$$\begin{aligned} f(u_i, v_i; x, y) &= H_i(0) \\ &= \frac{m+1}{2\pi} \cdot \frac{(Z_H - h)^2}{[(Z_H - h)^2 + (x - |u_i - X_1|)^2 + (y - |v_i - Y_1|)^2]^{(m+3)/2}}, \end{aligned} \quad (5.11)$$

where $u_i \in \{0, L\}$ and $v_i \in \{0, W\}$. Suppose the emission power of each LED lamp is the same, $P_{ti} = P_t$, then Eq. (5.11) becomes

$$\begin{aligned} P_r(x, y) &= P_t \sum_{i=1}^{N=4} f(u_i, v_i; x, y) \\ &= f(0, 0; x, y) + f(0, 5; x, y) + f(5, 0; x, y) + f(5, 5; x, y). \end{aligned} \quad (5.12)$$

The average power of each point on the indoor plane $z = h$ is

$$\bar{P}_r = \frac{1}{S} \iint_L P_r(x, y) dx dy = \frac{P_t}{S} \sum_{i=1}^{N=4} \iint_L f(u_i, v_i; x, y) dx dy, \quad (5.13)$$

where S is the area of the plane $z = h$ in the room, and L represents this area.

We use the variance D of the received power to represent the “average deviation” of the power of each point on the plane $z = h$. Then,

$$D = \frac{1}{S} \iint_L (P_r(x, y) - \bar{P}_r)^2 dx dy = \frac{1}{S} \sum_{i=1}^{N=4} \iint_L (f(u_i, v_i; x, y) - \bar{P}_r)^2 dx dy. \quad (5.14)$$

We calculate the partial derivatives of X_1 and Y_1 for Eq. (5.14). When $\frac{\partial D}{\partial X_1} = \frac{\partial D}{\partial Y_1} = 0$, we obtain the best X_1^* and Y_1^* , and the best LED layout can be determined.

5.5 Indoor Visible Light Channel

In an indoor optical wireless communication system, the link configuration of the transmitter and receiver can be of three types: direct link configuration, diffuse reflection link configuration, and semi-diffuse reflection link configuration [27, 28].

Direct link configuration means that there is a direct, unobstructed optical link between the transmitter and receiver, so the transmitter and receiver must be aligned with each other. In the direct link configuration system, a detector with a smaller field of view can be used, so it is less affected by the background light, and a higher communication rate and lower link loss can be obtained. The communication rate can reach several Gbit/s. A disadvantage is that the transmitter and receiver need to be aligned, and once the direct link is blocked, the communication will be interrupted; moreover, the flexibility of this configuration is poor.

In a system with a diffuse reflection link configuration, both the transmitter and receiver have a larger transmitting angle and receiving field of view angle, and their alignment is not required. The optical signal emitted by the transmitter reaches the receiver after multiple diffuse reflections from the reflector in the room. Communication will not be interrupted by an obstruction, and the flexibility is high. The disadvantage is that the link loss is serious and the multipath signal formed by diffuse reflection considerably reduces the bandwidth of the system; the communication rate is generally in tens of Mbit/s.

The semi-diffuse reflection link configuration inherits the advantages of both the direct and diffuse reflection links and effectively increases the bandwidth of the system without losing flexibility. In a system with a semi-diffuse reflection link configuration, the transmitter has a larger launch angle, and there are multiple direct and multiple diffuse reflection paths between the transmitter and receiver; this reduces the impact of shielding on communication. In addition, the direct path plays a leading role in the system. Compared with the diffuse reflection link configuration system, the system bandwidth is considerably improved, and the communication rate can reach several hundred Mbit/s.

In an indoor visible light communication system, the semi-diffuse reflection link configuration is adopted. Figure 5.11 shows a schematic diagram of the indoor visible

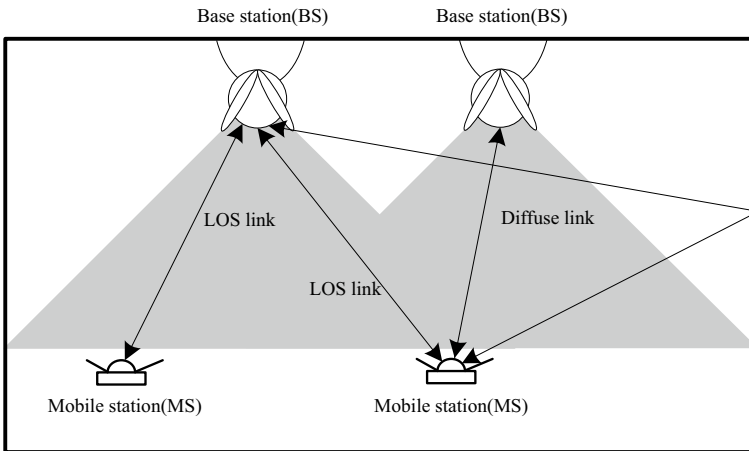


Fig. 5.11 Schematic diagram of indoor visible light communication link configuration

light communication link configuration. The transmitter (i.e., the modulated white LED array light source) is installed on the ceiling or wall with a large divergence angle. Multiple array light sources can cover all corners of the room with direct light signals, effectively reducing the impact of shadows and occlusion. The receiver (i.e., a photodetector located on a desktop or working platform) receives the optical signal from the transmitter. In the indoor visible light communication system, considering the needs of lighting, the direct link dominates the channel, but there is also a part of the diffuse reflection link.

In indoor visible light communication intensity modulation/direct detection (IM/DD) system, the data signal is modulated and loaded on the optical carrier of the LED, and the receiver, that is, the photodetector, outputs a current signal proportional to the power of the received optical signal. The input signal $x(t)$ of the channel is the instantaneous optical signal output by the white LED array lamp. The output signal $y(t)$ of the channel is the instantaneous current output by the photodetector. The magnitude of the current is proportional to the total power of the light signal received on the surface of the photodetector. There are multipath signals between the transmitter and receiver, resulting in frequency selective fading and spatial selective fading, so the amplitude of the received signal depends on the spatial position of the transmitting and receiving antennas. The channel characteristic can be expressed as a baseband linear system between the input signal $x(t)$, output signal $y(t)$, and channel impulse response $h(t)$. The background light, detector, and circuit noise can be considered additive white Gaussian noise $n(t)$ independent of $x(t)$, as shown in Fig. 5.12. The final baseband channel model can be expressed as

$$y(t) = \eta \frac{x(t)}{d^2} * h(t) + n(t), \quad (5.15)$$

where $*$ means convolution, η is the photodetector response sensitivity with a unit of $\text{A} \cdot \text{m}^2/\text{W}$, and d is the distance between the transmitter and receiver. Without loss of generality, the factor $1/d^2$ can be incorporated into $h(t)$ to obtain

$$y(t) = \eta x(t) * h(t) + n(t). \quad (5.16)$$

The indoor visible light communication system is a comprehensive system integrating lighting and communication. First, to achieve better lighting effects, multiple

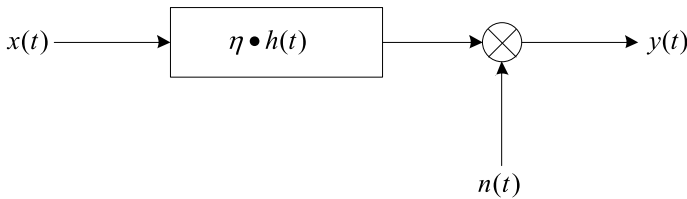


Fig. 5.12 Indoor visible light communication wireless optical channel model

white-light sources are usually installed indoors, and the layout of multiple light sources is closely related to the transmission model of the channel. It is necessary to study the multi-light source channel model and its relationship with the light source layout. Second, unlike the diffuse reflection configuration of infrared wireless communication, the line-of-sight link is dominant in the visible light communication system. At the same time, necessary measures must be taken to reduce or eliminate the impact of diffuse reflection light signals on the system bandwidth. $h(t)$ can generally be expressed as

$$h(t) = f(x) = \begin{cases} \frac{2t_0}{t^3 \sin^2(\text{FOV})}, & t_0 \leq \frac{t_0}{\cos(\text{FOV})} \\ 0, & \text{else} \end{cases}, \quad (5.17)$$

where t_0 represents the minimum delay time. Considering the safety of human eyes, the maximum power is expressed as

$$D_{\max} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) dt. \quad (5.18)$$

Unlike radio frequency communication, the signal-to-noise ratio of optical wireless communication is proportional to the square of the received signal.

$$\text{SNR} = \frac{R^2 H^2(0) P_r^2}{R_b N_0} \quad (5.19)$$

The channel DC gain is expressed as

$$H(0) = \int_{-\infty}^{\infty} h(t) dt. \quad (5.20)$$

The radiation intensity of the system can be expressed as

$$R_0(\phi) = \begin{cases} \frac{m_1+1}{2\pi} \cos^{m_1}(\phi), & -\frac{\pi}{2} \leq \phi \leq \frac{\pi}{2} \\ 0, & \phi \geq \frac{\pi}{2}, \end{cases} \quad (5.21)$$

where m_1 is the Lambertian order of beam directionality, and the radiation is strongest when $\phi = 0$. The Lambertian order has the following relationship with the LED half power angle:

$$m_1 = \frac{-\ln 2}{\ln(\cos \phi_{1/2})}. \quad (5.22)$$

The DC gain of the direct channel can be expressed as

$$H_{\text{los}} = \begin{cases} \frac{A_r(m_1+1)}{2\pi d^2} \cos^{m_1}(\phi) T_s(\psi) g(\psi) \cos \psi, & 0 \leq \psi \leq \psi_c \\ 0, & \text{else.} \end{cases} \quad (5.23)$$

5.6 Receiver and Detection Technology

In indoor visible light communication systems, a major factor affecting system performance is the interference of background light. Background light primarily comes from sunlight passing through windows and artificial light sources, such as incandescent and fluorescent lamps. Since the visible light communication system uses white LEDs for lighting as the communication light source, only the influence of sunlight should be considered. In the optical link, the receiver converts signal light and background light into electrical signals for judgment, and the background light in the conversion process will cause the photodetector to produce some shot noise, which affects the signal-to-noise ratio of the receiver. Another factor affecting system performance is ISI caused by multipath dispersion. The light signal emitted by the light source undergoes multiple reflections from ceilings, walls, furniture, etc., which will cause the signal reaching the receiver to expand to varying degrees in the time domain and cause ISI.

The influence of these two should be fully considered in the design of the receiver to minimize the disruption caused. The composition diagram of the optical receiver in the application of IM/DD system is shown in Fig. 5.13. The receiver consists of a receiver front end, linear filter, and decision recovery circuit. The front end of the receiver is composed of a photodetector and amplifier to realize the conversion and amplification of the photoelectric signals. The linear filter performs filtering processing on the received electrical signal to initially reduce the impact of ISI. The decision recovery circuit makes a decision on the sampled electrical signal to recover the sent data.

5.6.1 Receiver Front End

The front end of the receiver is composed of a photodetector and amplifying circuit, which can realize photoelectric conversion and amplification of photocurrent signals.

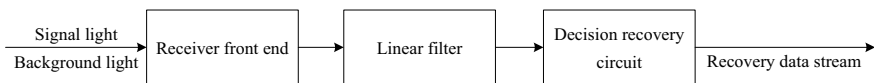


Fig. 5.13 Diagram of the receiver structure

Figure 5.14 shows a typical transconductance amplifier circuit of a positive-intrinsic-negative (PIN) photodetector. The signal light and background light are irradiated on the photodetector, and the generated photocurrent is output through the amplifying circuit and then sent to the judgment circuit for judgment.

Under the influence of photons, the photodetector absorbs the photons of the incident light to generate photocarriers, and electron–hole pairs are generated inside the detector. Under the influence of the electric field, the electron–hole pair forms a photocurrent proportional to the incident light power. For an ideal PIN photodetector, the photoelectric conversion process is a Poisson process [29–31]. The average number of photons output during time T with and without optical signal can be expressed as [32]

$$\mu_{\text{on}} = \frac{\xi P_s T}{hf} h(t - T) + \frac{\xi P_b T}{hf}, \quad (5.24)$$

$$\mu_{\text{on}} = \frac{\xi P_b T}{hf}, \quad (5.25)$$

where ξ is the quantum efficiency of the photo detector, P_s is the optical power of the received signal, P_b is the optical power of the background light, h is the Planck constant, and f is the frequency of the optical carrier.

In an indoor visible light communication system, the background light is primarily the sunlight transmitted through windows. In particular, when the receiver is close to a window, the influence is very large. Therefore, the extent of the influence of the background light on the system is related to the orientation of the windows, the time of day, and the location of the receiver in the room. For a window facing west, the average irradiance of sunlight is 57 W/m^2 [33]. In other optical wireless communication systems, optical filters are usually used to reduce the influence of the background light. However, in the visible light communication system, the overlap area of the wavelengths of the source white LED and the sunlight is considerable, so the effect of the optical filter in the visible light communication system is not notable.

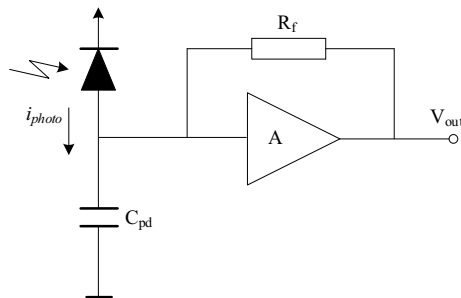


Fig. 5.14 Positive-intrinsic-negative (PIN) photodetector transconductance amplifier circuit

In the receiving circuit, thermal and shot noise from the photodetector and amplifier circuit affect the output of the receiver front end. Figure 5.12 shows the equivalent circuit diagram of the receiver front end. In Fig. 5.15, I is the average photocurrent, which is equal to the number of photoelectrons multiplied by the electron charge q [34]:

$$I_{\text{on}} = \left[\frac{\xi P_s T}{hf} h(t - T) + \frac{\xi P_b T}{hf} \right] q, \quad (5.26)$$

$$I_{\text{off}} = \left(\frac{\xi P_b T}{hf} \right) q, \quad (5.27)$$

where $i_s(t)$ and $i_t(t)$ represent the shot and thermal noise in the receiving circuit, respectively. The circuit thermal noise is primarily caused by the electronic components of the receiver and is inevitable in all communication systems. It is primarily caused by the drift of the impedance components and active devices in the amplifier circuit. The shot noise caused by the photocurrent of the photodetector is also the main noise source of the wireless optical link. Shot noise is primarily caused by the discontinuity of the energy and charge of the photodiode. Generally, thermal noise and shot noise can be considered additive white Gaussian noise [35].

The voltage across the transimpedance is equal to the product of impedance and current. Since shot noise and thermal noise are zero mean, the mean value of the integrator output is [36]

$$\mu_{\text{on}} = E[Y_{\text{on}}] = \left[\frac{\xi P_s T}{hf} h(t - T) + \frac{\xi P_b T}{hf} \right] R_f q, \quad (5.28)$$

$$\mu_{\text{off}} = E[Y_{\text{off}}] = \left(\frac{\xi P_b T}{hf} \right) R_f q. \quad (5.29)$$

The shot noise is determined by the current passing through the photodetector. Therefore, after passing through the integrator, the output variance caused by the shot noise is [37]

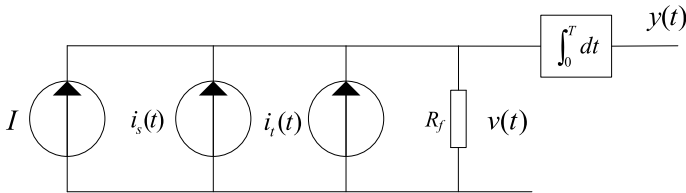


Fig. 5.15 Front-end equivalent circuit in receiver

$$\text{Var}[Y_{\text{on}}]_s = \left[\frac{\xi P_s T}{hf} h(t - T) + \frac{\xi P_b T}{hf} \right] R_f^2 q^2 + I_{\text{bg}} R_f^2 q, \quad (5.30)$$

$$\text{Var}[Y_{\text{off}}]_s = \left(\frac{\xi P_b T}{hf} \right) R_f^2 q^2 + I_{\text{bg}} R_f^2 q, \quad (5.31)$$

where I_{bg} is the dark current of the photodetector.

The output variance caused by thermal noise is [38]

$$\text{Var}[Y]_t = \frac{2kT_K I_{\text{nbf}}}{T R_f}, \quad (5.32)$$

where k is Boltzmann's constant, T_K is the thermodynamic temperature, and I_{nbf} is the noise bandwidth factor.

Since shot noise and thermal noise are independent of each other and can be linearly superimposed, the variance of the total output of the integrator is [39]

$$\sigma_{\text{on}}^2 = \text{Var}[Y_{\text{on}}] = \left[\frac{\xi P_s T}{hf} h(t - T) + \frac{\xi P_b T}{hf} \right] R_f^2 q^2 + I_{\text{bg}} R_f^2 q + \frac{2kT_K I_{\text{nbf}}}{T R_f}, \quad (5.33)$$

$$\sigma_{\text{off}}^2 = \text{Var}[Y_{\text{off}}] = \left(\frac{\xi P_b T}{hf} \right) R_f^2 q^2 + I_{\text{bg}} R_f^2 q + \frac{2kT_K I_{\text{nbf}}}{T R_f}. \quad (5.34)$$

In the visible light communication system using IM/DD technology, the white LED emits light to indicate that there is a signal and does not emit light to indicate that there is no signal. The receiver photodetector converts the received optical signal into an electrical signal and sends it to a demodulator for demodulation. The received signal-to-noise ratio describes the ratio of the power of the useful signal to the noise signal and determines the performance of the system.

5.6.2 Receiving Array Design

The receiving array adopts a hexagonal planar array design, as shown in Fig. 5.16. The side length of the hexagon is d , the photodetector is located at the center of each hexagon, and the distance between adjacent photodetectors, which are numbered in sequence, is $\sqrt{3}d$. To increase the diversity gain as much as possible so that the channels obtained by each detector are independent of each other, the distance between the detectors should meet certain conditions. Moreover, the selection of the detector spacing is related to the field of view (FOV) of the detector, as shown in Fig. 5.17.

Fig. 5.16 Schematic diagram of a receiving array design

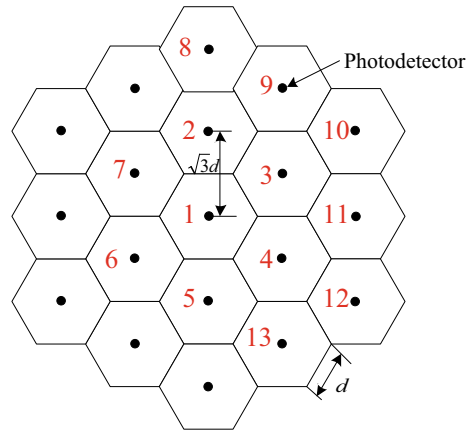
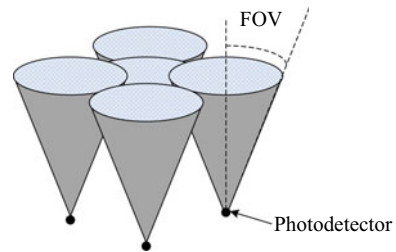


Fig. 5.17 Photodetector coverage area



5.7 Uplink of Visible Light Communication

The lack of uplink is the biggest shortcoming in visible light communication. Generally, there are the following schemes for choosing an uplink.

5.7.1 Radio Frequency Uplink

At present, the radio frequency communication technology is quite mature, and the IEEE 802.11 protocol suite, that is, Wi-Fi, is the de facto standard widely used in indoor wireless communication. The use of visible light broadcast channel as the downlink and Wi-Fi technology as the uplink [40] not only takes advantage of the characteristics of high-speed data transmission of visible light but also uses Wi-Fi to cover a large range and is not restricted by line-of-sight transmission. This eliminates the shortcomings of visible light blocking, which causes signal interruption.

However, the integration of radio frequency technology into visible light communication loses the advantages of visible light communication: high confidentiality and immunity to electromagnetic interference.

5.7.2 Infrared Uplink

A typical visible light communication scheme that uses infrared light for uplink data transmission [41] uses the blue light in the RGB three-color LED as the signal source of the system downlink and the infrared LED with a central wavelength of 850 nm as the uplink. The uplink and downlink are modulated by discrete multitone mode to form a wavelength division duplex system. The receiver uses an avalanche photodiode detector with 473 nm (bandpass 10 nm) and 805 nm long-pass filters placed on the front end. They are used to eliminate the interference between the uplink and downlink and can also suppress background noise.

Because of the low cost and simple structure of the transmitter drive circuit, infrared LEDs have long been used for communication. Infrared Data Association (IrDA) is a worldwide organization that formulates and improves the infrared communication standards.

However, the infrared band of 780–950 nm is close to the red band of visible light, and the human eye is very sensitive to it. Therefore, due to safety considerations, the transmit power is often not very high. This limits the coverage of the uplink and communication quality of the transmission link.

5.7.3 Laser Uplink

A typical visible light communication scheme that uses lasers for uplink data transmission [42] uses LED as the signal source of the system downlink and 1550 nm laser beam as the signal carrier of the data uplink. The two work together to form one full-duplex wavelength-division multiplexing (WDM) communication system. The receiving end also uses a narrow-band filter on the front of an avalanche photodiode detector to filter out background noise and improve the signal-to-noise ratio.

Although the 1550 nm wavelength is beyond the visible range of human vision, the corresponding safety standards must be followed when transmitting wirelessly, which further limits the transmission distance and link quality.

5.7.4 Visible Light Uplink

There are two schemes for using visible light for uplink and downlink data transmission. The first one uses time division duplex mode and through accurate timing determines the respective working hours of the uplink and downlink for information transmission. For example, forty LEDs are used to form an array as the signal emission source of the downlink, and one LED is used for the uplink transmission [43]. Another way of using wavelength division duplexing is to use different visible light bands to construct uplink and downlink. For example, in the RGB three-color LED,

the red (R) and green (G) channels are respectively modulated by 32QAM-OFDM for the downlink transmission of the system, and the blue (B) channel is simultaneously modulated with 32QAM-OFDM signal [43] for use as the system uplink.

Visible light communication primarily considers communication while illumination. When using visible light for uplink communication, both solutions emit a certain intensity of light beams in space, and these light beams are not necessary for illumination. They will cause a certain visual disturbance to people, which considerably limits its use scenarios.

5.7.5 Isomorphic Uplink

As shown in Fig. 5.18, the indoor visible light single-source full-duplex communication system consists of two parts: the active end and the reverse end. The downlink is the visible light communication link from the active to reverse end, and the uplink is the reply communication link from the reverse to active end. In this system, the active end is typically fixed to the ceiling of a room, and is transmitting information while illuminating the room. The reverse end, located on various information terminals, receives information from the downlink and sends data back to the active end. The light source for modulation for the entire system is provided by the active end. Relying on the downlink modulated optical signal as the carrier, the uplink modulates the downlink optical signal for the second time using an inverse modulator. The uplink data is modulated onto the reflected optical signal and reflected back to the active end. At the active end, the modulated optical signal is processed to extract the uplink signal. Then, a full-duplex communication with unidirectional light source is realized.

5.7.5.1 Active End Structure

The active end is composed of an Ethernet card, FPGA main controller, analog to digital converter (ADC) circuit, digital to analog converter circuit (DAC), LED drive circuit, LED drive circuit, photoelectric receiving circuit and signal processing circuit. Here, the FPGA

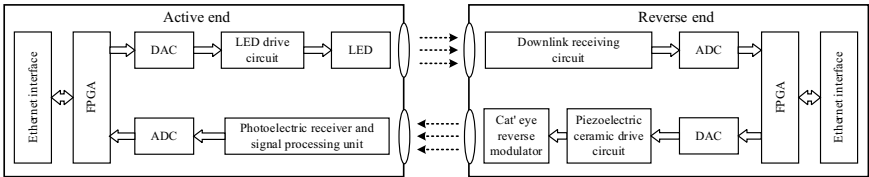


Fig. 5.18 Full-duplex communication system with a single light source

main controller implements the Ethernet card driver, downstream signal modulation, upstream signal demodulation, and ADC/DAC chip driver. In the active end, the downlink source data is input to the FPGA main controller through the Ethernet card for modulation. Then, the modulated digital signal is sent to the DAC circuit to be converted to an analog modulated signal, and sent to the LED drive circuit to control the LED to generate downlink modulated optical signal. The uplink secondary modulated optical signal is received by the photoelectric receiving module and converted into an electrical signal. The signal processing module is used to process the secondary modulated signal to extract the uplink signal. Then, the uplink signal is converted to a digital signal by the DAC circuit and demodulated by the FPGA. Finally, the demodulated uplink data is sent to the uplink sink via the Ethernet card.

5.7.5.2 Uplink System

The uplink is accomplished by the reverse end, which includes a defocusing cat's eye reverse modulator and its drive circuit, photoelectric receiving circuit, ADC/DAC circuit, FPGA main controller and Ethernet card. Here, the FPGA main controller implements the Ethernet card driver, downlink signal demodulation, uplink signal modulation, and ADC/DAC chip driver. The data transmission process at the reverse end is described as follows. The downlink optical signal is converted to an electrical signal by the photoelectric receiving circuit and then sent to the ADC circuit for sampling. The quantized digital signal is read into FPGA for demodulation. After that, the demodulated data is sent to the downlink sink through the Ethernet card. The uplink data source is input to the FPGA through the Ethernet card. Then, the modulated signal from FPGA is converted into an analog modulated signal by DAC. After that, the modulated signal is input into the inverse modulator drive circuit to perform a modulation for the second time on the downlink optical signal. Finally the secondary modulated optical signal is reflect back to the active end.

5.8 Visible Light Communication Positioning

With the rise and development of mobile Internet and Internet-of-Things technology, location-based services have shown huge commercial and military application prospects. As the premise and key to the realization of location-based services, location technology has attracted more attention. In fact, the satellite positioning technology represented by the Global Positioning System (GPS) can provide high-precision positioning services outdoors through the built-in GPS signal receiving module in the mobile terminal, and the civilian-level positioning accuracy is less than 10 m. However, owing to factors such as building occlusion, electromagnetic

shielding, and multipath fading, traditional satellite positioning systems completely “fail” in large shopping malls, hospitals, indoor locations, tunnels, underground parking lots, and other specific places. This has led to the use of indoor wireless positioning technologies, such as infrared and ultrasonic waves, Bluetooth, ultra-wideband, radio frequency identification, Wi-Fi, and visible light communication, have developed rapidly.

5.8.1 Received Optical Signal Strength Positioning

In a visible light communication system, a photodetector is used in the receiving device to detect the intensity of the visible light signal. Naturally, the received signal strength indicator (RSSI) method in traditional radio positioning can be used for visible light communication positioning.

In a visible light communication positioning system that only considers the downlink, the signal propagation path includes a direct path and reflection path. Since the direct visible light reaches the optical receiver, the power accounts for 95% of the total power [44], so the channel parameter measurement of the direct path is primarily studied. The typical downlink channel model is the Lambertian model. Based on this, the distance between each point light source and the photodetector can be calculated, and then the basic trilateral positioning method can be used to locate the receiver. With indoor multi-light sources, there are often more than three point light source signals received by the receiving end. Based on this, an overdetermined equation set is established, and the maximum likelihood estimation method is used to estimate the position of the receiver to further improve the positioning accuracy [44].

5.8.2 Fingerprint Identification Positioning

The fingerprint recognition positioning method [45] is consistent with the principle of the non-parametric method in traditional received signal strength indicator (RSSI) positioning. First, through offline surveys, the received signal strength (RSS), time of arrival (TOA), or time difference of arrival (TDOA) associated with the location of the mobile device is obtained to establish a fingerprint database. Then through online positioning, the real-time measured user signal is matched with the characteristic information in the database to obtain the location information of the target.

5.8.3 LED Identity Positioning

The LED identity positioning method [46] loads the position-related ID data on the corresponding LED light modulator. The LED sends out a light signal carrying the ID data controlled by the modulator. The receiving terminal analyzes the detected LED light signal and obtains corresponding location information from the acquired ID data to achieve passive positioning. The theoretical positioning accuracy of this method is half of the distance between adjacent LED signal sources.

5.8.4 Visible Light Imaging Positioning

Visible light imaging positioning technology [47] has received considerable research attention in recent years. This method uses the LED lighting array as the visible light communication transmitting part. The three-dimensional coordinates emitted from at least four LEDs in the array are received through two optical lenses. Then, the information is demodulated by two image sensors, and the LEDs received in the image sensor are used with the distance geometric relationship of the image to calculate the position of the target. Using this method in a $1.8 \times 1.8 \times 3.5$ m test space, using two six-million-pixel image sensors, the positioning error can be less than 0.15 m.

5.9 Summary and Prospects

It is generally believed that visible light communication is a short-range communication, which can be applied to indoor environments (e.g., hospitals and cabins), as well as vehicle-to-road and vehicle-to-vehicle communication. Visible light communication is used for short-distance access, and long-distance transmission relies on other communication means. There are also reports on the application of visible light communication to satellite-to-ground communication. Visible light communication needs to overcome two technical bottlenecks in the future. (1) Visible light communication is currently suitable for one-way broadcast communication; thus, research on the uplink equivalent to visible light is the direction of future efforts. (2) Background light impacts visible light communication; thus, background light suppression technology must be developed.

5.10 Questions

- 5.1 Briefly describe the composition of the visible light communication system.
- 5.2 Briefly describe the three channels of indoor visible light communication.
- 5.3 What are the main applications of visible light communication?

5.11 Exercises

- 5.1 In indoor visible light communication, the multipath effect will inevitably exist, and the multipath effect is mainly caused by the following two factors, one is the distance difference between the light source and the receiver path, and the other is caused by multiple reflections of light through the indoor wall. As shown in Fig. 5.19, consider the pulse delay width caused when two LEDs are not equal to the receiver.
- 5.2 The visible light communication system is shown in Fig. 5.20. Find the channel impulse response of the system.
- 5.3 As shown in Fig. 5.21, the distance between the LED and the receiving point is d , and ψ is the field of view of the receiver. Find the optical power received by the receiving point.
- 5.4 For a visible light system, considering non-direct viewing reception, only one reflection is considered, and the light receiving power of one reflection at a certain point in the room is calculated.
- 5.5 Calculate the channel impulse response of a light source, and give the least square estimation and minimum mean square error estimation of the channel.
- 5.6 Try to analyze the signal-to-noise ratio distribution of signals received by the visible light communication system at different locations as shown in Fig. 5.22.

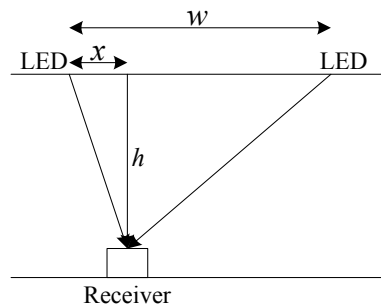


Fig. 5.19 Model of indoor visible light communication

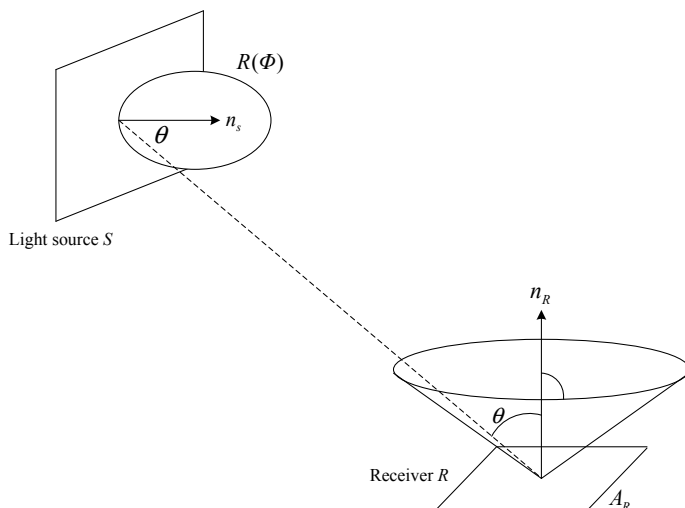
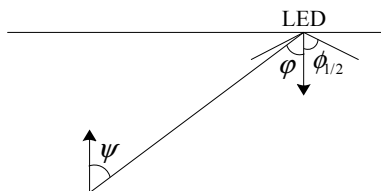
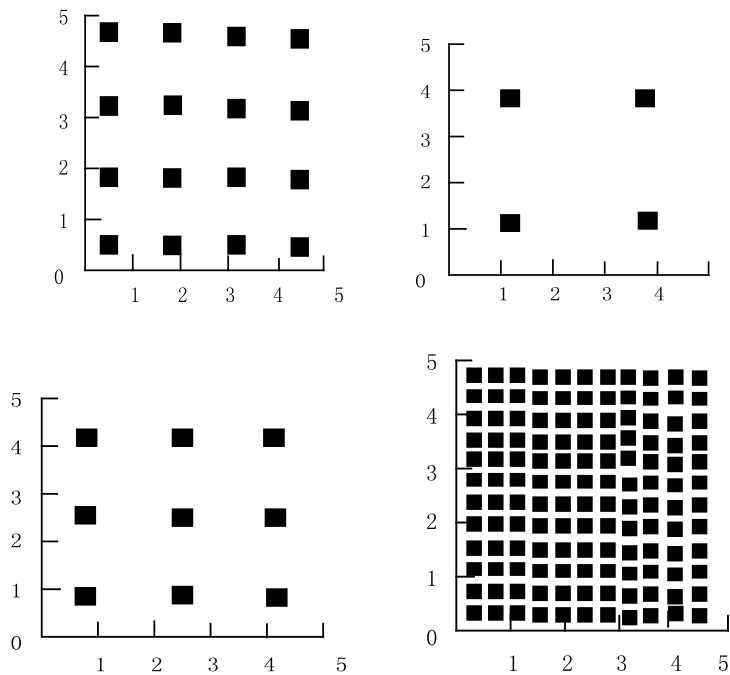


Fig. 5.20 Visible light communication system

Fig. 5.21 Model of visible light communication system



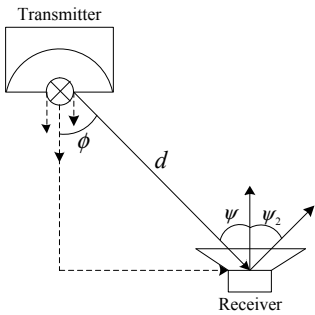
- 5.7 Try to find an optimization problem description that satisfies a certain signal-to-noise ratio under the conditions of Exercise 5.6.
- 5.8 Try to analyze the signal-to-noise ratio of the visible light MIMO system.
- 5.9 Try to find the signal-to-noise ratio of the visible light communication system as shown in Fig. 5.23.



Parameter	Value
Receiver field of view	50°
Transmit half angle	60°
PD effective detection area	1 cm ²
Transmit power	1 W
Wall reflection coefficient	0.8

Fig. 5.22 Visible light communication system at different locations and the parameters

Fig. 5.23 Model of visible light communication system



References

1. Nakamura S (1997) Present performance of InGaN based blue/green/yellow LED. Proceedings of SPIE Conf on light emitting diodes: research, manufacturing, and applications 26–35
2. Liu J, Li P, Hao Y (2005) Research and development of high-brightness GaN-based blue and white LEDs. *Chin J Quant Electron* 22(5):673–679
3. European Commission. Phasing out incandescent bulbs in the EU[EB/OL]. <http://ec.europa.eu>
4. Lee K, Park H (2011) Channel model and modulation schemes for visible light communications. IEEE 54th international midwest symposium on circuits and systems, pp. 1–4
5. Komine T, Haruyama S (2003) Nakagawa M. Bidirectional visible-light communication using corner cube modulator. Proceedings of wireless and optical communication
6. Hou J, O'Brien DC (2006) Vertical handover-decision-making algorithm using fuzzy logic for the integrated radio-and-OW system. *IEEE Trans Wirel Commun* 5(1):176–185
7. Lopez HF, Poves E, Perez JR et al (2006) Low cost diffuse wireless optical communication system based on white LED. Proceedings of 2006 IEEE tenth international symposium on consumer electronics St
8. Langer KD (2008) Optical wireless communications for broadband access in home area networks. Proceedings of the 10th anniversary international conference on transparent optical networks (ICTON), vol 4, pp 149–154
9. Kavehrad M (2007) Broadband and room service by light. *Scient Amer*, 82–87
10. Schubert EF, Kim JK, Hong L et al (2006) Solid-state lighting a benevolent technology. *Rep Prog Phys* 69:3069–3099
11. Chu M, Wu Q, Wang J et al (2009) Calculation of white LED's limit lumen efficiency. *Chin J Luminescence* 30(1):77–80
12. Wu H, Wang C, He S (2008) Study on the color rendering of white LED coated with red and green phosphors. *Acta Optica Sinica* 28(9):1777–1782
13. Sun CC, Lee TX, Tsung X et al (2006) Precise optical modeling for LED lighting verified by cross correlation in the midfield region. *Opt Lett* 31(14):2193–2195
14. Ke XZ, Wu JL, Yang SJ (2021) Research progress and prospect of atmospheric turbulence for wireless optical communication. *Chin J Radio Sci* 36(3):323–339
15. Zhai C (2021) Turbulence spectrum model and fiber-coupling efficiency in the anisotropic non-Kolmogorov satellite-to-ground downlink. *Results Phys* 29:104685
16. Ding DQ, Ke XZ (2006) Visible light communication and research on its key techniques. *Semiconductor Optoelectron* 27(02):114–117
17. Xu SW, Wu Y, Wang XF (2020) Visible light positioning algorithm based on sparsity adaptive and location fingerprinting. *Acta Optica Sinica* 40(18):1806003
18. Zhao L, Peng K (2017) Optimization of light source layout in indoor visible light communication based on white light-emitting diode. *Acta Optica Sinica* 37(7):0706001
19. Ding DQ, Ke XZ (2010) Research on generalized mathematic radiation model for White LED. *Acta Optica Sinica* 30(9):2536–2540
20. Yin P, Xu XP, Jiang ZG et al (2018) Design and performance analysis of planar concentrators as optical antennas in visible light communication. *Acta Optica Sinica* 38(4):0406004
21. Zhao JQ, Xu YF, Li JH et al (2016) Turbulence channel modeling of visible light communication under strong background noise and diversity receiving technologies. *Acta Optica Sinica* 36(3):0301001
22. Vetelino FS, Young C, Andrews L et al (2007) Aperture averaging effects on the probability density of irradiance fluctuations in moderate to strong turbulence. *Appl Opt* 46(11):2099–2108
23. Vetelino FS, Young C, Andrews L (2007) Fade statistics and aperture averaging for Gaussian beam waves in moderate-to-strong turbulence. *Appl Opt* 46(18):3780–3789
24. Ansari IS, Yilmaz F, Alouini MS (2015) Performance analysis of free-space optical links over Malaga (M) turbulence channels with pointing errors. *IEEE Trans Wirel Commun* 15(1):91–102

25. Wu ML, Ma FK, Liu WK (2020) Noise suppression method in medium and long distance outdoor visible light communication. *Laser Optoelectron Progress* 57(13):130601
26. Ding D, Ke X (2007) Design and simulation of light source layout of VLC system. *Opto-Electron Eng* 34(1):131–134
27. Iniguez RR, Idrus SM, Sun Z (2008) Optical wireless communications IR for wireless connectivity. Auerbach Publications, BocaRaton
28. O'Brien DC, Leminh H (2008) Home access networks using optical wireless transmission. *Proceedings of IEEE PIMRC*, pp 1–5
29. Jungnickel V, Pohl V, Nönnig S et al (2002) A physical model of the wireless infrared communication channel. *IEEE J Sel Areas Commun* 20:631–640
30. Rodríguez S, Pérez-Jiménez R, López-Hernández FJ et al (2002) Reflection model for calculation of the impulse response on IR-wireless indoor channels using ray-tracing algorithm. *Microw Opt Technol Lett* 32:296–300
31. Rodríguez S, Pérez-Jiménez R, González O et al (2003) Concentrator and lens models for calculating the impulse response on IR-wireless indoor channels using a ray-tracing algorithm. *Microw Opt Technol Lett* 36:262–267
32. Gagliardi RM, Karp S (1995) Optical communications. John Wiley & Sons, New York
33. Moreira AJC, Valadas RT, Oliveira-Duarte AM (1997) Optical interference produced by artificial light. *Wireless Netw* 3:131–140
34. Rahaim MB, Vegni AM, Little TDC (2011) A hybrid radio frequency and broadcast visible light communication system. *Proceedings of the GLOBECOM workshops*, pp 792–796
35. Cossu G, Corsini R, Khalid A M et al (2014) Bi-directional 400 Mbit/s LED-based optical wireless communication for non-directed line of sight transmission. *Proceedings of the optical fiber communication conference*
36. Zheng Z, Liu L, Hu WW et al (2014) Analysis of uplink schemes for visible-light communication. *ZTE Technol J* 6:8–11
37. Liu YF, Yeh CH, Chow CW et al (2012) Demonstration of bi-directional LED visible light communication using TDD traffic with mitigation of reflection interference. *Opt Express* 20(21):23019–23024
38. Chi N, Wang YQ, Wang YG et al (2014) Ultrahigh-speed single red-green-blue light emitting diode-based visible light communication system utilizing advanced modulation formats. *Chinese Opt Lett* 12(1):10605
39. Komine T, Nakagawa M (2004) Fundamental analysis for visible light communication system using LED lights. *Cons Electron IEEE Trans* 50(1):100–107
40. Zhang W, Kavehrad M (2012) Comparison of VLC-based indoor positioning techniques. *Opt Express* 20(21):23019–23024
41. Xu Y, Huang X, Li R et al (2014) Research on indoor positioning technology based on LED visible light communication. *China Light Lighting* 4:11–15
42. Yang A, Wu Y, Wang Y et al (2014) An indoor positioning method based on visible light tags: China, CN103823204A[P]
43. Dambul KD, O'brien D, Faulkner G (2011) Indoor optical wireless MIMO system with an imaging receiver. *IEEE Photon Technol Lett* 23(2):97–99
44. Ding Y, Xu N, Tu X et al (2014) Experimental research on optical power distribution of indoor visible light communication. *Chin J Quant Electron* 3:379–384
45. Fang R, Xu B (2014) Research on adaptive equalization technology of indoor visible light communication system. *Data Commun* 2(34–36):39
46. Wen X, Wang J, Xu Z et al (2013) Analysis and simulation of indoor visible light power distribution. *J Military Commun Technol* 1:73–76
47. Huang L, Feng G, Li H et al (2014) Research on channel estimation of indoor visible light communication. *Optoelectron Technol* 34(4):255–259

Chapter 6

Underwater Laser Communication



Sea water is a good conductor. Therefore, the skin effect of an electromagnetic wave (i.e., a shear wave), affects its transmission in seawater so that radio waves can hardly be transmitted underwater. The penetration depth of an electromagnetic wave is directly related to its wavelength: the penetration depth of a short wave is small, and the penetration depth of a long wave is larger. Even in an ultra-long wave communication system, the penetration depth of seawater is extremely limited (the maximum depth is only 80 m). Thus, laser provides a new way for underwater communication.

6.1 Overview of Underwater Laser Communication

Sea water is not only a natural barrier to conceal underwater navigation but also an obstacle to the propagation of radio waves. The early communication means of submarines was via long waves. Long wave communication usually requires laying large-scale long wave transmitting antenna on the land, and the submarine receives the information from the land command in a safe depth underwater. Although the ability of a long wave to penetrate seawater is strong, the ground-based antenna must be huge, and the communication rate is low.

It has been found that the attenuation of blue-green light in the water is notably less than that of other wavelengths. Thus, the submarine can use blue-green light communication in cruise depth or deeper sea water, which not only ensures the concealment of underwater navigation but also does not affect its normal activities. Furthermore, the laser communication has the advantages of high communication rate, good directivity, and strong anti-interference and anti-interception abilities [1].

In 1963, Dimtley et al. found that there is a blue-green light transmission window at 450–550 nm in seawater and that its attenuation is much smaller than that of other light bands.

Since 1980, the US Navy has performed six large-scale blue-green light communication experiments at sea. It has been shown that the blue-green laser communication can work normally under severe conditions, such as heavy rain and sea water turbidity. At the end of 1983, the former Soviet Union also conducted a laser communication experiment near Sevastopol, the main base of the Black Sea fleet, sending blue laser beams to space orbit mirrors and then transmitting them to underwater ballistic submarines [2].

The blue (460 nm), cyan (490 nm), and green (520 nm) of LuxeonIII LED were selected by Australian National University. The SLD-70BG2A photodiode with high sensitivity to blue and green light was used in the receiver circuit. The communication speed reached 57.6 kbit/s [2].

The Woods Hole Institute of Oceanography of the United States has developed a low-power deep-sea underwater optical communication prototype based on a LED. The communication rate of 10 Mbps was realized by using keying modulation technology (e.g. OOK).

The research group of Keio University of Japan conducted research on underwater optical wireless communication based on visible light LED. Their simulation results show that the transmission characteristics of underwater optical channel are related to wavelength and seawater turbidity [3].

A research group from the US naval aviation system command discussed the influence of seawater scattering on the communication rate of 10–100 Mbps in PSK modulated underwater optical wireless communication. The experimental results show that the sea water turbidity has an important impact on the channel modulation bandwidth and phase [3].

In 1983, the United States developed three submarine laser communication systems: space-based laser-to-submarine, satellite laser, and satellite mirror. In the satellite borne laser scheme, the laser transmitter is placed on a satellite. In the satellite mirror scheme, the laser is placed on the ground, and the optical signal is reflected through a mirror on a satellite. In 1986, a P-3C aircraft equipped with blue-green laser successfully sent signals to a submarine under ice by using the blue-green laser communication technology. In 1988, the United States completed the conceptual verification of the blue light communication system. In 1989, the United States began to study and improve the performance of laser communication between aircraft or satellite platforms and underwater submarines [4].

From 1989 to 1992, the US also implemented a submarine laser communication satellite program, aiming to realize satellite-to-submarine laser communication in geosynchronous orbit [4].

The influence mechanism of seawater scattering on coded modulation technology in real environment is still not comprehensively understood.

6.2 Underwater Laser Communication System

The underwater laser communication system comprises three parts: transmitting system, underwater channel, and receiving system.

6.2.1 Principle of Underwater Laser Communication

The mechanism of underwater optical wireless communication is as follows. The information to be transmitted is loaded onto the modulator after being encoded by the encoder and transformed into current changing with the signal to drive the light source (converting the electrical signal into optical signal). Then, the beam is transmitted in the channel as a parallel beam through the lens; the parallel beam transmitted by the lens is merged in the form of a point light source at the receiving end for light detection. On the device, the optical signal is converted into electrical signal by the optical detection device, and then the signal is adjusted. Finally, the original information is demodulated by the decoder. An underwater laser system is shown in Fig. 6.1.

6.2.2 Underwater Channel

Seawater contains dissolved substances, suspended solids, and various active organisms. Owing to the inhomogeneity of the matter and its suspension in seawater, light attenuation in the underwater propagation process is caused by absorption and scattering. The attenuation characteristics of seawater vary with the sea area, water depth, and season [5].

According to the various functions and protection objectives of seawater quality, it is divided into four categories:

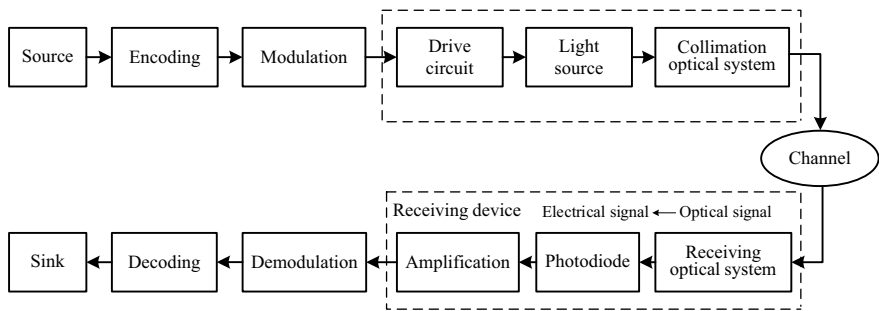


Fig. 6.1 Underwater laser system

The first category is suitable for marine fishery waters, marine nature reserves, and rare and endangered marine life reserves. The second category applies to aquaculture areas, sea baths, and marine sports or recreation areas, where the human body is in direct contact with the seawater. This category also includes industrial water use areas directly related to human consumption. The third category is applicable to the general industrial water use area and coastal scenic tourist areas. The fourth category is applicable to marine port waters and development and operation areas.

The optical properties of seawater are related to water medium, dissolved matter, and suspended matter. There are many types of dissolved and suspended substances, including inorganic salts, dissolved organic compounds, active marine zooplankton, bacteria, debris, and mineral particles. The transmission of light beam in seawater is much more complicated than that in the atmosphere; therefore, it is difficult to describe it with a single mathematical model. The influence of light wave transmission under water can be summarized into the following three aspects.

- (1) Attenuation: the attenuation of light in seawater comes primarily from the influence of absorption and scattering, which is usually represented by the absorption coefficients of seawater molecules, phytoplankton, and suspended particles and the scattering coefficients of molecules and suspended particles.
- (2) Beam diffusion: the light beam emitted from the light source will expand horizontally in the vertical direction during transmission, and its diffusion diameter is related to water quality, wavelength, transmission distance, and underwater divergence angle.
- (3) Multipath scattering: when light propagates in seawater, it is scattered by scattered particles and deviates from the optical axis, forming multiple scattering.

6.2.3 Characteristics of Underwater Laser Communication

Optical communication technology can overcome the shortcomings of underwater acoustic communication, such as narrow bandwidth, high environmental impact, low applicable carrier frequency, and long transmission delay [5], as follows.

- (1) Owing to the high frequency of a light wave and its strong information carrying capacity, large capacity wireless communication links can be established.
- (2) The transmission rate of a light wave in the water medium can reach a gigabyte, which enables fast underwater transmission of a large amount of data.
- (3) Optical communication has strong anti-electromagnetic interference ability and is not affected by seawater temperature and salinity.
- (4) The communication beam has good directionality. To intercept it, another receiver aimed at the transmitter within the line of sight is required; this will cause a communication link interruption. Such communication link interruption accident can be discovered rapidly.

- (5) With the continuous breakthrough of key technologies of semiconductor light source, the visible spectrum optoelectronic devices with small size, low price, and high efficiency are abundant. Moreover, owing to the length of the light wave, the size of the transceiver antenna is small, and its weight can be significantly reduced.

6.3 Submarine Laser Communication

6.3.1 *Forms of Submarine Laser Communication*

Submarine laser communication system can be divided into three schemes: land-based, space-based, and air-based [6].

- (1) Land-based system. A strong pulse laser beam is sent out by the land base station and is reflected to the target sea area by a mirror on a satellite to realize the communication with a submarine. In this way, the satellite borne mirror can be used to expand the beam into a wide beam to achieve a wide range of communication; it can also be controlled into a narrow beam to communicate in scanning mode.
- (2) Space-based system. A high-power laser is placed on a satellite to complete the abovementioned communication functions. The ground can control and contact the equipment on the satellite through an electric communication system. Moreover, the satellite with the best position can communicate with the submarine owing to communication between satellites.
- (3) Air-based systems. A high-power laser is placed on an aircraft. When the plane flies over the predetermined sea area, the laser beam sweeps the target sea area with a beam of a certain shape to complete the broadcast communication to the underwater submarine.

6.3.2 *Transmission of Each Dielectric Layer*

6.3.2.1 Cloud Transmittance

The occurrence probability and coverage rate of stratus, stratocumulus, altocumulus, and cirrus clouds above the ocean are above 80%. The stratus and stratocumulus clouds are water clouds with an average height of 0.4–2 km and an average thickness of 300–500 m. The liquid water content of a water cloud is concentrated in the range of 0.1–0.5 g/m³. The diameter of liquid water particles in a water cloud is 10–50 μm. High level clouds and altocumulus clouds belong to ice–water mixed clouds. Their average height ranges from several hundred to several thousand meters, and their thickness is mostly concentrated in 200–500 m. The liquid water content therein is generally between 0.03–0.1 g/m³. Cirrus is the main component of an ice cloud. In

Table 6.1 Cloud extinction coefficient β_c

Cloud type	β_c/m^{-1}
Nimbus	0.128
Altostratus	0.108
Stratus II	0.100
Dense Cumulus	0.069
Stratus I	0.067
Cumulonimbus	0.044
Stratocumulus	0.045
Clear day cumulus	0.021

middle and high latitudes, cirri generally appear at the altitude of 5–8 km. In low latitude areas, the average height of cirrus clouds is 10–14 km. The thickness of cirrus clouds ranges from several hundred to several thousand meters [6].

In the 450–550 nm visible light band, the atmospheric attenuation is primarily caused by the Mie scattering, whereas the Rayleigh scattering of molecules is negligible. Extinction coefficient is a physical quantity that characterizes the attenuation degree of an electromagnetic wave caused by the medium. It is equal to the relative value of attenuation of electromagnetic wave intensity due to absorption and scattering when electromagnetic wave propagates a unit of distance in the medium. The extinction coefficients of various clouds are shown in Table 6.1. Using the Van de Hulst formula to calculate the spectral transmittance L_c of cloud, we obtain

$$L_c = \frac{F \times 1.69}{\tau(1 - \langle \cos \theta \rangle) + 1.42} \quad (\tau \geq 10), \quad (6.1)$$

where L_c is the ratio of sea surface illumination with and without clouds, τ is the cloud optical thickness, and the relationship between τ and cloud physical thickness D is $\tau = \beta_c D$. Moreover, $\langle \cos \theta \rangle$ is the average cosine value of the scattering angle: for visible light, $\theta = 34^\circ$; therefore, $\langle \cos \theta \rangle \approx 0.83$. F is a function related to the incident angle of the light above the cloud. When $\tau < 10$,

$$L_c = F(1 - 0.046\tau). \quad (6.2)$$

6.3.2.2 Influence of Interface

When the light enters the sea from air, the transmittance of the interface is $L_{aw} = L_{aw1} L_{aw2}$, where L_{aw1} is the interface transmittance determined by the discontinuity of refractive index, and L_{aw2} is the interface transmittance determined by sea foam and stripe. When $\tau \geq 10$, the incident light at the interface is diffuse light, and when the sea surface wind speed is $v < 8\text{m/s}$, $L_{aw} \approx 0.83$.

6.3.2.3 Sea Water Transmittance

Owing to the scattering of seawater and the scattering caused by suspended particles in seawater, the size distribution of suspended particles varies considerably with water quality. The scattering of particles in seawater is two to three orders of magnitude stronger than that in the atmosphere [6].

(1) Transmittance of downward irradiance

In the case of a cloudy sky, the diffuse light of clouds is incident to the sea water. For diffuse light, the attenuation of seawater obeys the exponential law, that is, the transmittance of downward spectral irradiance is

$$L_w = \exp(-K_d Z), \quad (6.3)$$

where K_d is the downward diffusion attenuation coefficient (with a unit of m^{-1}), and its value depends on the water quality and is related to the depth. Refer to Table 6.2 for values, where Z is the depth. The values of K_d in Table 6.2 is derived from depth. In general, the value of K_d decreases as the depth increases. If the depth is divided into j layers, Eq. (6.3) can be rewritten as

$$L_w = \exp\left(-\sum_{i=1}^j K_{di} Z_i\right). \quad (6.4)$$

(2) Underwater emissivity distribution

The emissivity N pointing to any point in the water is a function of zenith angle ϕ , azimuth angle θ , and depth Z : $N(\phi, \theta, Z)$. When calculating the received light power or energy, a factor $f(\phi_r)$ related to the emissivity distribution and field of view angle must be multiplied:

$$f(\phi_r) = \frac{\int_0^{2\pi} d\theta \int_0^\phi N(\phi, \theta) \cos\phi \sin\phi d\phi}{\int_0^{2\pi} d\theta \int_0^\pi N(\phi, \theta) \cos\phi \sin\phi d\phi}, \quad (6.5)$$

where ϕ_r is the field angle. When $\phi_r > 90^\circ$, the numerical value of $f(\phi_r)$ does not increase much, so the field angle does not need to be very large. When atomic filter is used, the field angle can be considered 90° , and $f(\phi_r)$ is approximately 0.85.

Table 6.2 Downward diffuse attenuation coefficient (unit: m^{-1}), wavelength $\lambda = 459 \text{ nm}$

Sea water Class I	Sea water Class II	Sea water Class III
0.032	0.063	0.120

6.3.2.4 Background Light Power of Detector

In addition to receiving the transmitter signal, the receiver receives the natural light from the sky and the background light radiation interference produced by marine organisms. This interference reduces the signal-to-noise ratio of the system and increases the bit error rate of the system. In the daytime, the main background radiation sources are the sun and sky light. At night, it is moonlight, starlight, bioluminescence, and zodiacal light. If the sun and moon are in the field of view of the receiver, they can be treated as point sources, whereas skylight and clouds can be regarded as extended sources filling in the field of view. Since the irradiation of natural light is incoherent after multiple scattering by clouds, the background light incident to the sea surface is diffuse light, which is absorbed and scattered by seawater and reaches the receiver. At this time, the light field distribution of the background light is the same as that of the signal [7].

The background light power is denoted by

$$P_b = H_\lambda \Delta\lambda AT L_c L_w L_{aw} f(\phi_r), \quad (6.6)$$

where A is the receiving antenna area with a unit of m^2 , T is the total transmittance of the receiving optical system and filter, and $\Delta\lambda$ is the bandwidth of the filter with a unit of nm. The unit of P_b is W.

6.3.3 Time Spreading

The propagation of light beam in a medium will produce the multipath effect, causing signal spreading in space and time, because of the different optical paths of photon transmission caused by multiple scattering. Therefore, the received signal is not the original standard pulse waveform, but a complex signal after time spreading, energy attenuation, and drowning in strong noise.

Due to the multiple scattering of clouds, when $\tau < 10$, the scattered light will dominate. Then, the output signal waveform at the bottom of the cloud changes and is notably extended in time. L. B. Stotts's equation can be used to calculate the time spreading [8].

Denoting Δt_c to represent the time width of the half power point of the output pulse, we obtain [9]

$$\Delta t_c = \frac{D}{c} = \left\{ 0.3/\omega_0 \tau \theta_c^2 \left[(1 + 2.25\omega_0 \tau \theta_c^2)^{1.5} - 1 \right] - 1 \right\}, \quad (6.7)$$

where ω_0 is the ratio of the scattering coefficient to extinction coefficient. For visible light, $\omega_0 \approx 1$, and θ_c is the average scattering angle of a cloud with typical value $\theta_c = 0.66$ rad. D is the cloud thickness, and c is the speed of light.

It can be seen from Eq. (6.7) that Δt_c primarily depends on the optical thickness and extinction coefficient. For stratocumulus and cumulonimbus, the extinction coefficient can be set to $\beta_c = 0.04\text{m}^{-1}$. When $\tau < 10$, the direct light plays a dominant role. It can be approximated that $\Delta t_c = 0$, and the time spreading caused by other factors, such as water scattering and optical path difference, is much smaller than Δt_c and can usually be ignored. Therefore, the total time spreading of the channel is $\Delta t = \Delta t_c$.

The scattering of laser in cloud changes the time characteristics of the signal. The peak value is significantly reduced, the rise time of pulse is delayed, and the total time spreading effect is produced. The time spreading is from tens to thousands of nanoseconds, and such a large dynamic range creates difficulties to the signal detection of the receiver and considerably reduces the signal-to-noise ratio of the communication system.

6.3.4 Energy Equation

The single pulse energy of the signal received by the detector is expressed as [10]

$$E_r = E_P(A/S)L_cL_wL_{aw}Tf(\phi_r), \quad (6.8)$$

where E_P is the unit pulse energy output by the transmitter optical system (with a unit of J), S is the underwater spot area at the receiving point (m^2), and L_{aw} is the atmospheric molecular scattering transmittance; $L_a = e^{-\beta_a H}$, where H is the aircraft flying height, and β_a is the atmospheric scattering coefficient that can be set as $\beta_a = 4 \times 10^{-5} \text{m}^{-1}$ in an approximate calculation. If the energy attenuation caused by various factors is expressed in dB, Eq. (6.8) becomes

$$\begin{aligned} 10 \lg(E_r/E_P) = & 10 \lg(A/S) + 10 \lg L_a + 10 \lg L_c \\ & + 10 \lg L_{aw} + 10 \lg L_w + 10 \lg T + 10 \lg f(\phi_r) \end{aligned} \quad (6.9)$$

When designing an underwater laser communication system, the attenuation of the signal must be considered and the transmission energy must be budgeted reasonably.

6.3.5 Trends of Submarine Laser Communication

Underwater laser communication is still in development, and the general development trend is shown in the following aspects. (1) The communication rate is constantly increasing. The problems of laser propagation in seawater and cloud with blue and green wavelengths under various weather conditions and ocean conditions have been solved. With breakthroughs in key technologies and experimental

success, the research focus of submarine laser communication has turned to improve the communication performance of the system, especially the communication rate. (2) Development of satellite-based communication. Compared with land- and space-based communication, satellite communication has irreplaceable advantages, such as large coverage, considerable distance from the ground battlefield, invulnerability to attack, and strong survivability, and thus can provide real-time and reliable communication support for submarines. (3) From theoretical research and experimental stage to practical direction: the key challenges of laser-to-submarine communication have been solved, which laid a foundation for the practical application of laser communication technology in submarines.

6.4 Summary and Prospects

Underwater laser communication has a strong application background and has been successfully implemented in Western developed countries. The attenuation of laser signal by water and suspended solids in water is an obstacle that must be overcome in underwater laser communication. The suppression of the attenuation and scattering of a laser signal by the underwater channel is a problem that should be addressed in the future.

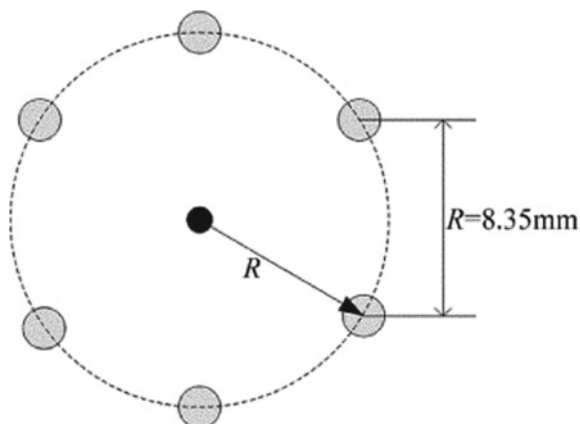
6.5 Questions

- 6.1 What are the main forms of laser to submarine communication.
- 6.2 Describe the main attenuations of laser in underwater transmission.
- 6.3 What is the difference between the multipath effect of underwater laser communication and that of atmospheric laser communication?
- 6.4 Briefly describe the development trends of underwater laser communication.

6.6 Exercises

- 6.1 For submarine laser communication, calculate (1) the mean square value of beam scattering angle, (2) the multipath time delay and (3) the average energy distribution time of receiving laser pulse signal.
- 6.2 Figure 6.2e is the layout of an LED lamp. Calculate the relative luminance distribution of light intensity on the receiving surface.
- 6.3 Analyze the relationships between blue-green laser power and communication distance.
- 6.4 For OOK laser modulation mode, whether there are “1” and “0” corresponding to the transmission information in each time unit width since only one bit is

Fig. 6.2 Layout of an LED lamp



- sent per time unit, the bit rate is equal to the transmission bandwidth. Try to analyze the decision threshold of underwater optical channel transmission.
- 6.5 When the underwater platform communicates with satellite, the laser pulse with modulation information is transmitted vertically. The laser pulse with divergence angle of 0° extends through the communication channel space and covers a certain area vertically above the underwater platform. When the satellite passes through this area, the received signal strength and signal-to-noise ratio are large enough within the limited range from the laser pulse center, so that the satellite can receive the laser pulse successfully after demodulation, the communication information can be obtained. Try to analyze the influence of chlorophyll, air sea interface and cloud on communication performance.
 - 6.6 Calculate the mean square value of scattering angle of underwater laser communication.
 - 6.7 Analyze the influence of sea surface fluctuation on underwater laser communication.
 - 6.8 Considering the geometric loss caused by the size of the receiver, space position and receiving plane angle, calculate the received optical power.

References

1. Liu Z, Liu J et al (2014) Simulation and analysis on time-domain broadening of underwater laser pulse transmission. J Changchun Univ Sci Technol (Nat Sci Edit) 04:56–59
2. Wang F, Rao J, Xiang X (2014) Research of performance of circular array light source in underwater wireless LED optical communication. Laser Technol 4:527–53.2
3. Shen N, Guo J, Zhang X (2012) Influence of bit error rate in laser underwater communication. Infrared Laser Eng 11:2935–2939
4. Li J, Ma Y, Zhou B, Wang H (2012) Ber analysis of wireless underwater optical OOk channel. Opt Optoelectron Technol 04:24–27

5. Liu J, Chen W (2010) Uplink laser communications between a satellite and a submerged platform. *Acta Photonica Sinica* 04:693–698
6. Zhou Y, Liu Q, Zhang X, Sun J (2009) A simulation method for time-domain broadening of laser pulse propagation underwater. *China laser* 36(01):143–147
7. Wang C, He Z (2011) Influence of ocean wave on wireless optical communication. *Comput Appl* 31(S1):19–22
8. Wei W, Zhang X, Rao J, Guangcheng W (2011) Study on computing the receiving optical power in underwater wireless optical communication. *China Laser* 38(09):103–108
9. Sun C (2007) Short range underwater visible optical communication. Ocean University of China, Qingdao
10. Sui M (2009) The key Technology research on underwater wireless optical communication system. Ocean University of China, Qingdao

Chapter 7

Ultraviolet Communication



Ultraviolet (UV) communication can work in the non-line-of-sight (NLOS) mode; thus, a UV communication system can easily adapt to complex terrain environment and overcome the limitation of wireless laser communication: requiring line-of-sight (LOS) mode. Additionally, compared with radio frequency communication, UV communication has the advantages of low eavesdropping, low discrimination, omnidirectionality, and strong anti-interference ability.

7.1 UV Light and Its Channel Characteristics

7.1.1 UV Light

UV light is a type of electromagnetic radiation with wavelength range of 10–400 nm. It is usually divided into near UV (315–400 nm), middle UV (280–315 nm), far UV (200–280 nm), and vacuum UV (10–200 nm) and has different characteristics depending on its wavelength. Specifically, O_3 in the upper atmosphere has a strong absorption effect on the spectrum with wavelength lower than 200 nm. Therefore, the UV transmission of this spectrum section is considerably limited, and it cannot be transmitted in the atmosphere, so it is called vacuum UV. Moreover, ozone in the atmosphere has strong absorption effect on the UV spectrum near the wavelength of 250 nm, so the UV radiation in this spectrum band hardly exists in the near-Earth atmosphere. The solar background is less than 10^{-13} w/m^2 , which is often called “solar-blind area”, with spectral range of 200–280 nm. The solar background radiation of the spectrum segment with wavelength of more than 280 nm is very strong, and the background light interferes with a working communication system.

7.1.2 Characteristics of UV Light

Owing to the existence of a large number of particles in the atmosphere, there is a large scattering phenomenon in the transmission process of UV radiation. This scattering characteristic enables the UV communication system to transmit signals in the NLOS mode, thus overcoming the weakness of other free space optical communication systems: the requirement of the LOS mode. Compared with traditional communication methods, UV communication has the following advantages:

- (1) Low eavesdropping rate: in the process of transmission, due to the absorption and scattering of atmospheric molecules and suspended particles, UV energy decays rapidly, so it is used for wireless communication with a limited range. Outside the communication range, eavesdropping is impossible even with a high-sensitivity UV detector.
- (2) Low resolution: because UV light is invisible, it is difficult for the naked eye to find the direction of the signal source. Furthermore, it is difficult to determine the location of the signal source from these scattering signals because the UV light primarily emits signals in the form of scattering.
- (3) Strong anti-interference ability: due to the absorption of ozone and oxygen in the atmosphere, the interference with UV light near the ground is very small. Furthermore, due to the scattering effect, UV light near the ground is evenly distributed, and the background signal can be removed by filtering at the receiving end. Finally, the optical signal is not affected by radio waves, and it is difficult to implement long-distance UV interference.
- (4) Omnidirectionality: there are large numbers of molecules and aerosol particles in the atmosphere, and they have a strong scattering effect on UV light. After multiple scatterings, UV photons can diffuse to all directions of local space. Therefore, the information can be received within the effective coverage range; it will not have the strong directionality of a laser.
- (5) NLOS communication: because of the strong scattering effect of UV light, it can bypass obstacles and realize NLOS communication naturally.
- (6) It works all day: the band range of ultraviolet communication is generally selected in the solar-blind area (200–280 nm). The near-earth radiation of the sun is weak even in the daytime, so there will not be a strong interference signal. Although the transmittance of UV light varies with season, altitude, climate, visibility, etc., the influence of these factors is the same as that of solar radiation. In a specific place and at a specific time, it can be regarded as a low-frequency background signal, which can be easily removed by a filter. Therefore, in general, in the face of complex and changeable environment, whether in cloudy, sunny, rainy, or foggy weather, or in smoky and dusty environment, UV communication can be executed smoothly.

7.1.3 UV Atmospheric Channel

7.1.3.1 UV Atmospheric Absorption and Scattering Characteristics

The UV transmission characteristics are primarily determined by the absorption and scattering of molecules in the atmosphere. Absorption of light radiation or light energy is a general property of matter. When light passes through a material, the electric vector of the light wave causes the charged particles in the material structure to vibrate under force, and part of the energy of the light is used to supply the energy needed for the forced vibration. Then, if the material particles collide with other atoms or molecules, the vibration energy may be converted into translational kinetic energy, which will increase the energy of the thermal motion of molecules, resulting in the heating of objects. Thus, part of the light energy is diminishing and converted into heat, and in the process of transmission, atmospheric molecules consume the energy of the UV light. Various components in the atmosphere can absorb light of different wavelengths in varying degrees, and O_3 has the strongest UV light absorption ability. The higher the concentration of O_3 , the more energy the atmosphere absorbs and the greater the transmission loss. Because of the absorption by O_3 , UV communication can only be used as short distance communication.

In a medium with uniform optical properties or an interface between two homogeneous media with different refractive indexes, the light is limited to certain directions regardless of whether it is refracted or reflected. In the other directions, the light intensity is equal to zero, and no light can be seen on the side of the beam. However, when light passes through a material with inhomogeneous optical properties, the scattering of the light on the side can be seen. The more obvious the optical inhomogeneity of the medium, the stronger the scattering. The scattering characteristic of UV radiation is the basis of UV communication. The scattering particles in the atmosphere are primarily atmospheric molecules and suspended particles. Their concentration, size, uniformity and other characteristics affect the transmission characteristics of UV light. Atmospheric light scattering can be divided into Rayleigh scattering and Mie scattering.

7.1.3.2 Transmission Characteristics of UV Communication

According to the working mode, the solar-blind UV communication system can be divided into LOS, quasi-LOS, and NLOS. NLOS has three working modes: NLOS (A), NLOS (B), and NLOS (C), as shown in Fig. 7.1.

The LOS communication refers to the communication in which the transmitter is in the field of view of the receiver and the optical path of the communication is free of any obstacles; this is similar to typical laser communication. NLOS means that the two points of view that need to communicate are blocked and cannot see each other, and more than 50% of the Fresnel area (the circular area around the line of sight)

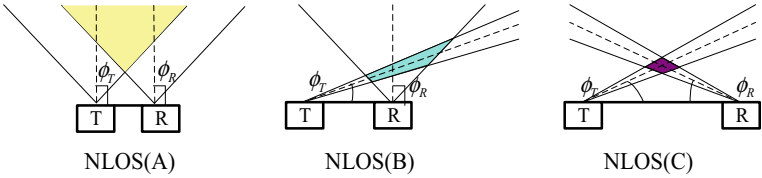


Fig. 7.1 Typical configurations of non-line-of-sight (NLOS) communication

is blocked. Due to the characteristics of solar-blind UV light intensity scattering, a non-direct viewing transmission path can be formed if the field angles of transmitter and receiver overlap.

According to the angle ϕ_T and ϕ_R between the horizontal plane and the optical axes of the transmitter and receiver, respectively, NLOS can be divided into Classes A, B, and C. In class B, ϕ_T and ϕ_R cannot be 90° at the same time, as shown in Fig. 7.1. According to the requirements of the UV communication system, it is easy to change ϕ_T and ϕ_R among the three working modes. The performance comparison of UV communication system in various working modes is shown in Table 7.1.

7.1.4 Characteristics of UV Atmospheric Channel

When the UV light of a carrier signal passes through atmospheric space, it is inevitably affected by various components of atmosphere, weather, and climate conditions; these are closely related to the quality of the communication and performance of the communication system [3].

7.1.4.1 Solar-Blind UV Light

More than 99% of the solar radiation spectrum is between 150 and 4000 nm. In this band, about 50% of the solar radiation energy is in the visible spectrum (wavelength 400–760 nm), 7% in the ultraviolet spectrum (wavelength <400 nm), and 43% in the

Table 7.1 Performance comparison of different ultraviolet working modes

Operation Mode	ϕ_T	ϕ_R	Omni-direction	Distance/km	Overlapping space	Bandwidth
LOS	/	/	None	2~10	Limited	Widest
QLOS	/	/	None	$<\text{LOS}$	Limited	Wide
NLOS(A)	$=90^\circ$	$=90^\circ$	Good	0~1	Infinite	Narrowest
NLOS(B)	$\leq 90^\circ$	$=90^\circ$	Fair	1.5~2	Limited	Wide
NLOS(C)	$<90^\circ$	$<90^\circ$	Poor	2~5	Limited	Wide

infrared spectrum (wavelength >760 nm) and the maximum energy is concentrated at the wavelength of 475 nm.

Figure 7.2 shows the specific division of UV spectra according to three different classification methods. The strong absorption of 20–280 nm UV light by the ozone layer (10–50 km) in the upper troposphere makes the solar background in the troposphere (especially near earth) lower than 10–13 w/m²; that is, there is almost no UV in the sunlight on the earth surface. This band is called the “solar-blind zone”. The horizontal and vertical transmittance of atmospheric environment is shown in Fig. 7.3. In the figure, we can see that the UV light attenuation at about 250 nm is very large, and there is a slight difference between the horizontal and vertical directions.

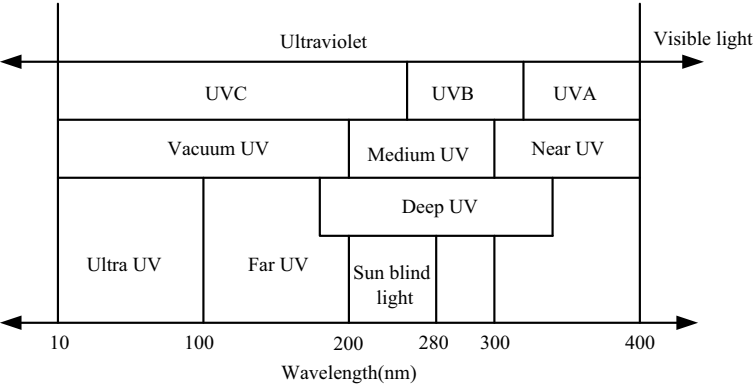
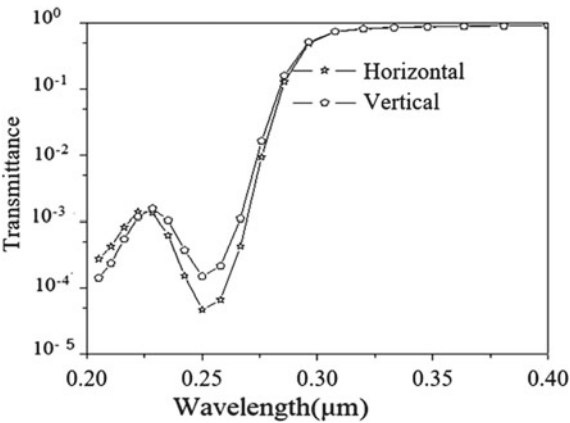


Fig. 7.2 Spectrum distribution of the ultraviolet band

Fig. 7.3 Transmittance of ultraviolet light



7.1.4.2 Main Factors Affecting UV Communication

NLOS UV communication uses the atmosphere as the transmission medium. When the UV light carrying signal is transmitted in space, the communication quality, communication system performance, and transmission range are inevitably affected by the concentration of O₃ in the atmosphere, and by the concentration, size, uniformity and working wavelength of scattering particles.

(1) Atmospheric absorption

When UV light passes through the atmosphere, various components in the atmosphere absorb it in different degrees. In the ultraviolet region (0.2–0.4 μm), there is a strong absorption band of ozone between 0.2 and 0.264 μm and weak absorption band between 0.3 and 0.36 μm. Sulfur dioxide and ozone have a strong UV light absorption ability. Although the content of ozone in the atmosphere only accounts for 0.01–0.1% of the total atmosphere, its absorption of solar radiation energy is significant.

(2) Atmospheric scattering

The main scatterers in the atmosphere are atmospheric molecules and aerosol particles. Due to their size differences, they have different scattering characteristics. The closer the atmospheric particles are to the UV light, the stronger the scattering. The molecular size is much smaller than the UV wavelength, which leads to Rayleigh scattering, whereas aerosol particles are larger than the UV light wave, which leads to Mie scattering. The results show that Rayleigh scattering plays a dominant role in the clear sky atmosphere. Therefore, in a theoretical calculation, only Rayleigh scattering is considered in sunny weather, while the scattering effect of suspended particles is ignored, which can be considered a reasonable approximation. The radii and concentrations of several main scattering particles in the atmosphere are listed in Table 7.2.

(3) Atmospheric turbulence

When a beam passes through vortices with different refractive indices, atmospheric turbulence effects, such as beam bending, drift, and propagation distortion, occur and cause the received light intensity to flicker and jitter.

Table 7.2 Radii and concentrations of atmospheric scattering particles

Type	Radius/μm	Concentration/cm ⁻³
Atmospheric molecule	10 ⁻⁴	10 ¹⁹
Aitken core	10 ⁻³ –10 ⁻²	10 ⁻⁴ –10 ²
Haze particle	10 ⁻² –1	10–10 ³
Fog drop	1–10	10–10 ²
Cloud drop	1–10	10 ⁻³ –10 ²
Raindrop	10 ² –10 ⁴	10 ⁻⁵ –10 ⁻²

7.2 Characteristics of NLOS UV Transmission

7.2.1 Ellipsoid Coordinate System

The analysis of UV single scattering link model is based on the ellipsoid coordinate system, as shown in Fig. 7.4. The surface of the ellipsoid is obtained by rotating the ellipse around its principal axis. The coordinates of any point on the ellipsoid are uniquely determined by radial coordinates ξ , angular coordinates η , and azimuth coordinates ϕ . The parameter definition of transformation from the rectangular coordinate system X–Y–Z to ellipsoid coordinate system is shown in Fig. 7.4.

For homogeneous scattering and absorbing media, the received scattering signal power is

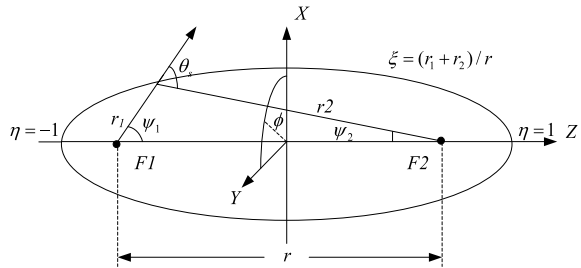
$$h(t) = \frac{Q_t A_r T_{of} c k_s \exp(-k_e c t)}{2\pi \Omega_t r^2} \times \int_{\eta_1(\frac{ct}{r})}^{\eta_2(\frac{ct}{r})} \frac{2G[\Phi(\frac{ct}{r}, \eta) p(\theta_s)]}{(\frac{ct}{r})^2 - \eta^2} d\eta, \quad (7.1)$$

where $h(t)$ is the impulse response of the scattering channel, θ_s is the scattering angle; Ω_t is the transmitting angle (solid angle), A_r is the receiving area, and $p(\cdot)$ is the scattering phase function. k_e , k_s and k_a are the extinction, scattering, and absorption coefficients of the medium, satisfying $k_e = k_s + k_a$. Finally, c is the speed of light, and T_{of} is the light transmission coefficient.

7.2.2 UV Scattering Communication

The UV single scattering link model based on the ellipsoidal coordinate system is shown in Fig. 7.5. When $t = 0$, the UV light with total energy E is transmitted at divergence angle φ_t (the transmit energy per unit solid angle is $\frac{E_t}{\Omega}$ with a unit of joules/cubic radian), solid angle $\Omega = 4\pi \sin^2(\theta_t/2)$, and reaches the scatterers in a distance of r_1 after travelling time $t = r_1/c$ along a transmission angle θ_t . Finally,

Fig. 7.4 Ellipsoidal coordinate system



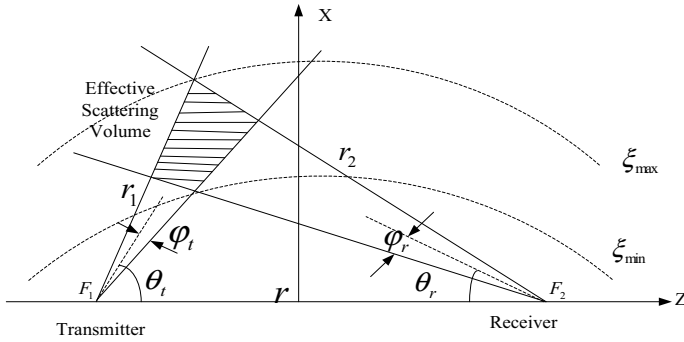


Fig. 7.5 Ultraviolet single scattering communication link model

the energy density is $\frac{E_t}{\Omega} \frac{e^{-kr_1}}{r_1^2}$, where k is the extinction coefficient, and e^{-kr_1} is the ratio of UV light energy after transmission in atmosphere in a distance r_1 to that in vacuum.

The volume differential dv of the scatter can be regarded as a second-order point light source, which is generated by the interaction between the UV light and medium. The energy density of the secondary point light source is $\left[\frac{E_t}{\Omega} \frac{e^{-kr_1}}{r_1^2} \right] \frac{k_s}{4\pi} p(\cos \theta_s) dv$ (unit: Joule/cubic radian), where k_s is the atmospheric scattering coefficient, and $p(\cos \theta_s)$ is the single scattering phase function.

Part of the energy of the secondary light source will be transmitted along the direction that can reach the receiving end. At $t = (r_1 + r_2)/c$ moment, the total energy density received by the receiving end from the secondary point light source is shown in Eq. (7.2).

$$\delta E_r = \left[\frac{E_t}{\Omega} \frac{e^{-kr_1}}{r_1^2} \right] \left[\frac{k_s}{4\pi} p(\cos \theta_s) dv \right] \frac{e^{-kr_2}}{r_2^2} \quad (7.2)$$

According to ξ , η , ϕ , and focal length r in the ellipsoidal coordinate system, the expression of the volume differential of scatterers can be obtained as

$$dv = (r/2)^2 (\xi^2 - \eta^2) d\xi d\eta d\phi. \quad (7.3)$$

Since $r_1 = \frac{r}{2}(\xi + \eta)$, and $r_2 = \frac{r}{2}(\xi - \eta)$, the total volume differential energy density of the receiver is shown in Eq. (7.4).

$$\delta E_r = \frac{E_t k_s}{4\pi \Omega} \frac{e^{-kr\xi}}{(r/2)(\xi^2 - \eta^2)} p(\cos \theta_s) d\xi d\eta d\phi. \quad (7.4)$$

Because $\xi = (r_1 + r_2)/r$ is $\xi = ct/r$, $d\xi = cdt/r$ is $\frac{d\xi}{dt} = \frac{c}{r}$, and the instantaneous differential volume received power density at the receiver is shown in Eq. (7.5).

$$\delta P_r = \frac{c E_t k_s}{2\pi \Omega r^2} \frac{e^{-kr\xi}}{(\xi^2 - \eta^2)} p(\cos \theta_s) d\eta d\phi \quad (7.5)$$

The total receiving power of the receiver is as follows:

$$P_r(\xi) = \begin{cases} 0, & \xi < \xi_{\min} \\ \int_{\eta_1(\xi)}^{\eta_2(\xi)} \int_{\phi_1}^{\phi_2} \frac{c E_t k_s}{2\pi \Omega r^2} \frac{e^{-kr\xi}}{(\xi^2 + \eta^2)} p(\cos \theta_s) d\eta d\phi, & \xi_{\min} < \xi < \xi_{\max} \\ 0, & \xi > \xi_{\max} \end{cases} \quad (7.6)$$

UV scattering communication is divided into two processes: transmission from the transmitter to scatterer and transmission from the scatterer to receiver.

7.2.3 NLOS Scattering Characteristics

The bit error rate of non-direct viewing UV transmission system depends on modulation mode, detector type, transmitter power, path loss, scintillation index, and noise. When OOK modulation is used, the bit error rate of the system can be expressed as [4]

$$BER_{e\text{-OOK}} = \frac{1}{2} \operatorname{erfc} \left[\frac{1}{\sqrt{2}} \frac{RGP_r}{\left[\left(qG^2FR(P_r + P_{bg})R_b + qI_{dc}R_b + \frac{2K_bT_0F_iR_b}{R_L} \right) \right]^{\frac{1}{2}}} \right] + \left[q \left(G^2FR_b + I_{dc} \right) R_b + \frac{2K_bT_0F_iR_b}{R_L} \right]^{\frac{1}{2}}, \quad (7.7)$$

where $\operatorname{erfc}(\cdot)$ is the error function, K_b is the Boltzmann constant, T_0 is the absolute temperature, F_i is the noise figure, R_L is the load impedance, q is the electronic charge, I_{dc} is the dark current of the detector, the bit rate is $R_b = \frac{1}{T_b}$, T_b is the bit time width, P_r is the photocurrent, G is the gain, F is the noise figure of the photodetector, and P_{bg} is the background radiation intensity received.

The bit error rate of PPM modulation can be expressed as [4]

$$BER_{e\text{-PPM}} = \frac{L}{4} \operatorname{erfc} \left[\sqrt{\frac{\log_2 L}{2L}} \frac{RGP_r}{\left[\left(qG^2FR(P_r + P_{bg})R_b + 2qI_{dc}R_b + \frac{4K_bT_0F_iR_b}{R_L} \right) \right]^{\frac{1}{2}}} \right], \quad (7.8)$$

where L is the number of time slots modulated by PPM. The average SNR can be expressed as

$$\langle \text{SNR}_{T-\text{NLOS}} \rangle = \frac{\text{SNR}_{0,\text{NLOS}}}{\sqrt{\left(\frac{P_{r0}}{\langle P_r \rangle}\right)} + \sigma_y^2 \text{SNR}_{0,\text{NLOS}}}, \quad (7.9)$$

where P_{r0} is the received power without turbulence, $\langle P_r \rangle$ is the average power received, and $\text{SNR}_{0,\text{NLOS}}$ is the SNR without turbulence. If $P_{r0} \approx \langle P_r \rangle$, then

$$\text{SNR}_{0,\text{NLOS}} = \sqrt{\frac{y_0}{\frac{2Rhc}{\lambda}}}, \quad (7.10)$$

where y_0 is the signal power received without turbulence, R is the data rate, h is the Planck constant, c is the speed of light, and λ is the wavelength. The bit error rate under atmospheric turbulence can be expressed as [5]

$$\text{SNR}_{T,\text{NLOS}} = \frac{1}{2} \int_0^\infty f_y(y) \text{erfc}\left(\frac{\langle \text{SNR}_{T-\text{NLOS}} \rangle y}{2\sqrt{2}}\right) dy. \quad (7.11)$$

7.3 Solar-Blind UV NLOS Communication Network

Wireless solar-blind UV communication combines the characteristics of optical and wireless communication. It uses UV light as information carrier and free atmosphere as communication medium. The entire process does not require any wired channel. It can be widely used in exhibitions, short-term rented buildings, temporary outdoor workplaces, or during earthquakes and other emergencies. However, due to the serious absorption of ultraviolet light in the atmosphere, the communication distance of ultraviolet light is limited, so it can only work in a short distance communication mode. Therefore, a new direction is to combine UV communication with a wireless mesh network to expand the communication range.

7.3.1 Wireless Mesh Communication Network

In a wireless mesh network, mesh topology, which can also be considered a multipoint-to-multipoint network topology, is adopted. In this type of mesh network structure, each network node is connected by wireless multi-hop mode through other adjacent network nodes. According to the node function division, wireless mesh network can be divided into three types of network structures: peer-to-peer, hierarchical, and hybrid network structures.

7.3.1.1 Peer-to-Peer Wireless Mesh Networks

Peer-to-peer wireless mesh network, also known as terminal device mesh network, is the simplest network structure of a wireless mesh network. As shown in Fig. 7.6, all nodes in the graph are peer-to-peer individuals with identical characteristics, including the same MAC layer, routing, security, and management protocols. All nodes in the network have the function of a mesh router and are enhanced end-user equipment. Peer-to-peer wireless mesh network is a mobile ad hoc network. It is suitable for a situation with a small number of nodes, which does not require access to the core network. It provides a new communication condition for a situation in which there is no or inconvenient use of the existing infrastructure.

7.3.1.2 Hierarchical Wireless Mesh Network

Hierarchical wireless mesh network is also called infrastructure/backbone mesh network. As shown in Fig. 7.7, it can be divided into upper and lower parts. A mesh structure with self-configuration and self-healing function is formed between the upper layers of mesh routers. It comprises basic network facilities, which can be used for the connection of the lower layer of mesh clients. Moreover, the mesh router with the gateway function can be connected to the Internet.

7.3.1.3 Hybrid Wireless Mesh Network

The hybrid network structure is shown in Fig. 7.8. It is a hybrid structure of terminal device and infrastructure/backbone mesh network structures. In this architecture, the terminal node can support not only the common devices of wireless local area network but also mesh devices with routing functions. The devices can be interconnected in

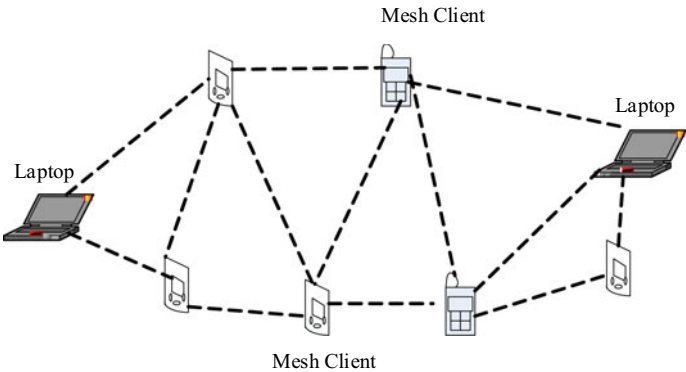


Fig. 7.6 Peer-to-peer wireless mesh network architecture

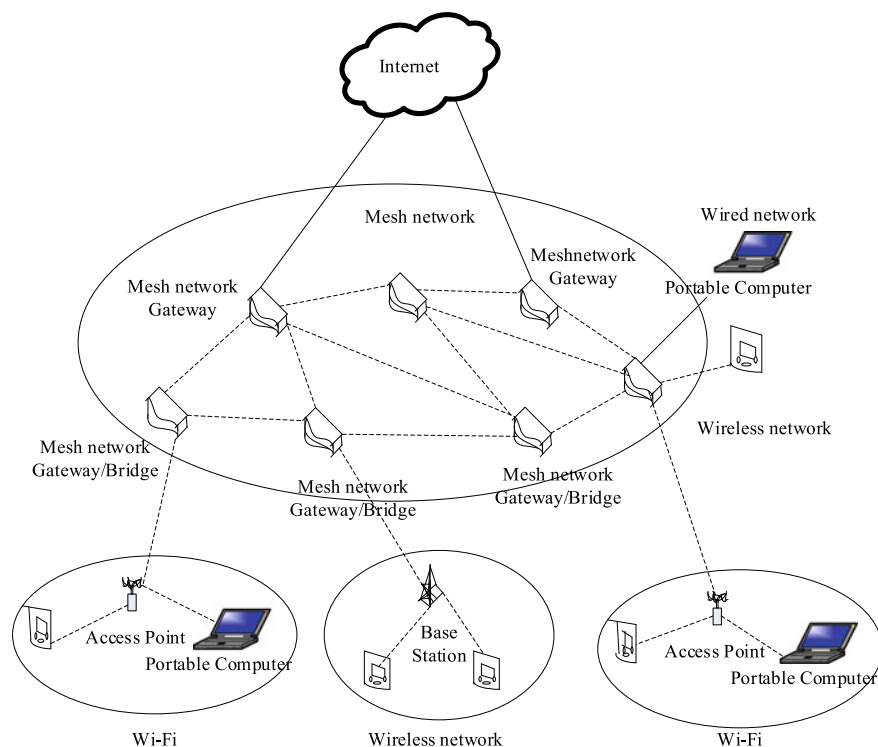


Fig. 7.7 Hierarchical wireless mesh structure

an ad hoc mode. In terms of the functions of the network, the terminal equipment can not only connect with other networks to realize wireless broadband access but also directly communicate with other users in this layer. Additionally, it can forward data as a router and send it to the destination node.

7.3.2 Wireless UV Mesh Communication Network

The wireless UV mesh communication network discussed in this chapter is composed of wireless UV communication and peer-to-peer wireless mesh network. Since the peer-to-peer mesh network is the simplest of the three network forms and can well reflect the multipoint-to-multipoint lattice network structure, it was chosen as the mesh network. As shown in Fig. 7.9, in the wireless UV communication network, if node *A* wants to send information to node *D*, it has multiple multi-hop paths to choose from, and an interruption of a certain link will not disrupt it. The network can be used in areas where infrastructure is not in place, communication cables cannot be installed, or temporary communication is required.

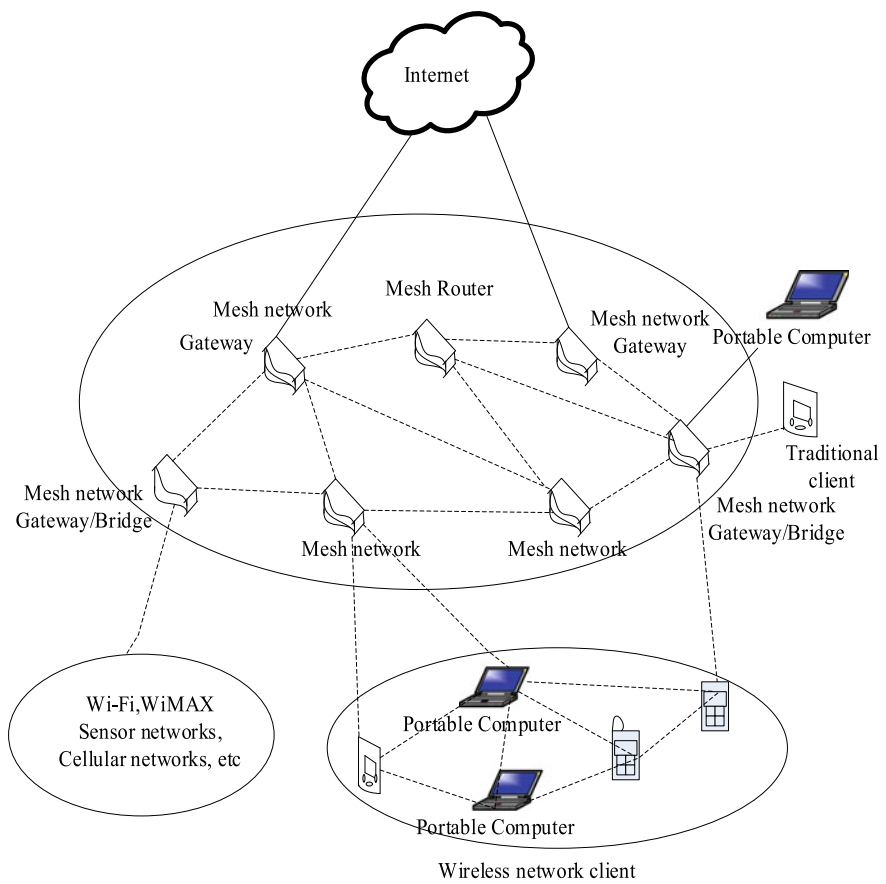


Fig. 7.8 Hybrid wireless mesh network structure

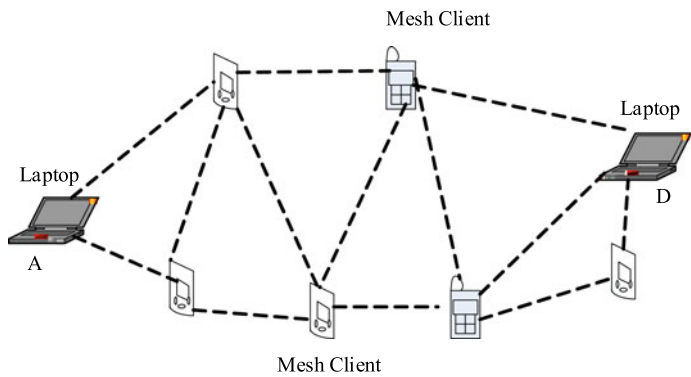


Fig. 7.9 Peer-to-peer wireless ultraviolet mesh communication network

The wireless UV mesh communication network is based on the multipoint-to-multipoint UV communication link. Therefore, the working mode of configuring communication for each node in the network must be based on the function of each node in the network. When the nodes in the network are in the state of group sending or searching, it can work in the NLOS (A) mode, because at this time, the nodes have the best omnidirectionality and can directly interact with many nodes in the surrounding coverage. When the network node wants to perform multi-node communication in a certain direction, it can work in the NLOS (B) mode, because the communication node has a certain omni-directionality under the premise of guaranteeing a certain bandwidth. When the network node needs to perform point-to-point NLOS broadband communication at a long distance in a certain direction, it can work in NLOS (C) mode, because the communication distance between communication nodes is the longest with the largest bandwidth.

Wireless UV mesh communication network can ensure smooth communication in a complex terrain environment, can be deployed quickly, is easy to install, and meets the modern communication requirements of strong security and mobile flexibility. Therefore, wireless UV communication network has a wide range of application prospects.

7.4 Summary and Prospects

UV communication generally refers to solar-blind UV communication. Due to the scattering effect of the atmosphere and suspended solids in the atmosphere, UV light can be transmitted in NLOS scenarios, and its attenuation is very large. The development direction of UV ad hoc network is to use the characteristics of NLOS transmission of UV light and expand the communication range by using the characteristics of an ad hoc network. The problems of unidirectional nodes and “deaf” nodes must be overcome in the networking of UV ad hoc networks. Finally, an important problem for UV communication to overcome is the suppression of the atmospheric channel distortion through channel equalization.

7.5 Questions

- 7.1 What wavelength range is used by solar-blind ultraviolet? Why is called solar-blind ultraviolet?
- 7.2 What are the main factors affecting UV communication?
- 7.3 Please briefly describe three modes of UV communication.
- 7.4 Please briefly describe the advantages of UV communication.
- 7.5 Please briefly describe the idea of UV communication networking.

7.6 Exercises

- 7.1 Try to analyze the bandwidth of UV NLOS communication according to the channel impulse response of UV channel [1].
- 7.2 Try to determine the integral limit in Eq. (7.1).
- 7.3 Try to determine the impulse response of the wireless UV communication system [1].
- 7.4 Please analyze the channel capacity of wireless ultraviolet communication system [1].
- 7.5 Try to analyze the phase function in UV NLOS communication [2].
- 7.6 Considering the single scattering and wavelength at 260 nm, try to calculate the energy and path loss received by the receiver under ellipsoid coordinate system [3].

References

1. Luo C, Li J, Chen X (2011) Channel analysis of wireless ultraviolet communication. *Progress Laser Optoelectron* 04:31–36
2. He H (2015) Study on volume scattering phase function in UV NLOS single scattering link. *Progress Laser Optoelectron* 03:113–117
3. Yang G, Li X, Chen M et al (2015) Ultraviolet atmospheric scattering in solar blind zone. *Sci Technol Eng* 02:236–240
4. Franz JH, Jain VK (2002) *Optical communication components and systems*, 1st edn. Narosa, India
5. Andrews LC, Phillips RL (2005) *Laser beam propagation through random media*. 2nd edn. SPIE Press

Chapter 8

Acquisition, Aiming, and Tracking Technology



Microwave antenna does not need precise antenna alignment owing to its high transmitting power and large divergence angle. In contrast, wireless laser communication requires good beam convergence; therefore, acquisition, pointing, and tracking become key problems.

8.1 Acquisition, Pointing, and Tracking System

8.1.1 Concepts

Acquisition, pointing, and tracking (APT) system is the basis of wireless laser communication and its most challenging component.

- (1) Acquisition is the determination and identification of a target in an uncertain area. The uncertain area is scanned until a beacon optical signal is received in the acquisition field of view; this lays the foundation for the subsequent alignment and tracking.
- (2) Pointing ensures that the communication transmitting end and receiving end antenna axes of view maintain precise coaxiality during the communication process. However, in a long-distance laser communication system, when the transmitter and receiver move relatively fast in the tangential direction, it is necessary to aim ahead at the transmitter.

Consider a model shown in Fig. 8.1, where a signal is transmitted by a moving station at r_1 and received by a ground station. When the signal is received, the moving carrier has arrived at the point r_2 . In the uplink transmission, the motion of r_1 and r_2 and the additional motion of r_3 must be considered. When the receiver on the moving carrier receives the signal, the carrier is already at the point r_3 . The angle between the receiving and transmitting vectors from the ground station is called the lead angle.

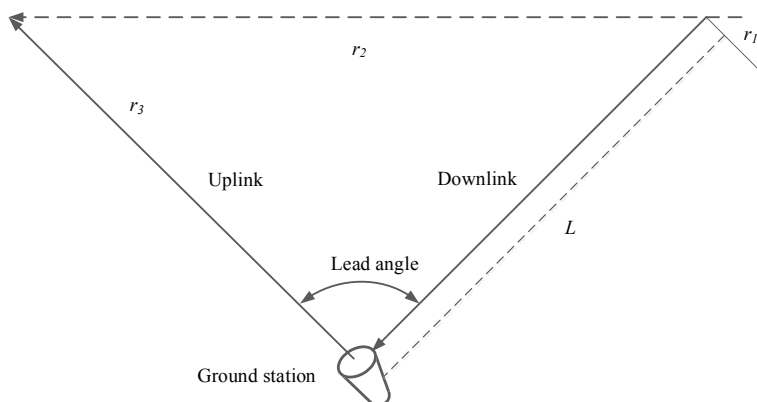


Fig. 8.1 Schematic diagram of advanced aiming

- (3) Tracking: the alignment error caused by the relative movement between the transmitter and receiver, the spot flicker and drift caused by atmospheric turbulence, and the influence of platform vibration is controlled within the allowable range through tracking after the acquisition and pointing are completed.

8.1.2 Operating Principle

The APT system of wireless laser communication is divided into two stages, as shown in Fig. 8.2. When the beacon light has not been captured by the receiver, an open-loop scanning is performed within the uncertain area, whose information is imported from the guidance information, that is, the position coordinate information of the receiver and transmitter. As the guidance information is received by the main control computer, the relative position of the receiver and transmitter is calculated and fed to the turntable to point to the target. Here, the speed measurement unit and angle control unit are the feedback units of the system and constitute a double closed-loop control system. The image processing unit as the spot position detection mechanism provides the spot position information for the system. Due to the deviation of guidance information and the error of the actuator, there will be errors after the turntable points, which, in general, cannot aim accurately at the target.

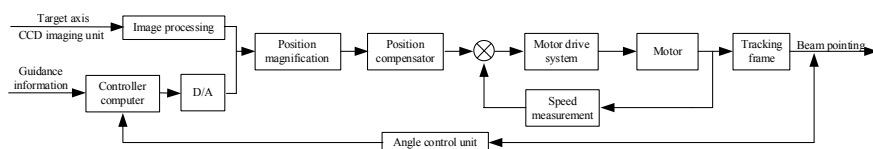


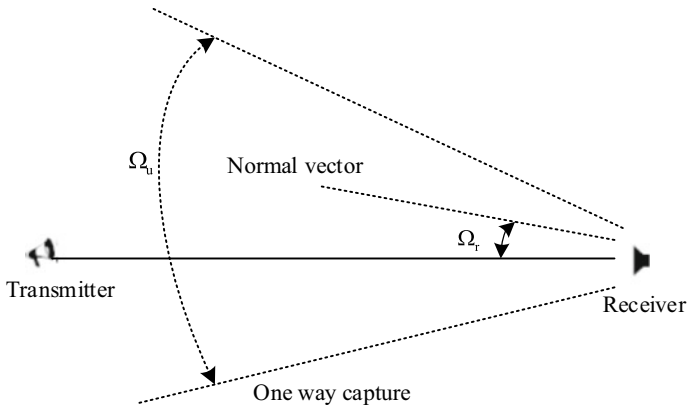
Fig. 8.2 Typical acquisition, pointing, and tracking system of a wireless laser communication

As the communication beams of transmitter and receiver are successfully captured, the respective beacon lights are kept in the field of view to achieve stable coarse tracking. Owing to the dynamic lag error caused by the relative motion and platform vibration, the coarse tracking servo with limited bandwidth is unable to meet the communication requirements. Therefore, it is necessary to proceed with the fine tracking unit to further suppress the coarse tracking residual. The axis centers of the coarse and fine tracking loops coincide, and the field of view of fine tracking should be greater than the maximum error of coarse tracking. Using the high resolution and bandwidth capability of fine tracking, the dynamic lag error and platform residual error can be effectively suppressed. The coarse tracking mechanism is generally a turntable system controlled by a servo motor, and the fine tracking mechanism is a piezoelectric ceramic galvanometer. Fine and coarse tracking cameras can detect the spot position with two types of accuracy to meet the different requirements of capture and tracking. The communication detector receives the signal beam from the opposite antenna collected by a Cassegrain system. The communication laser emits low-power and narrow signal beam, and the beacon light laser emits high-power and wide beacon beam. In the integrated transceiver, the optical splitter is used to achieve the isolation of receiving and transmitting, to prevent the mutual interference of the beam emitted by the local laser and the beam received by the antenna. Otherwise, the communication detector would have difficulty distinguishing the source of the received signal. A diagram of an APT system is shown in Fig. 8.3.

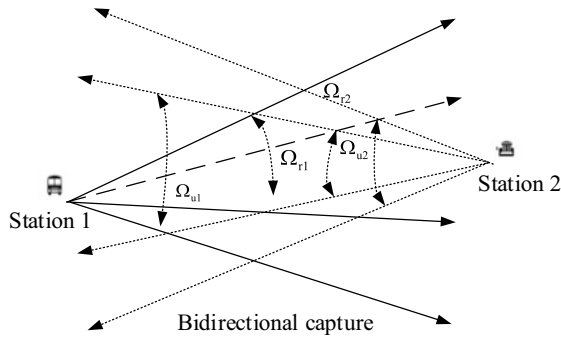
8.2 Automatic Acquisition

Space acquisition requires pointing the antenna in the direction of the field of arrival and adjusting the axis of view of the receiving antenna to be consistent with the angle of arrival of the beam. The angle of arrival is the angle between the elevation and azimuth of the incident beam. The angle between the angle of arrival and the normal vector can be within a specified solid angle. This allowable angle is also called search resolution angle and is denoted by Ω_r . The minimum resolution angle is also the diffraction limit angle; Ω_r is the diffraction limit angle: $\sin\Omega_r = 1.22\lambda/D$, where λ is the wavelength, and D is the aperture diameter. In a practical design, a relatively large resolution is required.

As shown in Fig. 8.4, the transmitter points to the receiver, and if the beam of the transmitter is aimed without error, the beam illuminates the receiving point. The receiver detects that the transmitter is within a certain uncertainty solid angle Ω_r , which is defined from the position of the receiver. The antenna of the receiver is aimed at the vertical direction of the receiving antenna within a predetermined field of view, and the antenna is aimed at a vertical direction of view of the receiver. Generally, $\Omega_r \ll \Omega_u$.



(a)



(b)

Fig. 8.4 Schematic diagram of space capture: **a** one-way acquisition and **b** two-way acquisition [2]

8.2.1 Open-Loop Acquisition Mode

To achieve acquisition with high probability, it is necessary for the divergence angle of the laser beam of the transmitter and the field-of-view angle of the receiver to cover the uncertain area by staring or scanning. Based on the different application of wireless laser communication system, two-way acquisition can be divided into the following three working modes.

(1) Gaze-gaze capture mode

For gaze acquisition mode, the divergence angle of beacon light should be greater than the uncertainty area of the open-loop acquisition, and the field of view of receiver

should be larger than the uncertainty area of the open-loop acquisition. When the receiver maintains a static attitude and only receives the optical signal, it is called gaze.

(2) Gaze-scan capture mode

According to the difference of beam divergence angle, field-of-view angle, and open-loop uncertainty region, this method can be divided into two subtypes:

- ① When the background light from the sky is strong, the receiving field-of-view angle is shrunk to reduce the influence of background light. Then, the acquisition field of view of the receiver is scanned. The transmitting power of the beacon light is increased to work in staring mode. This mode is suitable for ground-to-air or satellite-to-ground laser communication.
- ② When the background light from the sky is weak, the receiving field-of-view angle is increased, and the capture field of view of the receiver is staring at the uncertain area. Then, the uncertain area is scanned for the beacon light. This mode is suitable for interstellar laser communication [1].

The approximate expression of the capture time is [1]

$$T_{acq} \approx \frac{1}{(1+K)^2} \left[\left(\frac{\Omega_u}{\Omega_r} \right)^2 + \frac{1}{2} \left(\frac{\Omega_u}{\Omega_r} \right) \right] T_d N_t, \quad (8.1)$$

where Ω_u is the open-loop capture uncertainty area, Ω_r is the beam divergence angle, T_d is the dwell time, N_t is the number of scans, and K is the overlap coefficient of scanning. Due to the influence of the platform vibration, spot flicker, and spot drift, it is necessary for two near scanning fields to overlap to prevent a missed scan. A missed scan means that there is a scanning gap between the upper and lower scanning paths when the scanning line breaks. If the target position is in the scanning gap, a missed scanning occurs. As K increases, the capture time increases accordingly. Therefore, the value of K is generally set to 10–15%. T_d is the dwell time of the spot on the detector during the acquisition, and $T_d = T_S + T_F$, where T_S is the stepping time of the servo turntable, and T_F is the working frame period of the camera. In Eq. (8.1) the first term in the brackets $(\Omega_u/\Omega_r)^2$ is the number of steps required to scan the entire uncertain area, and the second term $(\Omega_u/\Omega_r)/2$ is the number of scans back to the center of the determined capture area after the scan. When the number of scanning steps is large, the second term can be ignored. Then, Eq. (8.1) can be written as [1]

$$T_{acq} \approx \frac{1}{(1+K)^2} \left(\frac{\Omega_u}{\Omega_r} \right)^2 T_d N_t. \quad (8.2)$$

3. Skip-scan capture mode

Owing to the limitations of beacon power, the divergence angle of laser beam cannot cover the entire area, and the capture field of view cannot be sufficiently large to

cover the entire uncertain area due to the influence of the background light from the sky. Therefore, it is necessary to capture the beam by skip-scan. Its working process is described as follows. The master optical transceiver works in a skip mode, and the slave optical transceiver works in a scan mode. In a working cycle, the main optical view axis of the transceiver is kept unchanged, and a complete scanning process by the optical transceiver is performed. When the beacon light is not captured in this working cycle, the master optical transceiver skips once and enters the next working cycle. That is, the view axis of the main optical transceiver jumps a certain distance, so that the scanning range of the next working cycle of the slave optical transceiver does not coincide with the scanning range of the previous working cycle. The number of steps of the main optical transceiver depends on the ratio of the captured uncertain area to the divergence angle of the beacon beam; the number of scans from the secondary optical transceiver depends on the ratio of the captured uncertain area to the captured field of view. Therefore, its approximate expression is [1]

$$T_{\text{acq}} \approx \frac{1}{(1-K)^2} \left(\frac{\Omega_u}{\Omega_{r1}} \right)^2 \frac{1}{(1-K)^2} \left(\frac{\Omega_u}{\Omega_{r2}} \right)^2 T_d N_t, \quad (8.3)$$

where Ω_{r1} and Ω_{r2} are the divergence angle of the beacon beam and the acquisition field-of-view angle, respectively.

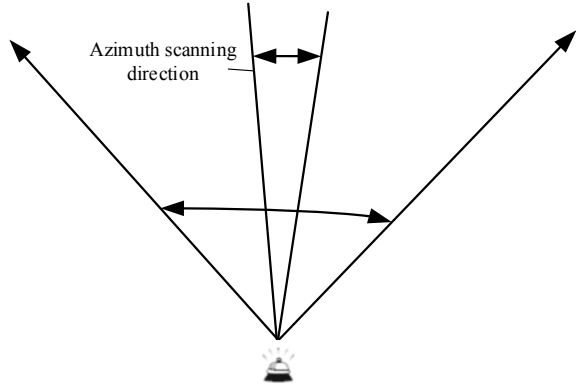
8.2.2 Scanning Modes

Because the field-of-view angle and beacon beam divergence angle of the optical transceiver are far smaller than the captured uncertainty area, it is necessary to perform open-loop scanning in the uncertain area. There are four common scanning methods: antenna scanning, focal plane scanning, focal plane array scanning, and step-by-step searching. In most laser systems, the optical telescope is adopted as the antenna with integrated transceiver. It is a high coaxial type, so antenna scanning is used at present, whereas focal plane scanning and focal plane array scanning are rarely used due to the limitation of volume.

8.2.2.1 Antenna Scanning

The receiving lens and optical detection system with a fixed field of view Ω_r are considered. The receiving system scans the entire uncertainty angle Ω_u , as shown in Fig. 8.5. In general, we only consider the scanning in azimuth. During scanning, the output of the photodetector is detected until it is confirmed that the beam has been received. This can be realized by setting a threshold value in the photodetector. Antenna scanning includes spiral scanning, raster scanning, and raster spiral composite scanning. A monochromatic point source beacon is considered to transmit

Fig. 8.5 Antenna scanning method



beacon light continuously, and the average count rate on the photo detector of the receiver is $ns = \alpha P_r$, where P_r is the optical power signal of the receiver. We assume a multimodal counting model with background additive noise count rate of nb . If the transmitting field is in the receiving field within T seconds, the average signal count $ks = nsT$ is generated. Thus, we can obtain the capture probability [2]:

$$\text{PAC} = \frac{\Gamma(K_T, K_s + K_b)}{\Gamma(K_T, \infty)}, \quad (8.4)$$

where $K_b = nbT$, K_t is the threshold, and $\Gamma(a, b)$ is a gamma function:

$$\Gamma(a, b) = \int_0^b e^{-t} t^{a-1} dt. \quad (8.5)$$

The error acquisition probability threshold is

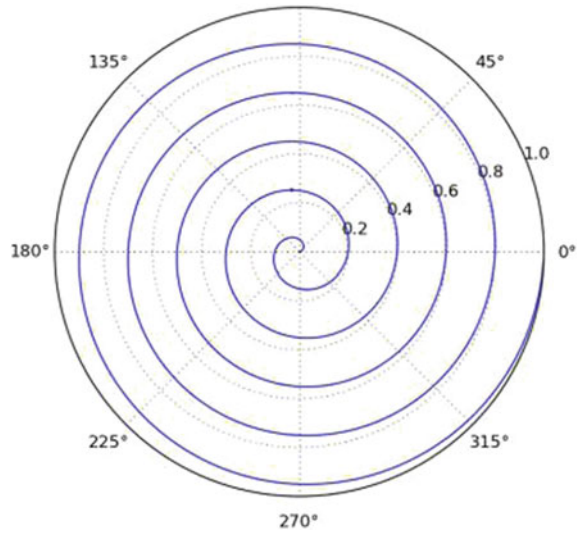
$$\text{PFC} = \frac{\Gamma(K_T, K_b)}{\Gamma(K_T, \infty)}. \quad (8.6)$$

The signal and noise counts are given as

$$K_s = n_s T, \quad (8.7)$$

$$K_b = n_b T = n_{b0} D_{sr} T, \quad (8.8)$$

where n_{b0} is the noise count rate of each spatial mode, and D_{sr} is the spatial modulus in the resolution field of view Ω_r . The time of the transmitter in the field of view is T , and the azimuth rotation rate is S_L (rad/s), which are connected by Eq. (8.9).

Fig. 8.6 Spiral scanning

$$T = \sqrt{(\Omega_r / \Omega_L)}. \quad (8.9)$$

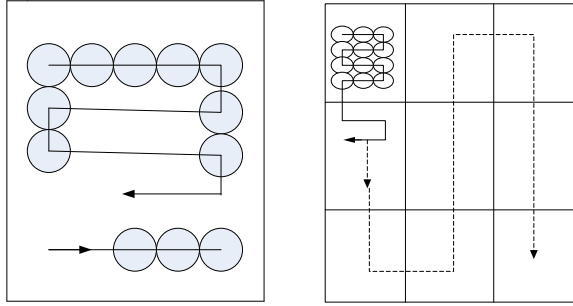
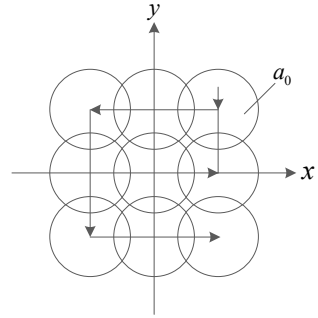
Theoretically, Ω_r should be wide enough to cover a single space module. To obtain the required K_s , Ω_r must be reduced. Therefore, the scanning will be slower and the acquisition will take longer.

(1) Spiral scanning

The spiral scanning method is shown in Fig. 8.6. This type of scanning starts from the central point, which has the highest probability in position, and gradually scans to the peripheral low probability area, forming a radial scanning mode. Owing to the constant angular velocity of the original scanning mode, the linear velocity is small when it is located near the center point. However, with the increase of scanning radius, the linear velocity gradually increases, which increases the circle spacing, which, in turn, increases the probability of a missed scanning. However, when the distance between the receiver and the transmitter is known, uniform and equidistant scanning lines can be formed in the plane of the receiving end by compensating the diagonal velocity, as shown in Fig. 8.6. The acquisition time of spiral scanning is less than that of rectangular scanning, but the driving current control is more complex.

(2) Raster scanning

There are two types of rectangular scanning, as shown in Fig. 8.7. If the detector has no travel limit and can scan the entire search range, the first method is adopted. If the detector has a travel limit and can only scan in a small range, it requires a slow scanner that can cover the entire search range to drive it. The function of the slow scanner can be performed by a telescope or tracking frame. Although the first method is simpler than the second method, the second method is often used in practical systems due to the limitation to detector travel.

Fig. 8.7 Raster scanning [2]**Fig. 8.8** Spiral grating scanning [2]

(3) Spiral grating scanning

The combination of grating scanning and spiral scanning is called spiral grating scanning, as shown in Fig. 8.8. This scanning mode combines the advantages of grating scanning with the advantages of scanning from high to low probability areas, similar to spiral scanning.

8.2.2.2 Focal Plane Scanning

The results of focal plane scanning and antenna scanning are the same, but there is a difference in the scanning mechanism. As shown in Fig. 8.9, a fixed optical lens transforms the uncertain field of view to the focal plane, scanning the focal plane with a single detector. In the focal plane scanning method, the receiving structure can be fixed, and a mechanically movable part is not required.

The acquisition probability of focal plane array scanning is [2]

$$PAC_1 = \sum_{k_1=0}^{\infty} P_{os}(k_1, K_s + K_b) \left[\sum_{k_1=0}^{k_1-1} P_{os}(k_2, K_b) \right], \quad (8.10)$$

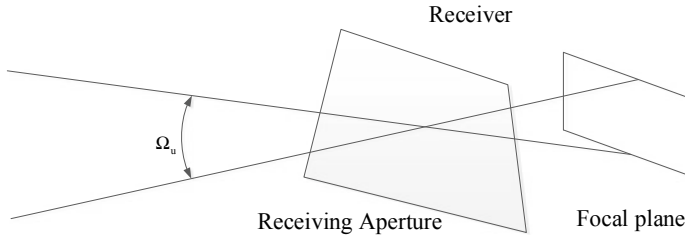


Fig. 8.9 Schematic diagram of a focal plane scanning mode with Ω_u the search work for the angle of the uncertain area

where K_s and K_b are given by Eqs. (8.7) and (8.8). Without losing generality, we use D_{su}/Q instead of D_{sr} ; D_{sr} is the spatial modulus, and Ω_r is the resolution field of view. The entire uncertain region can be divided into Q nonoverlapping corner Ω_r sub-regions; n_{bu} is defined as the noise count rate within the total solid angle Ω_u . Therefore,

$$K_s = n_s T, \quad (8.11)$$

$$K_b = \frac{n_{bu} T}{Q}. \quad (8.12)$$

The total focal plane search time is represented by T_t , and

$$T_t = QT, \quad (8.13)$$

where T is the time at each resolution position. When Q and the power level of the transmitter and background are given, the acquisition probability PCA_1 is related to the total search time T_t and observation time T . If T is not sufficiently long, the acquisition probability suddenly decreases.

8.2.2.3 Focal Plane Array Scanning

A fixed detector array is used to cover the focal plane. It enables the use of a parallel processing method, which can shorten the acquisition time. Each detector in the detector array processes a specific part of the uncertainty domain, as shown in Fig. 8.10. After the observation time is determined, the output of each detector can be collected to calculate and compare the transmitter positions. Because each detector works independently, the complexity of the receiver increases, but the acquisition time is shortened.

Considering an array with S detectors, the uncertainty angle can be divided into Ω_u/S resolution areas, and the focal plane is decomposed into S small areas:

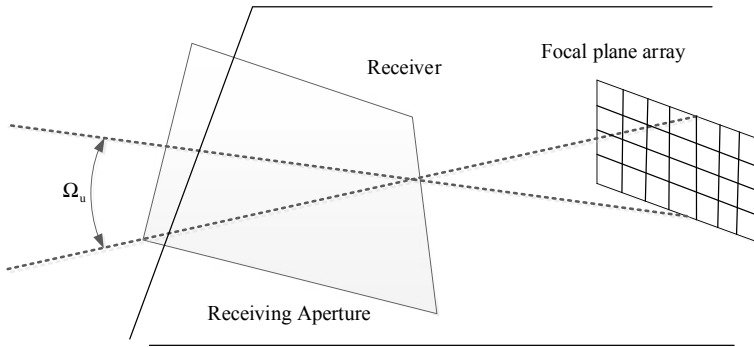


Fig. 8.10 Schematic diagram of a focal plane array scanning

$$\Omega_r = \frac{\Omega_u}{S}, \quad (8.14)$$

$$T_t = T. \quad (8.15)$$

To obtain a higher resolution, a larger array must be chosen, which will increase the complexity of the receiver. In engineering, a reasonable choice between the resolution and the search time must be made.

8.2.2.4 Sequential Scanning

If the number of detectors is large, parallel processing becomes very difficult. Sequential scanning can be used to find the maximum output at each time. Then, the acquisition time is slightly increased, but the receiver is considerably simplified. For a given observation time of T seconds, the array has S detectors. It takes ST seconds to complete a single scan and r iterations to complete all scans [2].

$$T_t = rST = \left(\frac{S}{\lg S} \right) T \lg \left(\frac{\Omega_u}{\Omega_r} \right) \quad (8.16)$$

For a square matrix, S must be the square of an integer to achieve the shortest search time. It can be proved that $S = 4$ is the optimal array for the shortest search time.

8.2.3 Performance of Acquisition

For the fast high-probability acquisition unit in space laser communication, the performance of each link is analyzed as follows:

(1) Pointing error of initial LOS axis

As shown in Fig. 8.2, the controller computer drives the coarse tracking turntable to point to the uncertain region according to the guidance information. The LOS axis of the communication transmitter is the initial LOS axis. The pointing error of the initial LOS axis is the uncertainty region of acquisition, which can be determined by [1]:

$$\sigma = \sqrt{\sigma_A^2 + \sigma_P^2 + \sigma_R^2 + \sigma_C^2 + \sigma_G^2 + \sigma_E^2}, \quad (8.17)$$

where σ_A is the platform attitude error, $\sigma_A = \sqrt{\sigma_{ap}^2 + \sigma_{ay}^2 + \sigma_{ah}^2}$, and σ_{ap} , σ_{ay} , and σ_{ah} are the azimuth, pitch, and yaw attitude measurement errors, respectively. Moreover, σ_P is the positioning error, $\sigma_P = \sqrt{\sigma_{px}^2 + \sigma_{py}^2 + \sigma_{pz}^2}$, and σ_{px} , σ_{py} , and σ_{pz} are the latitude, longitude, and elevation measurement errors, respectively. Finally, σ_R is the antenna installation error, σ_C is the computer calculation error, σ_G is the mechanical motion error of the servo turntable, and σ_E are other small errors.

(2) Relationship between the pointing error of initial LOS axis and uncertainty region of acquisition

The pointing error of initial LOS axis can be regarded as a random variable of Gaussian distribution with zero means in pitch and azimuth. Considering the distribution of the initial error at both positive and negative directions, the uncertainty region of acquisition should satisfy the following conditions:

$$\frac{\text{FOU}}{2} \geq 3\sigma. \quad (8.18)$$

If $\sigma_v = \sigma_h = \sigma$, then the probability of the target appearing in the uncertain region is

$$P_{\text{cov}} = \iint_{\text{FOU}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\theta_v^2}{2\sigma^2}\right) \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\theta_h^2}{2\sigma^2}\right) d\theta_v d\theta_h, \quad (8.19)$$

where θ_v and θ_h are the pitch and azimuth deviation angle, respectively. Equation (8.19) can be simplified as Rayleigh distribution of amplitude and uniform distribution of polar angle (1/2):

$$P_U = \int_0^{\frac{\text{FOU}}{2}} \frac{\theta}{\sigma^2} \exp\left(-\frac{\theta^2}{2\sigma^2}\right) d\theta = 1 - \exp\left(-\frac{\text{FOU}^2}{8\sigma^2}\right), \quad (8.20)$$

where $\theta = \sqrt{\theta_v^2 + \theta_h^2}$ is the angular deviation amplitude of the uncertain angle.

8.3 Automatic Tracking

8.3.1 Tracking System

Once the beacon light from the other side is detected in the acquisition field of view, both sides form a light tracking closed loop, and enter the coarse tracking stage. The presence of the spot is detected by the acquisition sensor, and the miss distance can be calculated. After servo compensation, the servo turntable is driven to move. As shown in Fig. 8.11, the implementation structure of coarse tracking is that the two-axis servo turntable drives the entire movement of the telescope unit. Owing to the limitations of the motor power, mechanical resonance frequency, and other factors, the coarse tracking servo bandwidth is limited, and the coarse tracking accuracy is not high. Therefore, the purpose of coarse tracking is to reliably enter the field of view of the precise tracking.

Let (θ_z, θ_e) be the visual vector line to the transmitter. As shown in Fig. 8.12, if (ϕ_z, ϕ_e) is the normal vector angle of the corresponding receiver plane, then the instantaneous angle error from the aiming receiver to the transmitter is [2]

$$\psi_z(t) = \theta_z - \phi_z, \quad (8.21)$$

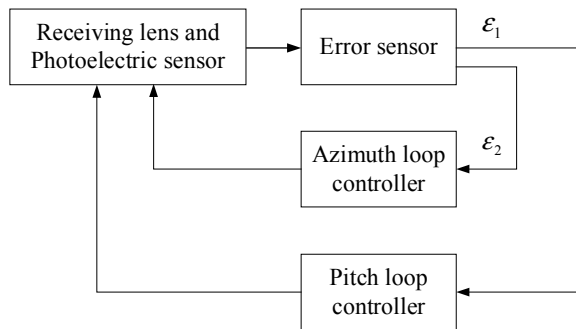
$$\psi_e(t) = \theta_e - \phi_e. \quad (8.22)$$

$\varepsilon_z(t)$ and $\varepsilon_e(t)$ are the error voltage generated by the optical sensor, control azimuth angle and pitch angle, respectively, used to correct the aiming angle, (θ_z, θ_e) . Thus,

$$\phi_z = \bar{\varepsilon}_z(t), \quad (8.23)$$

$$\phi_e = \bar{\varepsilon}_e(t). \quad (8.24)$$

Fig. 8.11 Structure of a tracking system in space [2]



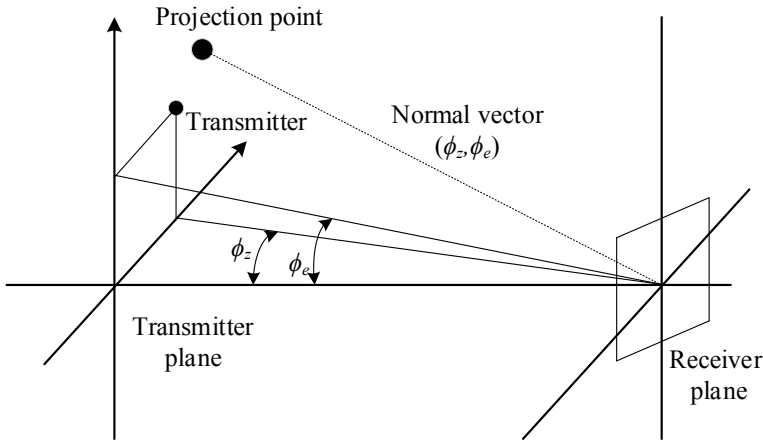


Fig. 8.12 Azimuth and pitch angle aiming [2]

The horizontal lines in Eqs. (8.23) and (8.24) represent the filtering effect of the control loop. By combining the formulas, we obtain the following results.

$$\begin{cases} \frac{d\psi_z}{dt} = \frac{d\theta_z}{dt} - \frac{d}{dt}[\bar{\varepsilon}_z(t)] \\ \frac{d\psi_e}{dt} = \frac{d\theta_e}{dt} - \frac{d}{dt}[\bar{\varepsilon}_e(t)] \end{cases} \quad (8.25)$$

8.3.2 Compound-Axis Control System

It is very difficult for an optical antenna to track quickly, accurately, and directly. Adopt a rotating frame structure, large aperture telescopes has a limited field of view of the rotating mirror. If the rotating frame and mirror are combined, a compound-axis system is formed. The compound-axis system comprises a mirror, which can move in pitch and azimuth, on the frame of large inertia to control the direction of transmitting and receiving optical axis. This mirror is called fast steering mirror, the axis of the mirror is called the sub-axis, and the axis of the main tracking frame is called the principal axis. The principal and sub-axis can be controlled separately to form the main and sub-system, respectively. The main system is working with a large range, but narrow bandwidth and low accuracy. The sub-system is working with a small range, but wide bandwidth, fast responsibility, and high precision. By adding the functions of the two systems, a fast and high-precision tracking in a wide range can be realized. Its structure is shown in Fig. 8.13.

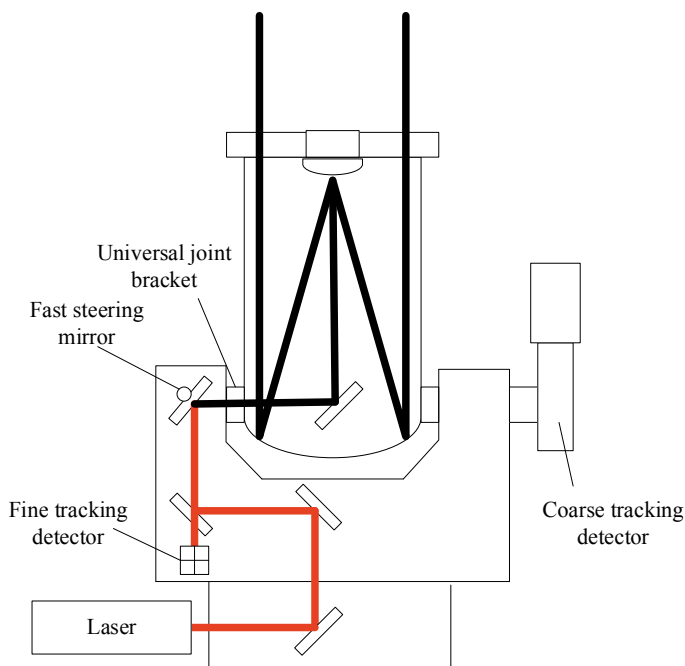


Fig. 8.13 Structure of compound-axis control system [3]

8.3.3 Accuracy of a Coarse Tracking Unit

The tracking accuracy of a coarse tracking unit is not only dependent on the characteristics of the system itself but also related to the characteristics of external excitation and noise sources. At the same time, the coarse tracking error is determined according to the tracking field of view, and it can be divided as follows. (1) Dynamic lag error: due to the relative motion between the two optical terminals, the angular velocity and angular acceleration of the line of sight will cause dynamic lag error for the position servo unit in the tracking process. (2) Platform vibration residual error: the satellite platform has strong random vibrations, which act on the LOS through the two-axis constant frame servo turntable. This will cause the random jitter of LOS, which is the main noise source affecting the tracking accuracy. (3) CCD spot centroid detection error: due to the difference of spot size, spot power spatial distribution, image signal-to-noise ratio, and spot detection algorithm, there will be errors in spot centroid detection. (4) Various torque interference errors: these include line disturbance torque, imbalance torque, and friction torque errors. The characteristics of these interference torque errors are different, and in engineering, they are treated approximately as Gaussian distribution. Overall, the total error of the coarse tracking servo unit is [1]

$$\delta_{CT} = \delta_1 + \sqrt{\delta_2^2 + \delta_3^2 + \delta_4^2}. \quad (8.26)$$

The error of CCD spot detection is randomly distributed, and the control error caused by it is expressed as [1]

$$\sigma_3^2 = \int_{-\infty}^{\infty} \left| \frac{G(j\omega)}{1 + G(j\omega)} \right| \phi_{ccd}(\omega) d\omega. \quad (8.27)$$

8.3.4 Fine Tracking Unit

The working principle of a galvanometer is that the actuator produces angular or linear displacement, and the mirror attached to the actuator is driven to realize the two-dimensional angular deflection. There are two common drives [4, 5]:

- (1) Electromagnetic galvanometer: four drivers are usually integrated into the electromagnetic galvanometer to form a two-dimensional push–pull operation. The driver is a voice coil motor: a type of DC motor, which converts electrical signal into linear displacement. Its working principle is that an electrified coil (conductor) placed in a magnetic field produces a force, the magnitude of which is proportional to the current. It has the characteristics of small volume, simple structure, high speed, high acceleration, and corresponding speed. According to the configuration of the driver, feedback, controller, and control algorithm, voice coil motor can reach 500–1000 Hz motion frequency, or even higher. The control accuracy can reach tens of nanometers.
- (2) Piezoelectric ceramic galvanometer: piezoelectric/electrostrictive micro displacement actuator is a type of energy exchange element that directly converts electrical energy into mechanical energy and produces micro displacement by using the inverse piezoelectric or electrostrictive effect of dielectric in an electric field. Currently, it is an ideal driving element in micro displacement technology. At present, PZT galvanometer is used in most space laser communication systems.

8.4 Fast Alignment Using Two-Dimensional Mirror

8.4.1 Introduction

Optical wireless communication has the advantages of high speed, good confidentiality and strong anti-interference ability [6]. It has gradually become the best way of

massive data transmission [7, 8]. The premise of establishing optical wireless communication is to maintain accurate alignment and real-time tracking of the transceiver visual axis of the optical transceiver on both sides of the communication [9]. Due to the characteristics of narrow laser beam and strong directivity, the signal beam requires extremely accurate pointing. When the laser signal is transmitted in the atmospheric channel, the atmospheric refractive index fluctuation caused by atmospheric turbulence leads to the beam expansion, beam drift, wavefront distortion, fluctuation of angle of arrival, and other phenomena, which have adverse impacts on the reception of optical signal, and even interrupt the communication link [10]. In optical wireless communication systems, three acquisition methods, gaze-gaze, gaze-scan, and skip-scan, are used to achieve coaxial alignment of the beam between the transmitting and receiving antennas [11]. This acquisition process causes significantly system delay in communication. In practice, a fast acquisition, pointing and tracking mechanism is required to establish a communication link [12, 13].

The compound-axis APT system scans the beam in a rectangular or spiral shape [14], where the receiver returns the data [15] through satellite positioning system and collects the image by the imaging device [16], so as to acquire the beam. A combination of a low bandwidth coarse tracking unit with a large field of view and high bandwidth fine tracking unit with a small range is used. Here, the coarse tracking unit makes the beacon spot reliably enter the fine tracking field of view [17], and the fine tracking unit further suppresses the residual error [18]. Using position sensitive sensor, four-quadrant detector [19, 20], charge coupled device [21–23] and other position feedback devices, Bao et al. proposed a flat field model and correction method based on the pixel response spot scale area of image sensor [24]. Zhang et al. proposed a point position detection method using four-quadrant detector combined with piecewise low-order polynomial least square fitting and Kalman filter [25]. Chen Gang et al. adopted the position sensitive detector to track and control the beam and applied them in laser communication system [26]. Nielson et al. adopts a double charge coupled device detection scheme with separated coarse and fine detection in design of the APT system in the terminal of optical wireless satellite communication system [27, 28].

As the actuator in the tracking system, the two-dimensional mirror is first used in adaptive optics system to compensate the wavefront tilt distortion [29], and then applied to the fields of beam scanning, beam positioning and target tracking. Two-dimensional mirror driven by piezoelectric ceramics has the characteristics of small scanning angle and high resonant frequency, but inherent defects of the delay compensation and creeping [30]. The two-dimensional mirror driven by voice coil motor has the characteristics of large scanning angle and resonant frequency [31]. Zuo et al. adopted the piezoelectric two-dimensional mirror to realize the fine tracking of 2.3 km beam by self-tuning control [32]. Toyoda et al. applied the composite-axis control technology with a nested coarse tracking and fine tracking units to the APT system of laser communication equipment, where the coarse tracking adopted a two-dimensional turntable structure, and the fine tracking adopted the piezoelectric two-dimensional mirror [33]. Mike et al. realized an accurate and stable LOS alignment of 50 km between the laser transmitter and optical detector, where a position

sensitive sensor and piezoelectric two-dimensional mirror were used to complete the tracking with a wide field of view [34]. To realize the beam tracking, Suite et al. installed a two-dimensional mirror laser communication test equipment controlled by the position sensitive sensor on the receiver [35].

To overcome the long alignment time of traditional APT system, this section introduces a method to realize fast beam alignment by using image calibration at the transmitter and two-dimensional mirror at the receiver. The beam scanning is realized by changing the pitch and azimuth of the two-dimensional steering mirror, and beam tracking is realized by feedback of the position information of the focal plane spot. Field experiments of 1.3 km and 10.3 km are performed respectively based on the IM/DD system.

8.4.2 Theoretical Model

The coaxial alignment of the optical wireless communication system is shown in Fig. 8.14a, which requires that the optical axes of the transmitting and receiving antenna completely coincide in space. Due to the beam drift caused by atmospheric

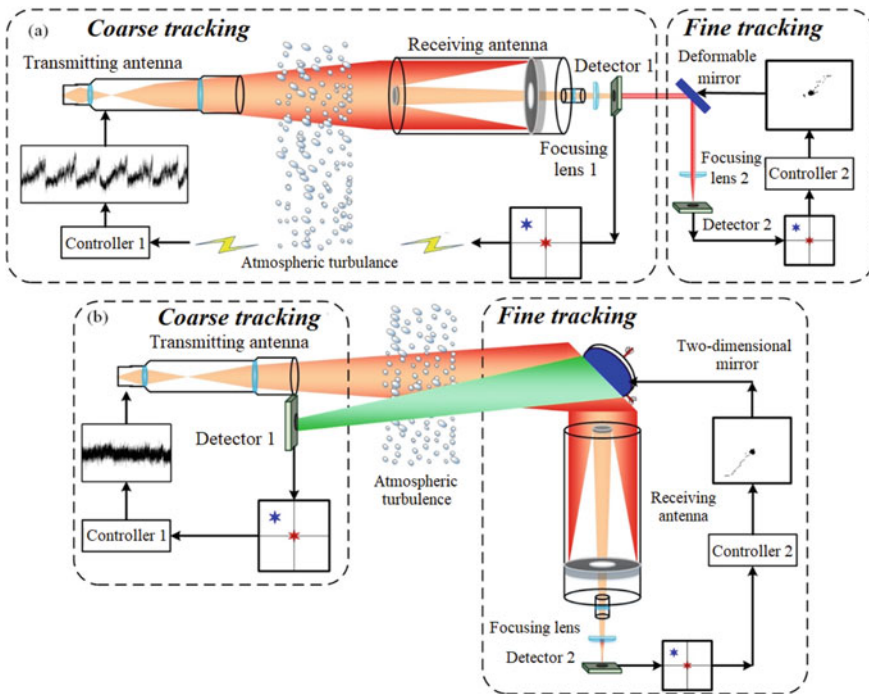


Fig. 8.14 Beam alignment diagram of optical wireless communication **a** coaxial alignment **b** two-dimensional mirror assisted alignment

turbulence, the transmit antenna should be adjusted according to the feedback from the detector to keep the beam in stable coarse alignment for a long time, as well as the fine alignment after that. Since the long-range feedback and position adjustment are affected by atmospheric turbulence, the traditional long-axis beam alignment process remains significant uncertainty.

Figure 8.14b shows an optical wireless communication system using two-dimensional mirror assisted alignment. The system consists of a coarse tracking from the transmitting antenna to the two-dimensional mirror and a fine tracking from the two-dimensional mirror to the receiving antenna. The calibration coarse alignment directly performs positioning and calibration using the calibration camera at the transmitter, and adjusts the transmitting antenna according to the captured image as the feedback. The fine alignment of the minor axis adjusts the two-dimensional mirror according to the feedback information from the detector at the back end of the receiving antenna. Since the calibration coarse alignment and minor axis fine alignment can be operated at a single end, and the calibration alignment at the transmitter does not require data feedback, this alignment method is convenient and less affected by the environment.

To investigate the scanning track of light on the surface after the two-dimensional mirror, a spatial coordinate system is established as shown in Fig. 8.15 [36]. Supposing that the vector of the incident light is \mathbf{A} , the angle between the outgoing light and normal of the two-dimensional mirror is α , a spatial coordinate system (x', y', z') of the incident light is established with the vector direction of the incident light as the x' axis. Denoting \mathbf{A}' the vector of the outgoing light, a spatial coordinate system (x, y, z) is established by taking the direction of the outgoing light as the z axis. Assume that the normal vector of the mirror is \mathbf{N} , the deflection angles of the two-dimensional reflector in the transverse and longitudinal directions are θ_1 and θ_2 , respectively.

Taking (x, y, z) as the principal coordinate system, the vector expressions for \mathbf{A} and \mathbf{N} are

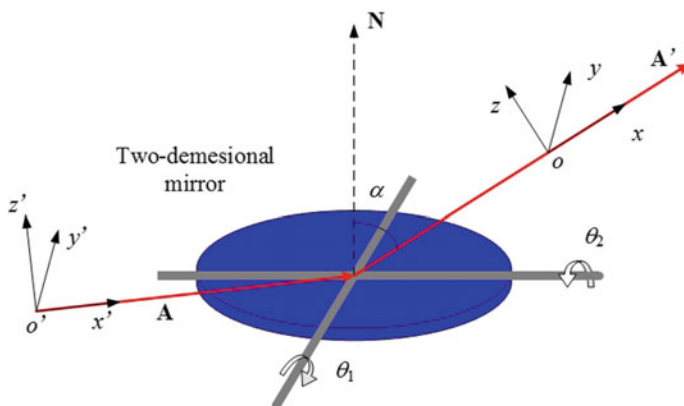


Fig. 8.15 Geometrical optical model of two-dimensional mirror

$$\mathbf{A} = \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix} = \begin{bmatrix} -\sin(2\alpha) \\ 0 \\ -\cos(2\alpha) \end{bmatrix}, \quad (8.28)$$

$$\mathbf{N} = \begin{bmatrix} N_x \\ N_y \\ N_z \end{bmatrix} = \begin{bmatrix} \sin(\alpha + \theta_1) \cdot \cos \theta_2 \\ \sin \theta_2 \\ \cos(\alpha + \theta_1) \cdot \cos \theta_2 \end{bmatrix}. \quad (8.29)$$

According to the reflection law, the angles between \mathbf{A} , \mathbf{A}' , and \mathbf{N} are equal, and

$$\mathbf{A}' = \mathbf{A} - 2(\mathbf{A} \cdot \mathbf{N}) \cdot \mathbf{N} = \mathbf{M}\mathbf{A}, \quad (8.30)$$

$$\mathbf{N}(\theta) = \mathbf{N} \cdot \cos \theta + \mathbf{C}(\mathbf{C} \cdot \mathbf{N}) \cdot (1 - \cos \theta) + (\mathbf{C} \times \mathbf{N}) \cdot \sin \theta. \quad (8.31)$$

where \mathbf{C} is the unit direction vector corresponding to any rotation axis. Using Eqs. (8.28), (8.29) and (8.31) in Eq. (8.30), the matrix representation of \mathbf{M} is obtained as

$$\mathbf{M} = \begin{bmatrix} 1 - 2N_x^2 & -2N_xN_y & -2N_xN_z \\ -2N_xN_y & 1 - 2N_y^2 & -2N_yN_z \\ -2N_xN_z & -2N_yN_z & 1 - 2N_z^2 \end{bmatrix}. \quad (8.32)$$

Using Eqs. (8.28) and (8.29) in Eq. (8.32), the matrix form of incident light can be obtained as

$$\begin{aligned} \mathbf{A}' &= \begin{bmatrix} A'_x \\ A'_y \\ A'_z \end{bmatrix} \\ &= \begin{bmatrix} \sin(2\alpha) \cdot (2 \cdot \cos^2 \theta_2 \cdot \sin^2(\alpha + \theta_1) - 1) + \cos(2\alpha) \cdot \cos^2 \theta_2 \cdot \sin(2\alpha + 2\theta_1) \\ \cos(2\alpha) \cdot \cos(\alpha + \theta_1) \cdot \sin(2\theta_2) + \sin(2\alpha) \cdot \sin(\alpha + \theta_1) \cdot \sin(2\theta_2) \\ \cos(2\alpha) \cdot (2 \cdot \cos^2 \theta_2 \cdot \cos^2(\alpha + \theta_1) - 1) + \sin(2\alpha) \cdot \cos^2 \theta_2 \cdot \sin(2\alpha + 2\theta_1) \end{bmatrix}. \end{aligned} \quad (8.33)$$

The coordinates of the scanning spot in xoy plane are

$$\begin{aligned} x &= L \frac{A'_x}{A'_z} \\ &= L \frac{\sin(2\alpha) \cdot (2 \cdot \cos^2 \theta_2 \cdot \sin^2(\alpha + \theta_1) - 1) + \cos(2\alpha) \cdot \cos^2 \theta_2 \cdot \sin(2\alpha + 2\theta_1)}{\cos(2\alpha) \cdot (2 \cdot \cos^2 \theta_2 \cdot \cos^2(\alpha + \theta_1) - 1) + \sin(2\alpha) \cdot \cos^2 \theta_2 \cdot \sin(2\alpha + 2\theta_1)}, \end{aligned} \quad (8.34)$$

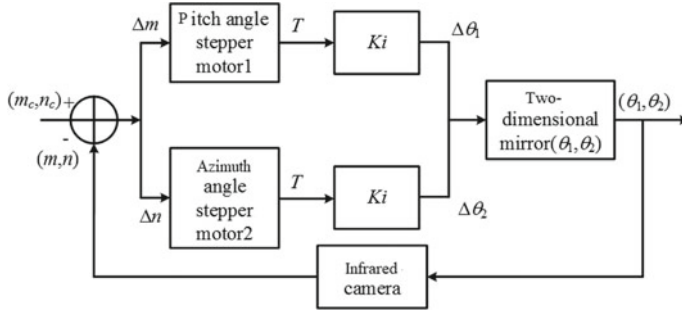


Fig. 8.16 Laser tracking algorithm block diagram using two-dimensional mirror

$$\begin{aligned}
 y &= L \frac{A'_x}{A'_z} \\
 &= L \frac{\cos(2\alpha) \cdot \cos(\alpha + \theta_1) \cdot \sin(2\theta_2) + \sin(2\alpha) \cdot \sin(\alpha + \theta_1) \cdot \sin(2\theta_2)}{\cos(2\alpha) \cdot (2 \cdot \cos^2 \theta_2 \cdot \cos^2(\alpha + \theta_1) - 1) + \sin(2\alpha) \cdot \cos^2 \theta_2 \cdot \sin(2\alpha + 2\theta_1)}.
 \end{aligned} \quad (8.35)$$

where L is the linear distance from the point of reflection of the beam to xoy plane.

Figure 8.16 is a schematic diagram of the laser tracking algorithm using a two-dimensional mirror. The infrared camera obtains the position of the light spot, converts the position into the adjustment angle of the two-dimensional mirror through calculation, and realizes the beam tracking and control by adjusting the two-dimensional mirror. Supposing that the position of the center point of the detector is (M_C, N_C) , and the imaging spot position of the infrared camera is $(m(k), n(k))$, the angles of pitch θ_1 and azimuth θ_2 applied to the two-dimensional mirror at $(k + 1)$ -th time are

$$\begin{bmatrix} \theta_1(k+1) \\ \theta_2(k+1) \end{bmatrix} = \begin{bmatrix} m(k) - m_c \\ n(k) - n_c \end{bmatrix} \cdot T \cdot K_i + \begin{bmatrix} \theta_1(k) \\ \theta_2(k) \end{bmatrix}, \quad (8.36)$$

where T is the unit conversion coefficient from motor angle to pixel displacement and K_i is the integral constant.

After the beam tracking is completed, the intensity modulated optical signal with information is coupled into the photodetector through focusing lens. Denoting V_π the half-wave voltage of the intensity modulator, the intensity modulation and direct detection of the optical signal can be realized by adjusting the DC bias voltage of the intensity modulator V_{bias} ,

$$V_{out}(t) = \left[\eta \cdot \int_s \left[E_{in}(t) \cos \left[\frac{\pi}{2V_\pi} (V_m(t) - V_{bias}) \right] e^{j \frac{\pi V_{bias}}{2V_\pi}} \right]^2 ds \right]^2 R, \quad (8.37)$$

where $V_m(t)$ is the source signal loaded into the modulator, $E_{in}(t)$ is the input optical signal of the modulator, s is the effective area of the detector, η is the responsibility of the detector, R is output impedance, and $V_{out}(t)$ is the output voltage signal of the detector.

8.4.3 Experiments

8.4.3.1 System Structure

Based on the theoretical model, an optical wireless communication IM/DD system using two-dimensional mirror is established to realize fast alignment, as shown in Fig. 8.17. At the transmitter, the source information is loaded to the light wave through external intensity modulation, amplified by optical fiber amplifier, and transmitted through the optical antenna after collimation. At the receiver, the initial calibration of the two-dimensional mirror is performed by the alignment platform and the calibration camera at the transmitter. This ensures the two-dimensional mirror is completely covered by the beam, and the coaxial alignment of the reflected beam with the receiving antenna is achieved by adjusting the two-dimensional mirror. The parallel light emitted from the receiving antenna passes through the optical prism and is divided into two beams. One beam is converged by the focusing lens and coupled into the photosensitive surface of the photodetector for system communication. The other beam is focused and used by the infrared camera to detect the spot position in real time to complete the beam tracking. Table 8.1 shows the experimental equipment and parameters.

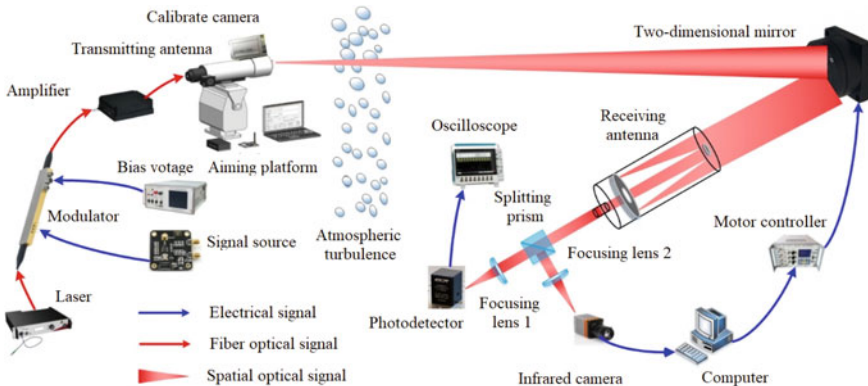


Fig. 8.17 Optical wireless communication IM/DD system with fast alignment of two-dimensional mirror

Table 8.1 Experimental equipment and parameters

Experimental equipment	Main parameter
Modulator	X-cut crystal Radio frequency drive voltage: $V_m = 4.5\text{ V}$ Half-wave voltage: $V = 5.5\text{ V}$
Antenna	Transmitting antenna: Cassegrain, Diameter = 105 mm Receiving antenna: Cassegrain, Diameter = 220 mm
Calibrate camera	Digital zoom up to $\times 128$
Two-dimensional mirror	Diameter = 280 mm Resonant frequency: 20 Hz Resolution: $2.73\text{ }\mu\text{rad}$
Infrared camera	Detector type: InGaAs Wavelength range: 0.9 mm ~ 1.7 mm Pixel resolution: 320×240 Pixel size: 1 pixel = 20 mm
Photo detector	Detector type: InGaAs Cut-off frequency: 30 kHz–1.5 GHz Effective area diameter: 400 mm

8.4.3.2 Field Experiments for a 1.3 km Communication Link

The IM/DD-based optical wireless communication fast alignment at a distance of 1.3 km field experiment is shown in Fig. 8.18, where the transmitter and receiver

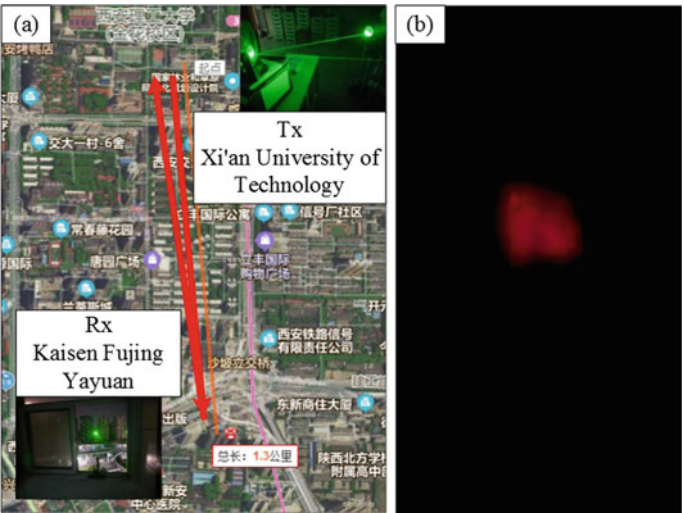


Fig. 8.18 Experiment of optical wireless communication at a distance of 1.3 km: **a** link diagram **b** far field spot

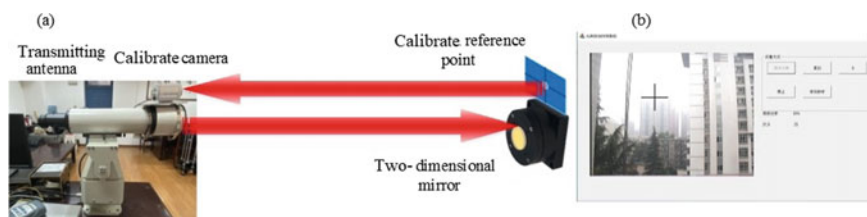


Fig. 8.19 Beam fast calibration: **a** structure diagram **b** calibrate camera interface

are located at Kaisen Fujingjiayuan on the South Second Ring Rd. of Xi'an and the Building #6 of Xi'an University of Technology, respectively.

Figure 8.19a is a structural diagram of fast calibration and tracking using a calibration camera. The calibration camera and transmitting antenna at the transmitting end are parallel coaxial optical systems during installation and adjustment. By adjusting the pitch and azimuth angles of the servo motor and the zoom system of the calibration camera, gradually find the reference coordinate position of the two-dimensional mirror at the receiving end. When the camera center coincides with the reference coordinate position of the two-dimensional mirror, according to the principle of optical path reversibility, that is, the beam is considered to have completed the coarse alignment of the long axis. Figure 8.19b is the upper computer interface where the imaging position center of the calibrated camera coincides with the reference coordinate position of the receiving end after adjustment.

After coarse alignment, the beam is roughly tracked according to the relative deviation between the calibration point and the imaging coordinate position of the two-dimensional mirror. Figure 8.20 shows the beam coarse tracking curve of the position coordinate of the two-dimensional mirror with the 1.3 km experimental link, in which the pitch angle is adjusted twice in 4 h and the azimuth angle is adjusted once in 4 h. The beam drift caused by atmospheric turbulence and the gravity subsidence and mechanical vibration of the optical mechanical structure make the coarse alignment unnecessary to be adjusted frequently in a short time.

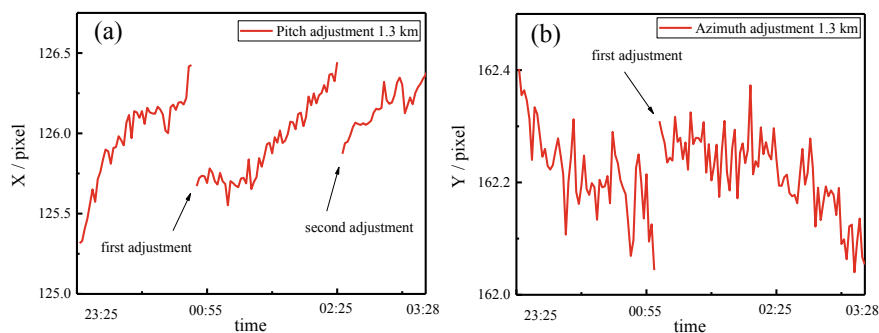


Fig. 8.20 Calibrate camera tracking curve at 1.3 km **a** pitch adjustment **b** azimuth adjustment

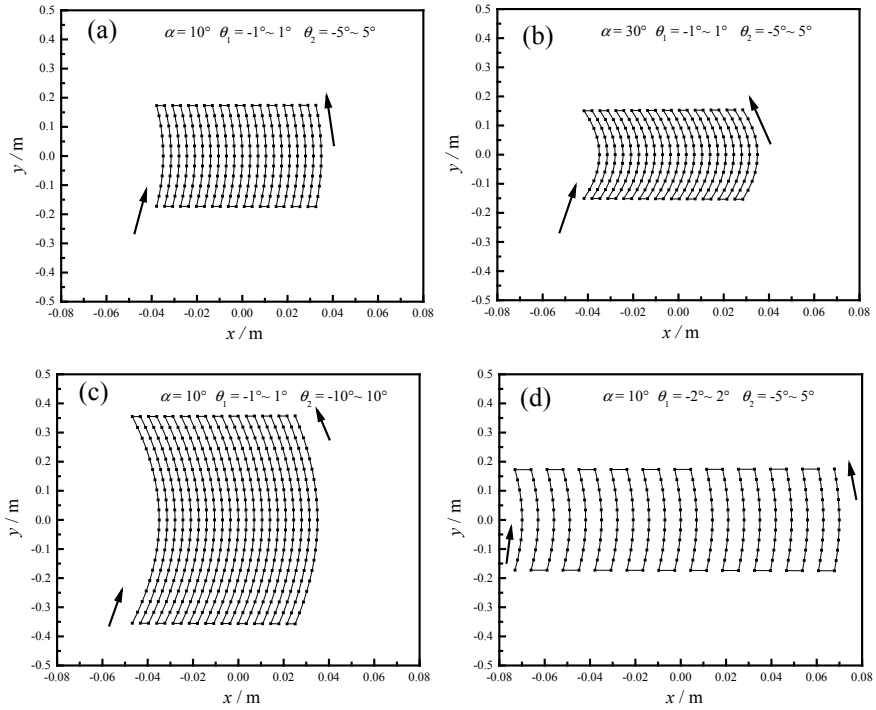


Fig. 8.21 Trajectory of outgoing spot under different incident angle and scanning angle **a** $\alpha = 10^\circ$ $\theta_1 = -1^\circ \sim 1^\circ$ $\theta_2 = -5^\circ \sim 5^\circ$ **b** $\alpha = 30^\circ$ $\theta_1 = -1^\circ \sim 1^\circ$ $\theta_2 = -5^\circ \sim 5^\circ$ **c** $\alpha = 10^\circ$ $\theta_1 = -1^\circ \sim 1^\circ$ $\theta_2 = -10^\circ \sim 10^\circ$ **d** $\alpha = 10^\circ$ $\theta_1 = -2^\circ \sim 2^\circ$ $\theta_2 = -5^\circ \sim 5^\circ$

Based on Eq. (8.34), the distance L from the center of the calibrated two-dimensional reflector to the detector surface is set to 1, and the trajectories depicted by the beam at the receiving surface are calculated for different incidence angles and different scanning angles θ_1 and θ_2 after adjustment, respectively. Figure 8.21 shows the scanning trajectory of the reflected beam under different parameters, and the scanning trajectory curve presents a cross-sectional diagram based on an arc shape, where the incident angle determines the arc of the scanning trajectory and the scanning angles θ_1 and θ_2 determine the scanning area of the scanning plane. Based on the incident angle and the distance L from the center of the two-dimensional reflector to the detector surface, a suitable range of scanning angle θ_1 and θ_2 scanning step can be selected to achieve fast spot capture in a short time.

When the spot is located at the center of the infrared camera, the mean variance and power spectral density (PSD) estimates of the change of the spot center in the X direction and Y direction are calculated respectively. The mean drift of the spot center in X direction is 161.7439 pixel, the variance is 12.5734 pixel, and the mean drift of the spot center in Y direction is 121.3437 pixel, the variance is 18.8653 pixel. Figure 8.22 shows the power spectral density estimation of the drift of the spot

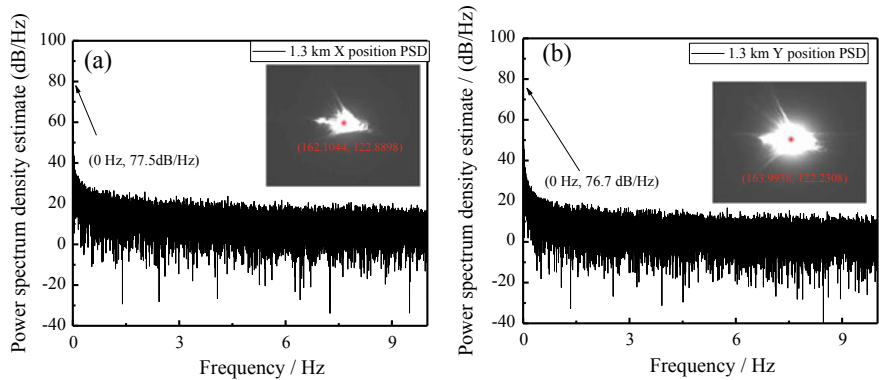


Fig. 8.22 Power spectral density estimation of spot center drift at a distance of 1.3 km **a** X position **b** Y position

center in *X* and *Y* directions. The power spectral component is mainly low-frequency component, indicating that the drift of the spot center is slow.

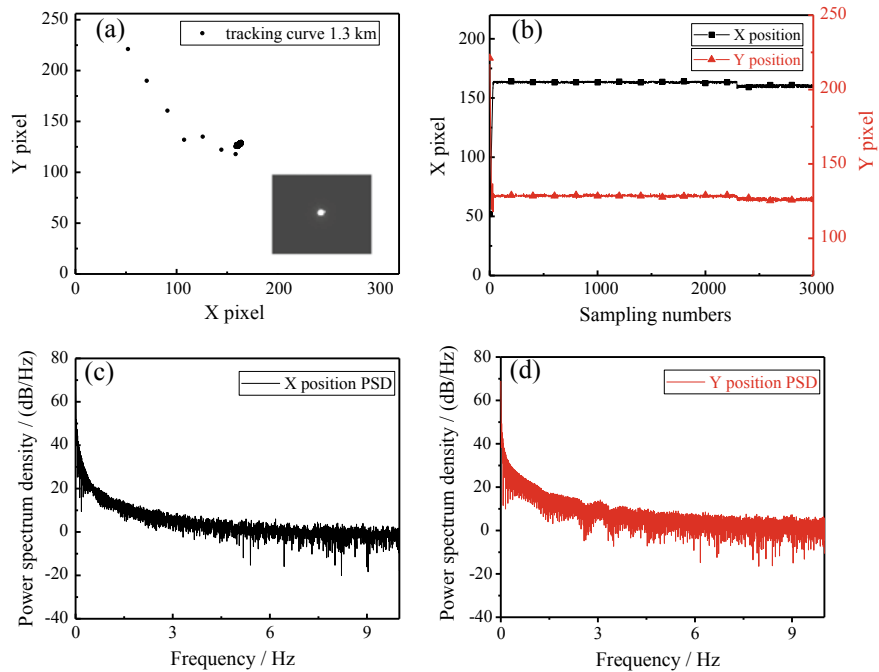


Fig. 8.23 Tracking waveforms for 1.3 km link **a** tracking curve **b** track curve in X and Y positions **c** X position power spectral density estimation **d** Y position power spectral density estimation

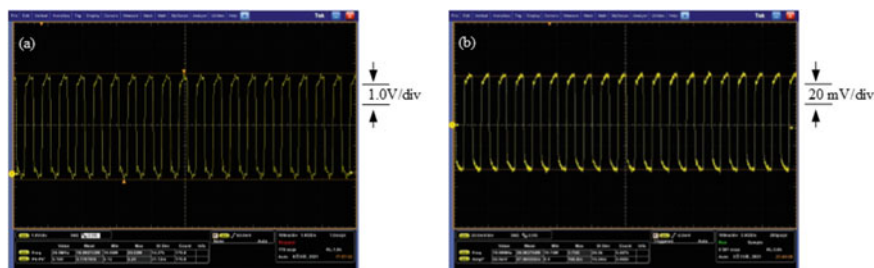


Fig. 8.24 Signal waveform for 1.3 km communication link **a** transmitting signal **b** receiving signal

For the 1.3 km link, Fig. 8.23 shows the tracks of beam tracking, the corresponding waveform in time domain and power spectral density estimation. The step angle of the two-dimensional mirror is adjusted to $10.92 \mu\text{rad}$. According to Eq. (8.36), with the increase of the number of iterations, the shape center of the spot is gradually adjusted from the first quadrant of the detection surface to the center position. After beam tracking, the variances of the spot centroid at the center of the camera in X and Y direction are 2.2770 pixel^2 and 1.3697 pixel^2 , respectively. The time elapsed from one spot acquisition to the final motor adjustment is 0.05 s , which is enough to compensate for the drift rate of the spot.

After the beam tracking is completed, the beam located in the detector branch is coupled into the photosensitive surface of the detector through the focusing lens. Figure 8.24b shows the output waveforms of the detector after tracking processing. When the transmission power is 15 mW , the output amplitude of the detector is 92.4 mV .

8.4.3.3 Field Experiments for a 10.3 km Communication Link

On the basis of the experiment for 1.3 km communication link, another field experiment of rapid alignment was performed for 10.3 km communication link. Figure 8.25a shows the experimental link diagram for a 10.3 km communication link. The transmitter and receiver are located at Xiaozhai Village, Bai Lu Yuan, Xi'an and the Building #6 in Xi'an University of Technology, respectively.

Figure 8.26 shows the coordinate tracking curve of the imaging position of the two-dimensional mirror after coarse alignment under the 10.3 km link. Within 6 h, the pitch angle was adjusted four times and the azimuth angle was adjusted twice. The beam drift caused by strong turbulence makes the coarse tracking frequency of the beam drift range at 10.3 km higher than 1.3 km.

Figure 8.27 shows the bidirectional alignment experiment of 10.3 km link optical wireless communication. After the beam emitted by laser a reaches the receiving end through atmospheric turbulence, adjust the two-dimensional mirror at the receiver to reflect the beam to the receiving Maksutov antenna. After converging by Maksutov antenna and focusing lens, detector a at the receiving end can detect the source

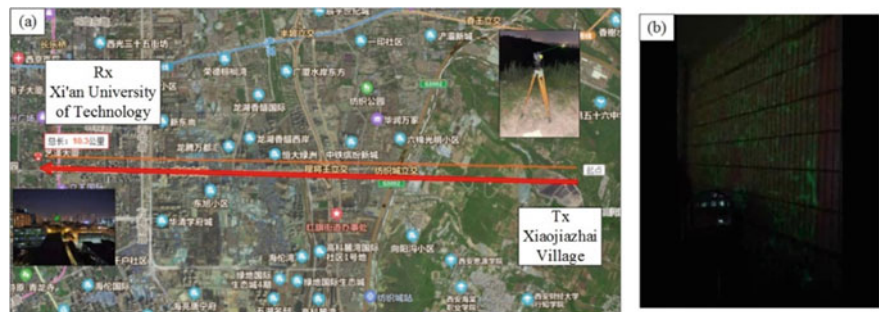


Fig. 8.25 Experiment of optical wireless communication for 10.3 km **a** link diagram **b** far field spot

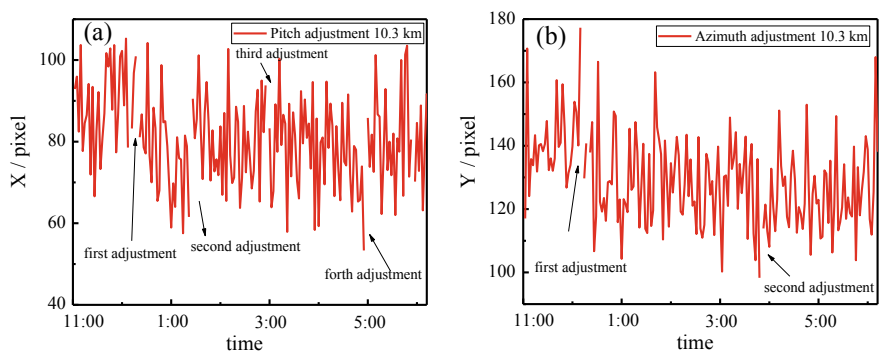


Fig. 8.26 Calibrate camera tracking curve at 10.3 km **a** pitch adjustment **b** azimuth adjustment

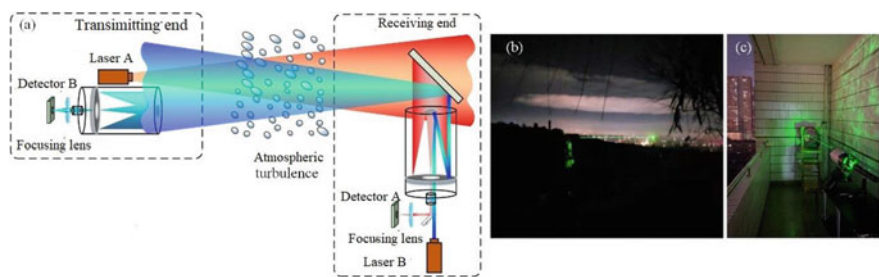


Fig. 8.27 14 Bidirectional alignment experiment (10.3 km): **a** system schematic diagram **b** transmitting end **c** receiving end

information from the transmitter; Laser B emits the light beam at the focus position of the receiving antenna, which is reflected by the two-dimensional mirror and reaches the transmitting end through atmospheric turbulence. The Maksutov antenna and the focusing lens at the transmitting end converge the light beam, and detector B can

detect the source information transmitted by the receiving end, so as to realize the two-way alignment of the light beam. By adjusting the two-dimensional mirror at the receiving end, because the optical path is reversible, the beam alignment from the receiving end to the transmitting end can be completed while completing the complete alignment from the transmitting end to the receiving end. Figure 8.28 shows the waveform output by the downlink detector and the spot detected by the uplink after bidirectional alignment.

As shown in Fig. 8.29, the mean variance and power spectral density estimation of the change of spot centroid in X and Y direction under the communication distance of 10.3 km are calculated respectively. The mean drift of spot centroid in X direction is 156.4072 pixel, the variance is 18.8653 pixel², the mean drift of spot centroid in Y direction is 118.0435 pixel, the variance is 10.5290 pixel², and the drift degree of spot in X direction is greater than that in Y direction. With the increase of communication distance, the low-frequency component in the power spectral density curve decreases compared with 1.3 km, and the enhancement of atmospheric turbulence significantly aggravates the spot drift rate.

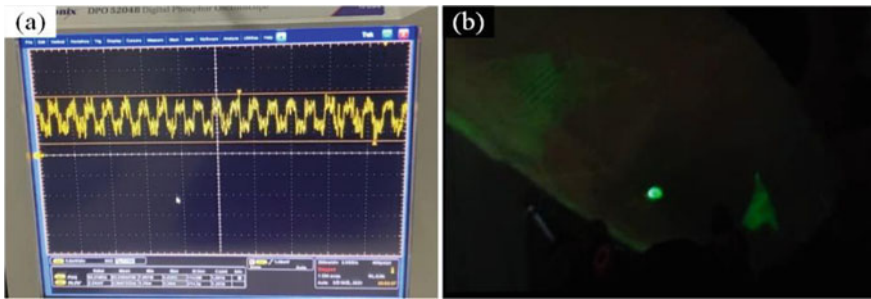


Fig. 8.28 Experimental results of bidirectional alignment (10.3 km): **a** downlink transmission signal waveform **b** uplink detection spot

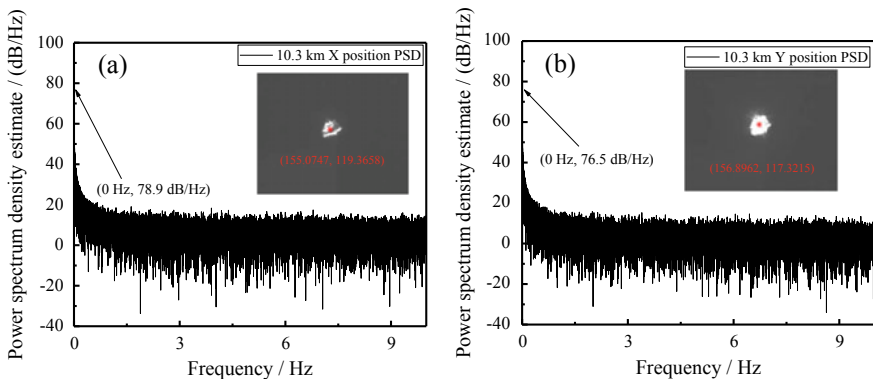


Fig. 8.29 Power spectral density estimation of spot center drift at 10.3 km **a** X position **b** Y position

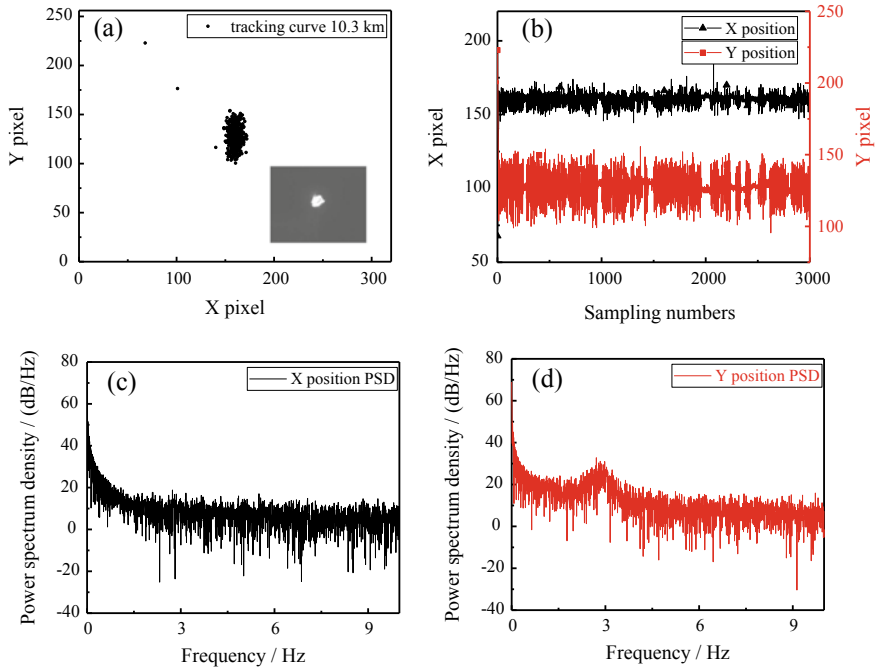


Fig. 8.30 Tracking waveform at 10.3 km **a** tracking curve **b** track curve in X and Y directions **c** X position power spectral density estimation **d** Y position power spectral density estimation

Figure 8.30 shows the centroid trajectory curve of beam tracking spot and the corresponding time domain waveform and power spectral density estimation at 10.3 km. The adjustment step angle of two-dimensional mirror is $10.92 \mu\text{rad}$. The variances of the spot centroid located in the center of the camera in the X direction and Y direction are 14.4970 pixel^2 and 8.0287 pixel^2 respectively. The adjustment range of the spot centroid should be greater than 1.3 km, and the proportion of low-frequency components in the power spectral density curve of the adjustment track should also be less than 1.3 km, which is due to the increase of the drift of the spot centroid caused by the increase of turbulence. The two-dimensional mirror needs to adjust a larger angle to adapt to the changes of the external environment.

Figure 8.31 is the waveform detected by the detector after tracking processing, in which the laser power output from the transmitter is 200 mW and the output amplitude of the detector is 74.4 mV.

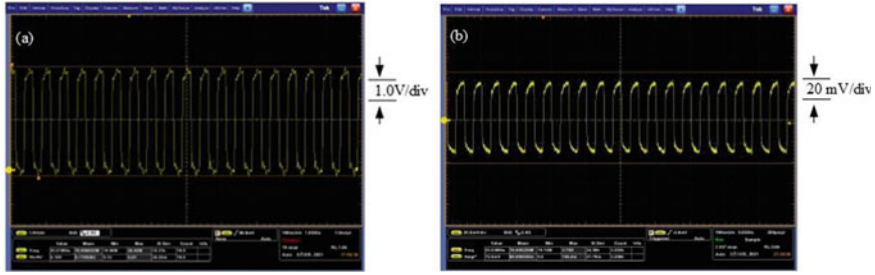


Fig. 8.31 Signal waveform at 10.3 km communication link **a** transmitting signal **b** receiving signal

8.5 Alignment Error

8.5.1 Attenuation Model of Optical Power

Consider a fundamental Gaussian beam propagating along z axis. The transmitter is located at $z = 0$, and its light field is expressed as Gaussian distribution, that is

$$U_0(r, 0) = A_0 \exp\left(-\frac{r^2}{W_0^2} - i\frac{kr^2}{2F_0}\right). \quad (8.38)$$

Assume that the receiver is located at $z = L$, the light field is

$$U_0(r, L) = \frac{A_0}{1 + ia_0L} \exp\left[-ikL - \frac{1}{2}\left(\frac{a_0kr^2}{1 + ia_0L}\right)\right], \quad (8.39)$$

where

$$a_0 = \frac{2}{kW_0^2} + i\frac{1}{F_0}. \quad (8.40)$$

The intensity distribution of a Gaussian beam is

$$I(r, L) = I_0 \cdot \frac{W_0^2}{W_L^2} \cdot \exp\left(-2\frac{r^2}{W_L^2}\right), \quad (8.41)$$

where W_L is the spot radius at $z = L$

$$W_L = W_0 \sqrt{1 + \left(\frac{\lambda L}{\pi W_0^2}\right)^2}. \quad (8.42)$$

The divergence angle is usually used to indicate the degree of divergence of the beam. Define the change rate of the spot radius of the Gaussian beam with the propagation distance as its divergence angle, denoted by θ_B

$$\theta_B = 2 \frac{dW_L}{dz} = \frac{2 \frac{\lambda L}{\pi W_0}}{\sqrt{\left(\frac{\pi W_0^2}{\lambda}\right)^2 + L^2}}. \quad (8.43)$$

As L approaches infinity ($L \rightarrow \infty$),

$$\theta = \lim_{z \rightarrow \infty} \frac{2W_L}{L} = \frac{2\lambda}{\pi W_0}. \quad (8.44)$$

8.5.1.1 Geometric Attenuation Model of Plane Wave

Regardless of the influence of turbulence, the received optical power model of the optical wireless communication system can be written as

$$P_R = P_T \cdot L_{geo} \cdot L_{atm}, \quad (8.45)$$

where P_T and P_R are the transmitted and received optical power, respectively, L_{geo} is geometric attenuation, $L_{atm} = e^{-\mu \cdot L}$ is the absorption and scattering attenuation caused by atmospheric, L are the communication distance, and μ is the atmospheric attenuation coefficient.

Assuming that the emitted light is a uniform plane wave, that is, the light intensity is uniformly distributed in a plane perpendicular to the transmission path, the geometric attenuation is

$$L_{geo,pl} = \left(\frac{D_R}{D_T + \theta \cdot L} \right)^2. \quad (8.46)$$

8.5.1.2 Geometric Attenuation Model of Gaussian Beam

For Gaussian beams, the geometric attenuation of the received optical power $L_{geo,gau}$ is

$$L_{geo,gau} = \frac{P_R}{P_T} = \frac{\int_0^{D_R/2} I(r, L) \cdot 2\pi r dr}{P_T}$$

$$= \frac{\frac{\pi I_0 W_0^2}{2} \cdot \left[1 - \exp\left(-\frac{D_R^2}{2W_L^2}\right) \right]}{P_T}. \quad (8.47)$$

Regarding the transmitter as a whole, the transmitted optical power is the output power of the optical antenna:

$$P_T = \int_0^{D_T/2} I(r, 0) \cdot 2\pi r dr = \frac{\pi I_0 W_0^2}{2} \cdot \left[1 - \exp\left(-\frac{D_T^2}{2W_0^2}\right) \right]. \quad (8.48)$$

Therefore, the geometric attenuation model of Gaussian beam without alignment error is

$$L_{geo, gau} = \frac{P_R}{P_T} = \frac{1 - \exp\left(-\frac{D_R^2}{2W_L^2}\right)}{1 - \exp\left(-\frac{D_T^2}{2W_0^2}\right)}. \quad (8.49)$$

8.5.2 Geometric Attenuation Model of Gaussian Beam with Alignment Error

(1) Alignment error model

Assuming that the displacements in X and Y axis caused by alignment error are independently and identically distributed random variables, which obey Gaussian distribution, the displacement at receiver is described by the Rayleigh distribution as

$$f_\rho(\rho) = \frac{\rho}{\sigma_\rho^2} \exp\left(-\frac{\rho^2}{2\sigma_\rho^2}\right). \quad (8.50)$$

(2) Geometric attenuation model with alignment error

Establishing a coordinate system with the center of the receiving aperture as the origin, In the case of pointing errors, the intensity distribution model of the Gaussian beam with alignment errors can be expressed as

$$I(r - \rho, L) = I_0 \cdot \frac{W_0^2}{W_L^2} \cdot \exp\left(-2\frac{(r - \rho)^2}{W_L^2}\right), \quad (8.51)$$

$$L_{err}(\rho) = \frac{P_R(\rho)}{P_T} = \frac{\int_A I(r - \rho, L)}{P_T}. \quad (8.52)$$

Under the assumption of rectangular aperture, an approximation is derived as

$$L_{err1}(\rho) \approx A_0 \exp\left(-\frac{2\rho^2}{W_{Leq}^2}\right), \quad (8.53)$$

$$A_0 = [erf(v)]^2, \quad v = (\sqrt{\pi} D_R/2)/(\sqrt{2} W_L), \quad W_{Leq}^2 = W_L^2 \frac{\sqrt{\pi} erf(v)}{2v \exp(-v^2)}. \quad (8.54)$$

Since the receiving aperture is circular in practice, after integral in polar coordinates, we obtain

$$L'_{err}(\rho) = L_{err}(\rho) \left[1 - \exp\left(-\frac{D_T^2}{2W_0^2}\right) \right], \quad (8.55)$$

$$\begin{aligned} L'_{err}(\rho) &= \int_0^{2\pi} \int_0^{D_R/2} \frac{2}{W_L^2} \cdot \exp\left[-2\frac{(r-\rho)^2}{W_L^2}\right] \cdot r dr d\theta \\ &= \frac{4}{W_L^2} \cdot \int_0^{D_R/2} \exp\left[-2\frac{(r-\rho)^2}{W_L^2}\right] \cdot r dr \\ &= \exp\left(-2\frac{\rho^2}{W_L^2}\right) - \frac{W_L^2}{4} \cdot \exp\left[-2\frac{(D_R/2-\rho)^2}{W_L^2}\right] \\ &\quad + \frac{4}{W_L^2} \cdot \rho \cdot \int_{-\sqrt{2}\rho/W_L}^{\sqrt{2}(D_R/2-\rho)/W_L} \exp(-x^2) \frac{W_L}{\sqrt{2}} dx. \end{aligned} \quad (8.56)$$

where $x = \sqrt{2}(r - \rho)/W_L$.

According to the value of $D_R/2 - \rho$ in the Eq. (8.56), the integral term is discussed in three cases:

- A. When $D_R/2 - \rho > 0$, that is, the center of the deviated spot is within the receiving aperture, then

$$\begin{aligned} &\rho \cdot \int_{-\sqrt{2}\rho/W_L}^{\sqrt{2}(D_R/2-\rho)/W_L} \exp(-x^2) \cdot \frac{W_L}{\sqrt{2}} dx \\ &= \frac{W_L \rho \sqrt{\pi}}{2\sqrt{2}} \cdot \left[erf\left(\sqrt{2}\rho/W_L\right) + erf\left(\sqrt{2}(D_R/2 - \rho)/W_L\right) \right] \end{aligned} \quad (8.57)$$

- B. When $D_R/2 - \rho < 0$, that is, the center of the deviated spot is drifted out of the receiving aperture, then

$$\begin{aligned} &\rho \cdot \int_{-\sqrt{2}\rho/W_L}^{\sqrt{2}(D_R/2-\rho)/W_L} \exp(-x^2) \cdot \frac{W_L}{\sqrt{2}} dx \\ &= \frac{W_L \rho \sqrt{\pi}}{2\sqrt{2}} \cdot \left[erf\left(\sqrt{2}\rho/W_L\right) - erf\left(\sqrt{2}(\rho - D_R/2)/W_L\right) \right] \end{aligned} \quad (8.58)$$

- C. When $D_R/2 - \rho = 0$, that is, the center of the deviated spot is at the edge of the receiving aperture, then

$$\begin{aligned}
& \rho \cdot \int_{-\sqrt{2}\rho/W_L}^{\sqrt{2}(D_R/2-\rho)/W_L} \exp(-x^2) \cdot \frac{W_L}{\sqrt{2}} dx \\
&= \frac{W_L \rho}{\sqrt{2}} \cdot \left\{ \int_{-\sqrt{2}\rho/W_L}^0 \exp(-x^2) dx \right\} = \frac{W_L \rho \sqrt{\pi}}{2\sqrt{2}} \cdot \operatorname{erf}\left(\sqrt{2}\rho/W_L\right). \quad (8.59)
\end{aligned}$$

In Eqs. (8.57)–(8.59), $\operatorname{erf}(x) = (2/\sqrt{\pi}) \int_0^x e^{-u^2} du$ is the error function.

Let $u = \sqrt{2}\rho/W_L$ and $v = \sqrt{2}(D_R/2 - \rho)/W_L$, then the accurate and analytical model of the geometric attenuation of Gaussian beams with alignment error is

$$L_{err}(\rho) = \begin{cases} \frac{\exp(-u^2) - \exp(-v^2) + \sqrt{\pi} \cdot u \cdot [\operatorname{erf}(u) + \operatorname{erf}(|v|)]}{\left[1 - \exp\left(-\frac{D_T^2}{2w_0^2}\right)\right]}, & D_R/2 > \rho \\ \frac{\exp(-u^2) - \exp(-v^2) + \sqrt{\pi} \cdot u \cdot [\operatorname{erf}(u) - \operatorname{erf}(|v|)]}{\left[1 - \exp\left(-\frac{D_T^2}{2w_0^2}\right)\right]}, & D_R/2 \leq \rho. \end{cases} \quad (8.60)$$

8.5.3 Average Geometric Attenuation Model with Alignment Error

The received optical power of optical wireless communication with alignment error is expressed as

$$P_R = P_T \langle L_{err} \rangle L_{atm}, \quad (8.61)$$

where

$$\langle L_{err} \rangle = \int_0^\infty L_{err}(\rho) \cdot f_\rho(\rho) d\rho, \quad (8.62)$$

and $f_\rho(\rho)$ is the alignment error in Eq. (8.50).

Using Eq. (8.50) into Eq. (8.60), after simplification, we obtain

$$\begin{aligned}
\langle L_{err} \rangle &= \int_0^\infty L_{err2}(\rho) \cdot f_\rho(\rho) d\rho \\
&= \frac{A}{2\sigma_\rho^2} - \frac{1}{2\sigma_\rho^2} \cdot \exp\left(-\frac{C}{2}\right) \cdot \left\{ A \exp\left(-\frac{B^2}{A}\right) + \sqrt{AB} \left[\sqrt{\pi} \cdot \operatorname{erf}\left(\frac{B}{\sqrt{A}}\right) + 1 \right] \right\}, \\
&+ 2 \sum_{n=0}^\infty A_n(r)^{n+1} + \frac{2D_R}{W_L} \sum_{n=0}^\infty C_n(q)^n \quad (8.63)
\end{aligned}$$

where $A = \frac{2\sigma_\rho^2 W_L^2}{W_L^2 + 4\sigma_\rho^2}$, $B = \frac{2\sigma_\rho^2 D_R}{W_L^2 + 4\sigma_\rho^2}$, $C = \frac{D_R^2}{W_L^2 + 4\sigma_\rho^2}$, $W_L = W_0 \sqrt{1 + \left(\frac{\lambda L}{\pi W_0^2}\right)^2}$, $A_n = \frac{(-1)^n (n+1)}{(2n+1)}$, $r = \frac{4\sigma_\rho^2}{W_L^2}$, $C_n = \sum_{m=0}^{2n+1} \frac{C_{2n+1}^m (-1)^{n+m} 2^{3m/2}}{n!(2n+1)2^n} \Gamma\left(\frac{m+2}{2}\right) \left(\frac{\sigma_\rho}{D_R}\right)^m$, and $q = \frac{D_R^2}{W_L^2}$.

8.6 Summary and Prospects

APT system is needed for point-to-point long-distance communication, whereas strict alignment is not required for visible and ultraviolet communication. APT system is a mature technology, but many technologies remain to be explored in laser communication. At present, most of the APT systems are aided by navigation and positioning system, and the development of a laser communication APT system without assistance is necessary. The time required for APT to complete alignment is generally seconds. Simple, reliable, and well-performing dynamic APT system remains to be developed in the future.

8.7 Questions

- 8.1 Briefly describe the concept of advanced aiming.
- 8.2 How many scanning methods are there? Try to describe their principles.
- 8.3 Briefly describe the working principle of APT.
- 8.4 Briefly describe the principle of compound-axis control system.
- 8.5 Draw the principle diagram of the space tracking system and briefly describe the principle of automatic tracking.
- 8.6 How many ways are there for antenna scanning? Briefly describe its working principle.

8.8 Exercises

- 8.1 Consider an orbiting satellite at an altitude of approximately 1000 miles, which orbits the earth every two hours.
 - (1) Suppose that the total pointing error of satellite earth trunk line is $50 \mu\text{rad}$, and the height uncertainty is ± 10 miles. Try to determine the appropriate leading angle and the beam width required for communication.
 - (2) Assuming that both stations are in parallel motion to each other, and each station has its own independent aiming and velocity errors, derive the equation of the upper and lower leading angles.

8.2 Consider an optical point source transmitter working at frequency of 10^{14} Hz, and located within a field of view of $1^\circ \times 1^\circ$. The receiving area is $3 \times 3 \text{ cm}^2$.

- (1) Calculate the diffraction limited spatial modulus must be searched.
- (2) If the resolution is $50' \times 50'$, find the Q value in $Q = \frac{\Omega_u}{\Omega_r}$.

8.3 Consider a deviation circle with a radius of r_0 , which falls into a symmetrically positioned quarter detector. Derive following equation according to the area of the deviation circle:

$$E[\xi_2(t)] = aG_d e I_S (\pi r_0^2) \left\{ 1 - \frac{2}{\pi} \cos^{-1} \left(\frac{\psi_z f_c}{r_0} \right) + \frac{2\psi_z f_c}{\pi r_0} \left[1 - \left(\frac{\psi_z f_c}{r_0} \right)^2 \right]^{\frac{1}{2}} \right\}$$

8.4 Try to analyze the system error rate change caused by the spot drift in laser communication [37].

8.5 In a satellite communication system, the distance between two adjacent satellites is 4000 km. If 1550 nm laser is used, what is the path loss in free space [4]?

References

1. Tong S, Jiang H et al (2010) Space laser communication technology and system. National Defense Industry Press, Beijing
2. Ke X, Xi X (2004) Introduction to wireless laser communication. Beijing University of Posts and Telecommunications Press, Beijing
3. Ma J (1989) Basic problems of acquisition, tracking and pointing system. Optoelectron Eng 16(3):1–42
4. Lu H, Zhao W, Hu H et al (2011) Influence of spot offset on the performance of space laser communication system. High Power Laser Part Beam 04:895–900
5. Wen C, Wang J, Li Y (2007) Performance analysis of receiving antenna in satellite optical communication. J PLA Univ Sci Technol (Nat Sci Edn) 3:211–215
6. Wu F, Yu S, Ma Z et al (2014) Correction of pointing angle deviation and in-orbit validation in satellite-ground laser communication links. Chin J Lasers 41(06):605008
7. Wu F, Yu S, Zhou J et al (2013) Analysis of the stability of bidirectional beam tracking in inter-satellite optical communication. Chin J Lasers 40(11):1105003
8. Li Q, Liu L, Ma XF et al (2019) Development of multi-target acquisition, pointing and tracking system for airborne laser communication. IEEE Trans Industr Inf 15(3):1720–1729
9. Qian F, Jia J, Zhang L et al (2014) Defective pixel correction of spot-detecting camera in satellite-to-ground laser communication ATP system. Chin J Lasers 41(5):505007
10. Ke X, Wu J, Yang S (2021) Research progress and prospect of atmospheric turbulence for wireless optical communication. Chin J Lasers 36(3):323–339
11. Ke X, Lei S, Yang P (2016). Beam coaxial alignment detection in atmospheric laser communication. Chin J Radio Sci 43(06):606003
12. Ke X, Zhang P (2021) A follow-me control system and follow-me control method for wireless optical communication CN201910339487.9. 07–20
13. Ke X, Lu N, Zhao L (2010). A beam automatic capture device and beam capture method CN201010185116.9. 10–06

14. Zhang M, Li B, Tong SF (2020) A new composite spiral scanning approach for beaconless spatial acquisition and experimental investigation of robust tracking control for laser communication system with disturbance. *IEEE Photon J* 12(6):7906212
15. Xiao Y, Dong R, Xiong Z (2011) Research on fixed-point acquisition in optical communication based on GPS. *Semicond Optoelectron* 32(6):833–835
16. Xiao Y, Ai Y, Song R et al (2012) Experiment of non-maneuvering target tracking based on ATP system. *Infrared Laser Eng* 41(9):2439–2443
17. Tan L, Wu S, Han Q et al (2012) Coarse tracking of periscope-type satellite optical communication terminals. *Opt Precis Eng* 20(2):270–276
18. Ke X, Wang J (2017). A four-quadrant detector-based spot alignment method CN201611244466.1. 05–31
19. Chen S, Wang L, Wang J (2017) Effects of digital to analog converter resolution on ATP system tracking accuracy. *Chin J Lasers* 44(8):806004
20. Chen MW, Yang YP, Jia XT et al (2013) Investigation of position algorithm and method for increasing the linear measurement range for four-quadrant detector. *Optik* 124:6806–6809
21. Liu Y, Jiang H, Tong S (2011) Study on stabilizational tracking technology for atmospheric laser communication system. *Chin J Lasers* 38(5):505005
22. Li S, Zhang Y (2017) Annular facula detection and error compensation of four-quadrant photoelectric detector in space laser communication. *Chin J Lasers* 44(11):1106005
23. Vo QS, Zhang XD, Fang FZ (2019) Extended the linear measurement range of four-quadrant detector by using modified polynomial fitting algorithm in micro-displacement measuring system. *Opt Laser Technol* 112:332–338
24. Bao JY, Xing F, Sun T et al (2019) CMOS imager non-uniformity response correction-based high-accuracy spot target localization. *Appl Opt* 58(16):4560–4568
25. Zhang WG, Guo W, Zhang CW (2019) An improved method for spot position detection of a laser tracking and positioning system based on a four-quadrant detector. *Sensors* 19(21):4722
26. Chen G, Dong Z, Geng J et al (2004) 155/622 Mb/s multiple transmitter laser communication systems. *Chin J Lasers* 31(5):583–587
27. Nielson T (1995) Pointing, acquisition, and tracking system for the free-space laser communication system SILEX. *Proc SPIE Int Soc Opt Eng* 2381:194–205
28. Nielsen TT, Oppenhaeuser G (2002) In-orbit test result of an operational optical intersatellite link between ARTEMIS and SPOT4. *SILEX Proc SPIE Int Soc Opt Eng* 46(35):1–15
29. Liu W, Yao KN, Huang DN et al (2016) Performance evaluation of coherent free space optical communications with a double-stage fast-steering-mirror adaptive optics system depending on the Greenwood frequency. *Opt Express* 24(12):13288–13302
30. Chang QB, Chen WS, Liu JK et al (2021) Development of a novel two-DOF piezo-driven fast steering mirror with high stiffness and good decoupling characteristic. *Mech Syst Signal Process* 159:107851
31. Dubra A, Massa JS, Paterson C (2005) Preisach classical and nonlinear modeling of hysteresis in piezoceramic deformable mirrors. *Opt Express* 13(22):9062–9070
32. Zuo T, Huang H, Xiao Y (2012) Fine tracking system design of space optical communication based on self-tuning control. *Chin J Scient Instrum* 33(5):1181–1186
33. Jono T, Toyoshima M, Takahashi N et al (2002) Laser tracking test under satellite microvibrational disturbances by OICETS ATP system. *Proc SPIE Int Soc Opt Eng* 4714:97–104
34. Borrello M (2005) A multi stage pointing acquisition and tracking (PAT) control system approach for air to air laser communications. *Amer Control Conf*, 3975–3980
35. Suite MR, Burris HR, Moore CI et al (2004) Fast steering mirror implementation for reduction of focal-spot wander in a long-distance free-space communication link. *Proc SPIE Int Soc Opt Eng* 5160:439–446
36. Pokorny P (2014) One-mirror and two-mirror three-dimensional optical scanners—position and accuracy of laser beam spot. *Appl Opt* 53(12):2730–2740
37. Held KJ, Barry TD (1996) Precision, acquisition and tracking system for the free space laser communication system[P]. *SILEX Proc SPIE* 2381:194–205

Chapter 9

Partially Coherent Optical Transmission



In a wireless laser communication system, when the beam is transmitted through the atmospheric channel, there will be light intensity flicker, beam expansion, spot drift, and angle of arrival fluctuation. Such turbulence effects reduce the energy coupling efficiency on the photodetector surface, increase the bit error rate of the laser communication system, and deteriorate the performance of the system. The partially coherent light transmission can reduce beam expansion, light intensity flicker, and spot drift and suppress the impacts of atmospheric turbulence.

9.1 Basic Parameters of a Light Beam

9.1.1 Emission Beam

A light beam transmission link with a beam expander at the transmitter is shown in Fig. 9.1. Assuming that a collimated Gaussian beam is emitted, the beam parameters are expressed as

$$\Theta_0 = 1, \Lambda_0 = \frac{2L}{kW_0^2}, \quad (9.1)$$

where k is the light wave number, W_0 is the beam radius (i.e., the point when the amplitude drops to $1/e$), and L is the transmission distance from the transmitter to the pupil plane of the receiving system. Denoting W_1 the radius of the beam incident to the receiving lens and F_1 the curvature radius of the wavefront, the beam parameters are

$$\begin{aligned} \Theta_1 &= \frac{\Theta_0}{\Theta_0^2 + \Lambda_0^2} = 1 + \frac{L}{F_1}, \quad \bar{\Theta}_1 = 1 - \Theta_1, \\ \Lambda_1 &= \frac{\Lambda_0}{\Theta_0^2 + \Lambda_0^2} = \frac{2L}{kW_1^2}, \end{aligned} \quad (9.2)$$

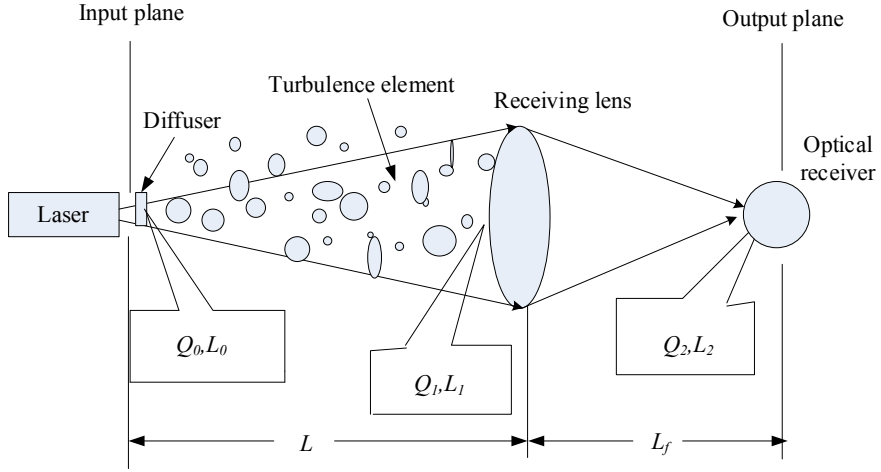


Fig. 9.1 Beam parameters and transmission geometry of collimated beam passing through a beam expander

Through a receiving lens with aperture W_G and focal length F_G , the beam parameters on the detector plane can be expressed as

$$\begin{aligned}\Theta_2 &= \frac{L}{F_f} \left[\frac{\frac{L}{F_f} - \frac{L}{F_G} + \bar{\Theta}_1}{\left(\frac{L}{F_f} - \frac{L}{F_G} + \bar{\Theta}_1\right)^2 + (\Lambda_1 + \Omega_G)^2} \right] = 0, \\ \Lambda_2 &= \frac{L}{F_f} \left[\frac{\Lambda_1 + \Omega_G}{\left(\frac{L}{F_f} - \frac{L}{F_G} + \bar{\Theta}_1\right)^2 + (\Lambda_1 + \Omega_G)^2} \right] = \frac{L}{F_f(\Lambda_1 + \Omega_G)},\end{aligned}\quad (9.3)$$

where $\Omega_G = 2L/kW_G^2$ is the dimensionless parameter, which represents the receiving lens with a finite size. The relationship between the beam radius W_2 and the image plane Λ_2 is

$$\Lambda_2 = \frac{2L_f}{kW_2^2}. \quad (9.4)$$

9.1.2 Mutual Interference Function

The beam in Fig. 9.1 can be described by a Gaussian Schell model (GSM) or a phase screen of a Gaussian spectral model.

9.1.2.1 Gaussian Schell Model

If the beam expander is placed on the aperture of the coherent laser transmitter, the beam after expanding is expressed as

$$\tilde{U}_0(\mathbf{s}, 0) = U_0(\mathbf{s}, 0) \exp[i\varphi(\mathbf{s})], \quad (9.5)$$

where $U_0(\mathbf{s}, 0)$ is the field incident to the beam expander, \mathbf{s} is the vector on its cross-section, and $\varphi(\mathbf{s})$ is the random phase of zero mean. Assuming that the correlation function of GSM beam is related to the $\exp[i\varphi(\mathbf{s})]$ random phase factor of the beam expander, it can be described by a Gaussian function:

$$\begin{aligned} B(\mathbf{s}_1, \mathbf{s}_2, 0) &= \langle \tilde{U}_0(\mathbf{s}_1, 0) \tilde{U}_0^*(\mathbf{s}_2, 0) \rangle \\ &= U_0(\mathbf{s}_1, 0) U_0^*(\mathbf{s}_2, 0) \langle \exp\{i[\varphi(\mathbf{s}_1) - \varphi(\mathbf{s}_2)]\} \rangle \\ &= U_0(\mathbf{s}_1, 0) U_0^*(\mathbf{s}_2, 0) \exp\left(-\frac{|\mathbf{s}_1 - \mathbf{s}_2|^2}{2\sigma_c^2}\right), \end{aligned} \quad (9.6)$$

where σ_c^2 represents the estimated correlation width of the beam expander, which describes the partial coherence characteristics of the effective source. If $\sigma_c^2 \gg 1$, the source field is completely coherent light. If σ_c^2 is much smaller than the wavelength of the source field, the effective emission source is incoherent light. The coherence of the source can be expressed by the coherence parameters of the source:

$$\zeta_s = 1 + \frac{W_0^2}{\sigma_c^2}, \quad (9.7)$$

which is used to describe the number of “speckle units” on the beam expander. For $\sigma_c^2 \gg W_0^2$, that is, the weak beam expansion, the number of speckle elements is one (i.e., coherent light). In the case of strong beam expansion, many speckle patterns appear, and each speckle unit is regarded as an independent source.

9.1.2.2 Analysis of Free Space on Pupil Plane

Based on the generalized Huygens–Fresnel principle, the autocorrelation correlation function (MCF) of a single collimated Gaussian light source transmitted to the pupil plane of the receiver is defined as

$$\Gamma_{\text{pp,diff}}(\mathbf{r}_1, \mathbf{r}_2, L) = \frac{k^2}{4\pi^2 L^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d^2 s_1 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d^2 s_2 \langle \tilde{U}_0(\mathbf{s}_1, 0) \tilde{U}_0^*(\mathbf{s}_2, 0) \rangle \quad (9.8)$$

$$\times \exp \left[\frac{ik}{2L} |\mathbf{s}_1 - \mathbf{r}_1|^2 - \frac{ik}{2L} |\mathbf{s}_2 - \mathbf{r}_2|^2 \right].$$

By substituting the correlation function in Eq. (9.6) into Eq. (9.8), we obtain

$$\Gamma_{\text{pp,diff}}(\mathbf{r}_1, \mathbf{r}_2, L) = \frac{k^2}{4\pi^2 L^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d^2 s_1 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d^2 s_2 \exp \left(-\frac{s_1^2 + s_2^2}{W_0^2} \right) \exp \left(-\frac{|\mathbf{s}_1 - \mathbf{s}_2|^2}{l_c^2} \right) \quad (9.9)$$

$$\times \exp \left[\frac{ik}{2L} |\mathbf{s}_1 - \mathbf{r}_1|^2 - \frac{ik}{2L} |\mathbf{s}_2 - \mathbf{r}_2|^2 \right]$$

where $l_c = \sqrt{2}\sigma_c$ is the correlation radius, and the transmitted light in Eq. (9.9) is a collimated beam of a unit amplitude. By integrating Eq. (9.9), we obtain

$$\Gamma_{\text{pp,diff}}(\mathbf{r}_1, \mathbf{r}_2, L) = \frac{W_0^2}{W_1^2(1 + 4\Lambda_1 q_c)} \exp \left[\frac{ik}{L} \left(\frac{1 - \Theta_1 + 4\Lambda_1 q_c}{1 + 4\Lambda_1 q_c} \right) \mathbf{r} \cdot \mathbf{p} \right] \quad (9.10)$$

$$\times \exp \left[\frac{2r^2 + \rho^2/2}{W_1^2(1 + 4\Lambda_1 q_c)} \right] \exp \left[-\left(\frac{\Theta_1^2 + \Lambda_1^2}{1 + 4\Lambda_1 q_c} \right) \left(\frac{\rho^2}{l_c^2} \right) \right],$$

where $\mathbf{p} = \mathbf{r}_1 - \mathbf{r}_2$, $\rho = |\mathbf{p}|$, $\mathbf{r} = (1/2)(\mathbf{r}_1 + \mathbf{r}_2)$, W_1 is the spot radius of the fully correlated beam on the pupil plane, and q_c are the dimensionless coherent parameters:

$$q_c = \frac{L}{kl_c^2}. \quad (9.11)$$

All the second-order statistics can be obtained using Eq. (9.10). For example, if $\mathbf{r}_1 = \mathbf{r}_2 = \mathbf{r}$, the average irradiance on the pupil surface is

$$\langle I(\mathbf{r}, L) \rangle_{\text{pp,diff}} = \frac{W_0^2}{W_1^2(1 + 4\Lambda_1 q_c)} \exp \left[\frac{2r^2}{W_1^2(1 + 4\Lambda_1 q_c)} \right]. \quad (9.12)$$

The spot radius of partially coherent light can be expressed as

$$W_{\text{pp,diff}} = W_1 \sqrt{1 + 4\Lambda_1 q_c}. \quad (9.13)$$

Considering the standard MCF, the complex coherence coefficient DOC can be obtained similarly as

$$\text{DOC}_{\text{pp,diff}}(\rho, L) = \frac{|\Gamma_{\text{pp,diff}}(\mathbf{r}_1, \mathbf{r}_2, L)|}{\sqrt{\Gamma_{\text{pp,diff}}(\mathbf{r}_1, \mathbf{r}_1, L)\Gamma_{\text{pp,diff}}(\mathbf{r}_2, \mathbf{r}_2, L)}} = \exp\left[-\left(\frac{\Theta_1^2 + \Lambda_1^2}{1 + 4\Lambda_1 q_c}\right)\left(\frac{\rho^2}{l_c^2}\right)\right]. \quad (9.14)$$

Therefore, the average speckle radius is

$$\rho_{\text{pp,speckle}} = \sqrt{\frac{l_c^2(1 + 4\Lambda_1 q_c)}{\Theta_1^2 + \Lambda_1^2}} = \frac{l_c}{W_0} W_{\text{pp,diff}}. \quad (9.15)$$

This means that the ratio of the number of speckles on the input and output planes at the source is a constant:

$$\frac{W_0^2}{l_c^2} = \frac{W_{\text{pp,diff}}^2}{\rho_{\text{pp,speckle}}^2}. \quad (9.16)$$

Therefore, under the limitation of strong diffusion ($l_c \rightarrow 0$), the average speckle radius is

$$\rho_{\text{pp,speckle}} = \frac{2\sqrt{2}L}{k W_0} = \frac{\sqrt{2}\lambda L}{\pi W_0}. \quad (9.17)$$

9.1.2.3 Random Phase Screen Model

The beam expander can be described by a thin random phase screen. The spatial power spectrum of the beam is used to describe the random phase screen. We assume that the Gaussian spectral function is used:

$$\Phi_S(\kappa) = \frac{\langle n_1^2 \rangle l_c^3}{8\pi\sqrt{\pi}} \exp\left(-\frac{l_c^2 \kappa^2}{4}\right), \quad (9.18)$$

where l_c is the transverse correlation radius, with σ_c^2 directly related to the parameters of the GSM beam, that is, $l_c^2 = 2\sigma_c^2$. The parameter $\langle n_1^2 \rangle$ represents the fluctuation index of the phase screen, and the parameter l_c represents the correlation radius of the beam expander.

9.1.2.4 Analysis of Random Medium on Pupil Plane

If the light wave at the transmitter is an amplitude free collimated beam, and the light wave parameters are expressed in Eq. (9.1), then the field on the pupil plane of the receiver can be expressed as

$$U(\mathbf{r}, L) = U_0(\mathbf{r}, L) \exp[\Psi_S(\mathbf{r}, L)], \quad (9.19)$$

where $\Psi_S(\mathbf{r}, L)$ is the complex phase fluctuation caused by the beam expander. Let $d_3 = 1$; the MCF of the beam can be obtained as follows:

$$\begin{aligned} \Gamma_{\text{pp,diff}}(\mathbf{r}_1, \mathbf{r}_2, L) &= \Gamma_0(\mathbf{r}_1, \mathbf{r}_2, L) \exp\left\{-4\pi^2 k^2 \Delta z \int_0^\infty \kappa \Phi_S(\kappa) \right. \\ &\quad \left. \times \left[1 - e^{-\Lambda_1 L \kappa^2/k} J_0(\kappa |\Theta_1 \mathbf{p} - 2i \Lambda_1 \mathbf{r}|)\right] d\kappa\right\}, \end{aligned} \quad (9.20)$$

where Δz is the thickness of the phase screen, and $\Gamma_0(\mathbf{r}_1, \mathbf{r}_2, L)$ is the MCF without diffusion

$$\Gamma_0(\mathbf{r}_1, \mathbf{r}_2, L) = \frac{W_0^2}{W_1^2} \exp\left(-\frac{2r^2}{W_1^2} - \frac{\rho^2}{2W_1^2} - i \frac{k}{F_1} \mathbf{p} \cdot \mathbf{r}\right), \quad (9.21)$$

$$\begin{aligned} \Gamma_{\text{pp,diff}}(\mathbf{r}_1, \mathbf{r}_2, L) &= \Gamma_0(\mathbf{r}_1, \mathbf{r}_2, L) \exp[\sigma_{\text{r,diff}}^2(\mathbf{r}_1, L) + \sigma_{\text{r,diff}}^2(\mathbf{r}_2, L)] \\ &\quad \times \exp[-T_{\text{diff}}(L)] \exp\left[-\frac{1}{2} \Delta_{\text{diff}}(\mathbf{r}_1, \mathbf{r}_2, L)\right], \end{aligned} \quad (9.22)$$

where $\sigma_{\text{r,diff}}^2(\mathbf{r}, L)$ represents the change of average illuminance, $T_{\text{diff}}(L)$ describes the average illuminance along the longitudinal or axial direction, and $\text{Re}[\Delta_{\text{diff}}(\mathbf{r}_1, \mathbf{r}_2, L)] = D_{\text{diff}}(\mathbf{r}_1, \mathbf{r}_2, L)$ is the wave structure constant (WSF). Based on the Gaussian spectrum of Eq. (9.18), each part of Eq. (9.22) can be expressed as

$$\begin{aligned} \sigma_{\text{r,diff}}^2(\mathbf{r}, L) &= 2\pi^2 k^2 \Delta z \int_0^\infty \kappa \Phi_S(\kappa) e^{-\Lambda_1 L \kappa^2/k} [I_0(2\Lambda_1 r \kappa) - 1] d\kappa \\ &= \frac{\sqrt{\pi} \langle n_1^2 \rangle k^2 l_c \Delta z}{2(1 + 4\Lambda_1 q_c)} \left\{ \exp\left[\frac{4\Lambda_1^2 r^2}{(1 + 4\Lambda_1 q_c) l_c^2}\right] - 1 \right\}. \end{aligned} \quad (9.23)$$

After normalization, we obtain

$$\frac{\sqrt{\pi} \langle n_1^2 \rangle k^2 l_c \Delta z}{1 + 4\Lambda_1 q_c} = 1. \quad (9.24)$$

Using Rytov approximation, we obtain the following results:

$$\sigma_{\text{r,diff}}^2(\mathbf{r}, L) = \frac{4\Lambda_1^2 \mathbf{r}^2}{(1 + 4\Lambda_1 q_c) l_c^2} = \frac{4\Lambda_1 q_c}{1 + 4\Lambda_1 q_c} \left(\frac{\mathbf{r}^2}{W_1^2}\right). \quad (9.25)$$

Similarly, the expression in Eq. (9.22), $T_{\text{diff}}(L)$, can be expressed as

$$T_{\text{diff}}(L) = 4\pi^2 k^2 \Delta z \int_0^\infty \kappa \Phi_S(\kappa) \left(1 - e^{-\Lambda_1 L \kappa^2 / k}\right) d\kappa = 4\Lambda_1 q_c. \quad (9.26)$$

Other quantities in Eq. (9.22) can be provided directly. If $\mathbf{r}_1 = \mathbf{r} - \mathbf{p}/2$, and $\mathbf{r}_2 = \mathbf{r} + \mathbf{p}/2$, we obtain

$$\begin{aligned} \Delta_{\text{diff}}^2(\mathbf{r}_1, \mathbf{r}_2, L) &= 4\pi^2 k^2 \Delta z \int_0^\infty \kappa \Phi_S(\kappa) e^{-\Lambda_1 L \kappa^2 / k} \\ &\times [I_0(2\Lambda_1 r_1 k) + I_0(2\Lambda_1 r_2 k) - 2J_0(\kappa |\Theta_1 \mathbf{p} - 2i\Lambda_1 \mathbf{r}|)] d\kappa \\ &= 2 \left(\frac{\Theta_1 + \Lambda_1^2}{1 + 4\Lambda_1 q_c} \right) \frac{\rho^2}{l_c^2} - \frac{4i\Theta_1 \mathbf{p} \cdot \mathbf{r}}{(1 + 4\Lambda_1 q_c) l_c^2}. \end{aligned} \quad (9.27)$$

Finally, combining Eqs. (9.25)–(9.27), we can obtain the following results:

$$\begin{aligned} \Gamma_{\text{pp,diff}}(\mathbf{r}_1, \mathbf{r}_2, L) &= \frac{W_0^2}{W_1^2(1 + 4\Lambda_1 q_c)} \exp \left[\frac{ik}{L} \left(\frac{1 - \Theta_1 + 4\Lambda_1 q_c}{1 + 4\Lambda_1 q_c} \right) \mathbf{r} \cdot \mathbf{p} \right] \\ &\times \exp \left[\frac{2\mathbf{r}^2 + \rho^2/2}{W_1^2(1 + 4\Lambda_1 q_c)} \right] \exp \left[- \left(\frac{\Theta_1^2 + \Lambda_1^2}{1 + 4\Lambda_1 q_c} \right) \left(\frac{\rho^2}{l_c^2} \right) \right] \end{aligned} \quad (9.28)$$

Using the approximation of $\exp[-T_{\text{diff}}(L)] \cong 1/[1 + T_{\text{diff}}(L)]$, both the spot radius and speckle radius can be obtained from the phase screen model.

9.1.3 Beam Spreading, Drift, and Intensity Fluctuation

9.1.3.1 Beam Spreading

As shown in Fig. 9.2, when the laser beam propagates in atmospheric turbulence, the beam will expand due to the influence of atmospheric refractive index fluctuation. Atmospheric turbulence results a part of the beam cannot reach the surface of the photodetector through the coupling lens, which increases the spot radius and lower the beam energy.

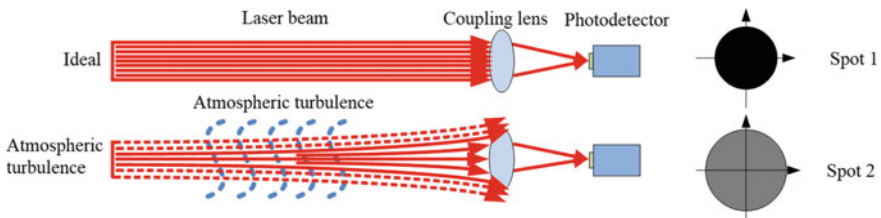


Fig. 9.2 Influence of beam spreading on atmospheric laser communication system

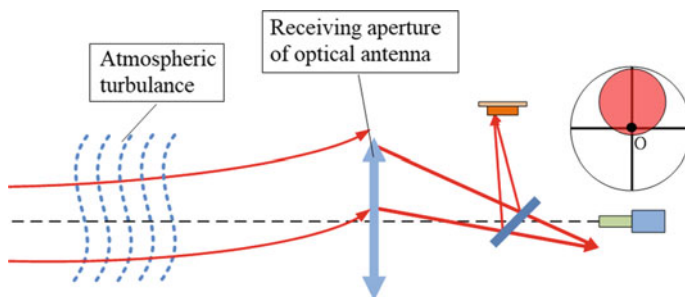


Fig. 9.3 Influence of spot drift on optical wireless communication system

9.1.3.2 Beam Drift

The wireless laser communication system is shown in Fig. 9.3. In the process of propagation, the beam drifts due to the influence of atmospheric turbulence; it causes the beam to move randomly in the direction perpendicular to the optical axis and deviate from the receiving surface. With further aggravation of turbulence, the transmitting beam may drift out of the receiving surface of the receiving antenna, causing the communication link to be interrupted.

9.1.3.3 Intensity Flicker

When the laser beam propagates in atmospheric turbulence, and the beam diameter is much larger than the turbulence scale, the beam cross-section contains several turbulence vortices. Each vortex independently scatters and diffracts the light beam on it, causing continuous changes of optical density on the detector plane in space and time, as shown in Fig. 9.4. This type of light intensity is called flicker intensity.

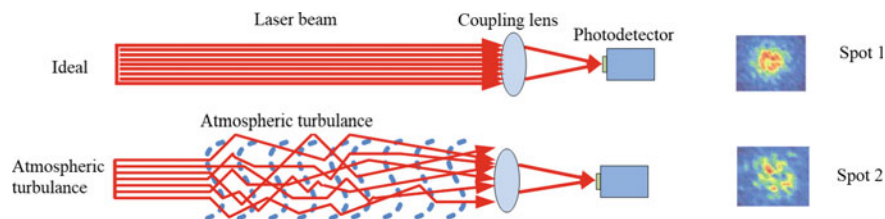
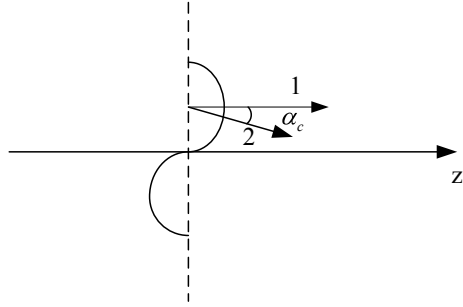


Fig. 9.4 Light intensity scintillation in an atmospheric laser communication system

Fig. 9.5 Local angle of arrival: 1-uniform phase front normal; and 2-equal phase normal



9.1.3.4 Fluctuation of Angle of Arrival

When a laser propagates in turbulent atmosphere, the fluctuation of atmospheric refractive index in different parts of the beam cross-section leads to different phase changes in different parts of the beam front. These changes lead to isophase surface with random fluctuation shape, as shown in Fig. 9.5. This phase deformation leads to the continuous change of the arrival angle of the beam wavefront, which is called the angle of arrival fluctuation effect.

9.2 Partially Coherent Light Model

Full coherence of light is not a necessary condition to produce good directivity. A partially spatially coherent light source can also produce far-field light intensity distribution, similar to laser. The type of beam with good directivity but only partial spatial coherence is called the partially coherent beam.

9.2.1 Description of Partially Coherent Light

9.2.1.1 Partially Coherent Light

The coherence of a beam includes its spatial and temporal coherence. In general, a partially coherent light source can be obtained by superimposing a random amplitude and phase distribution with time on a fully coherent field. It is assumed that the field distribution of the initial fully coherent light source is $U_0(x, y, 0)$, and the time-dependent amplitude and phase distributions are $t_A(x, y; t) = \exp[i\xi(x, y; t)]$. Therefore, the light field distribution of partially coherent light at the emission site can be expressed as [1]

$$U'(x, y, 0; t) = U_0(x, y, 0)t_A(x, y; t)$$

$$= U_0(x, y, 0) \exp[i\xi(x, y; t)], \quad (9.29)$$

where the partial coherence of the light field of the light source is expressed by $\xi(x, y; t)$.

9.2.1.2 Space-Time Domain

Suppose that $\rho = (x, y, z)$ represents the spatial position vector, and z is the beam propagation distance and a constant. $\vec{\rho}_1$ and $\vec{\rho}_2$ are any two points in the light field. $V(\vec{\rho}_1, t + \tau)$ and $V^*(\vec{\rho}_2, t)$ represent the complex amplitudes of space points $\vec{\rho}_1$ and $\vec{\rho}_2$ of partially coherent light at time t and $t + \tau$, respectively. In the space-time domain, the two light field points can be described by the mutual interference function $\Gamma(\vec{\rho}_1, \vec{\rho}_2, \tau)$. The definition of mutual interference function is [2]

$$\Gamma(\vec{\rho}_1, \vec{\rho}_2, \tau) = \langle V(\vec{\rho}_1, t + \tau) V^*(\vec{\rho}_2, t) \rangle, \quad (9.30)$$

where $\langle \cdot \rangle$ is the ensemble average. Assuming that the radiation field satisfies ergodicity, the ensemble average of the beam field can be obtained by averaging the calculation time:

$$\langle V(\vec{\rho}_1, t + \tau) V^*(\vec{\rho}_2, t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T V(\vec{\rho}_1, t + \tau) V^*(\vec{\rho}_2, t) dt, \quad (9.31)$$

where T is the measurement time. For $\vec{\rho}_1 = \vec{\rho}_2 = \rho$ and $\tau = 0$, the average light intensity at the space point $\vec{\rho} = (x, y, z)$ is

$$I(\vec{\rho}) = \langle V(\vec{\rho}, t) V^*(\vec{\rho}, t) \rangle = \Gamma(\vec{\rho}, \vec{\rho}, 0). \quad (9.32)$$

The spatial coherence of the light field can be described by the complex coherence degree $\gamma(\vec{\rho}_1, \vec{\rho}_2, \tau)$, which is called normalized mutual coherence function.

$$\gamma(\vec{\rho}_1, \vec{\rho}_2, \tau) = \frac{\Gamma(\vec{\rho}_1, \vec{\rho}_2, \tau)}{\sqrt{\Gamma(\vec{\rho}_1, \vec{\rho}_1, 0) \Gamma(\vec{\rho}_2, \vec{\rho}_2, 0)}} = \frac{\Gamma(\vec{\rho}_1, \vec{\rho}_2, \tau)}{\sqrt{I(\vec{\rho}_1) I(\vec{\rho}_2)}} \quad (9.33)$$

The light field can be divided into three categories according to the value of $|\gamma(\vec{\rho}_1, \vec{\rho}_2, \tau)|$:

$$\begin{cases} |\gamma(\vec{\rho}_1, \vec{\rho}_2, \tau)| = 0 & \text{Completely incoherent light} \\ 0 < |\gamma(\vec{\rho}_1, \vec{\rho}_2, \tau)| < 1 & \text{Partially coherent light} \\ |\gamma(\vec{\rho}_1, \vec{\rho}_2, \tau)| = 1 & \text{Completely coherent light} \end{cases}$$

It can be seen that the mode value of the normalized cross-correlation function, $|\gamma(\vec{\rho}_1, \vec{\rho}_2, \tau)|$, is between 0 and 1. The larger the value is, the higher the visibility of the interference fringes is and the stronger the coherence of the light field is. The temporal coherence of the light field is represented by $\gamma(\vec{\rho}, \vec{\rho}, \tau)$, which is called the auto-coherence function:

$$\Gamma(\tau) = \Gamma(\vec{\rho}, \vec{\rho}, \tau) = \langle V(\vec{\rho}_1, t + \tau) V^*(\vec{\rho}_2, t) \rangle. \quad (9.34)$$

The average intensity $I(\vec{\rho})$ can also be expressed by the autocorrelation function:

$$I(\vec{\rho}) = \langle V(\vec{\rho}, t) V^*(\vec{\rho}, t) \rangle = \Gamma(\vec{\rho}, \vec{\rho}, 0) = \Gamma(0) \quad (9.35)$$

The normalized autocorrelation function (complex autocorrelation degree) can be expressed as

$$\gamma(\tau) = \frac{\Gamma(\tau)}{\Gamma(0)} \quad (9.36)$$

and $\gamma(0) = 1$, $0 \leq \gamma(\tau) \leq 1$.

9.2.1.3 Space-Frequency Domain

In the space-frequency domain, the cross-spectral density is used to describe the coherence of a light field.

$$W(\vec{\rho}_1, \vec{\rho}_2, \omega) = \langle \hat{V}(\vec{\rho}_1, \omega) \hat{V}^*(\vec{\rho}_2, \omega) \rangle, \quad (9.37)$$

where $V(\vec{\rho}_1, \omega)$ and $\hat{V}(\vec{\rho}_1, \omega)$ are the field functions and their Fourier transforms.

$$\hat{V}(\vec{\rho}_j, \omega) = \int V(\vec{\rho}_j, t) \exp(i\omega t) dt, \quad j = 1, 2 \quad (9.38)$$

where ω is the frequency of the light field. Therefore, the relationship between the cross-spectral density function $W(\vec{\rho}, \vec{\rho}, \omega)$ and cross-correlation function $\Gamma(\vec{\rho}_1, \vec{\rho}_2, \tau)$ is

$$W(\vec{\rho}, \vec{\rho}, \omega) = \int \Gamma(\vec{\rho}_1, \vec{\rho}_2, \tau) \exp(i\omega \tau) d\tau \quad (9.39)$$

$$\Gamma(\vec{\rho}_1, \vec{\rho}_2, \tau) = \frac{1}{2\pi} \int W(\vec{\rho}, \vec{\rho}, \omega) \exp(-i\omega \tau) d\omega \quad (9.40)$$

Let $\vec{\rho}_1 = \vec{\rho}_2 = \vec{\rho}$; at a space point $\vec{\rho}$, the average light intensity of frequency ω is

$$I(\vec{\rho}, \omega) = W(\vec{\rho}, \vec{\rho}, \omega). \quad (9.41)$$

In the space–frequency domain, the temporal coherence is represented by the spectral density function $S(\omega)$, which is defined as

$$S(\omega) = W(\vec{\rho}, \vec{\rho}, \omega). \quad (9.42)$$

Even if the partially coherent light propagates in free space, the spectral density function $S(\omega)$ changes; this is called the Wolf effect [1]. The Complex degree of spatial coherence, also known as the spectral degree of coherence, can be obtained by normalizing the cross-spectral density.

$$\mu(\vec{\rho}_1, \vec{\rho}_2, \omega) = \frac{W(\vec{\rho}_1, \vec{\rho}_2, \omega)}{\sqrt{W(\vec{\rho}_1, \vec{\rho}_1, \omega)W(\vec{\rho}_2, \vec{\rho}_2, \omega)}} = \frac{W(\vec{\rho}_1, \vec{\rho}_2, \omega)}{\sqrt{S(\vec{\rho}_1, \omega)S(\vec{\rho}_2, \omega)}} \quad (9.43)$$

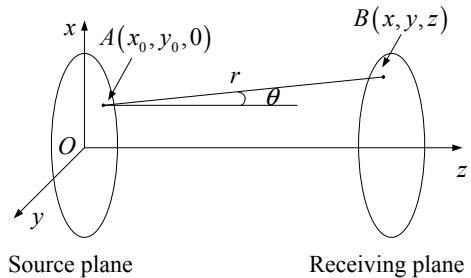
where the range of spectral degree of coherence is $0 \leq \mu(\vec{\rho}_1, \vec{\rho}_2, \omega) \leq 1$.

9.2.2 Partially Coherent Beam

9.2.2.1 Generalized Huygens–Fresnel Principle

The geometric schematic diagram of Huygens–Fresnel principle is shown in Fig. 9.6. In the light source plane $z = 0$, the light field $E_0(x_0, y_0)$ at any point $A(x_0, y_0, 0)$ can be regarded as a source of a spherical wave. The intensity of this wave source is proportional to the light field $E_0(x_0, y_0)$, and the amplitude in each direction can be represented by $K(\theta) = (1 + \cos \theta)/(2i\lambda)$. The field distribution $E(x, y, z)$ at any point $B(x, y, z)$ on the receiving surface is formed by superposition of all the spherical wave sources on the light source O [3]:

Fig. 9.6 The geometry of the Huygens–Fresnel principle



$$E(x, y, z) = \frac{1}{i\lambda} \iint E_0(x_0, y_0, 0) \frac{e^{ikr}}{r} \frac{(1 + \cos \theta)}{2} dx_0 dy_0. \quad (9.44)$$

When the light wave propagates in vacuum, the paraxial approximation is used.

$$E(x, y, z) = \frac{e^{ikr}}{i\lambda z} \iint E_0(x_0, y_0, 0) \exp\left\{\frac{ik}{2z}[(x - x_0)^2 + (y - y_0)^2]\right\} dx_0 dy_0 \quad (9.45)$$

When the beam propagates in turbulent atmosphere, the field change can be represented by a negative phase factor. When the light in the source plane propagates from $(x_0, y_0, 0)$ to (x, y, z) , the field change can be represented by a negative phase factor.

$$F_1 = \exp(\chi + iS_1) = \exp[\psi(\boldsymbol{\rho}', \mathbf{r})], \quad (9.46)$$

where χ and S_1 represent the weak fluctuation of amplitude and phase, respectively. At this time, Eq. (9.45) can be written as

$$E(x, y, z) = \frac{e^{ikr}}{i\lambda z} \iint E_0(x_0, y_0, 0) \exp\left\{\frac{ik}{2z}[(x - x_0)^2 + (y - y_0)^2]\right\} \times \exp[\psi(\boldsymbol{\rho}', \mathbf{r})] dx_0 dy_0 \quad (9.47)$$

9.2.2.2 Cross-Spectral Density Function

The two-point cross-spectral density function of partially coherent light is [4]

$$W_0(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \left\langle U'^*(\boldsymbol{\rho}_1) U'(\boldsymbol{\rho}_2) \right\rangle, \quad (9.48)$$

where $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2$ represents the two-dimensional vector at the light source plane. The cross-spectral density of partially coherent Schell beams is

$$W_0(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \sqrt{I_0(\boldsymbol{\rho}_1)} \times \sqrt{I_0(\boldsymbol{\rho}_2)} \mu_0(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1) \quad (9.49)$$

where $I_0(\boldsymbol{\rho})$ is the average intensity, and $\mu_0(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1)$ is the degree of coherence at $z = 0$.

When $I_0(\boldsymbol{\rho}) = A \exp\left(-\frac{|\boldsymbol{\rho}|^2}{2\sigma_s^2}\right)$ and $\mu_0(\boldsymbol{\rho}) = \exp\left(-\frac{|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2|^2}{2\sigma_g^2}\right)$, it is a partially coherent Gaussian Schell beam, and the cross-spectral density at $z = 0$ is

$$W^{(0)}(\rho_{s1}, \rho_{s2}, 0) = A \exp\left(-\frac{|\rho_{s1}|^2 + |\rho_{s2}|^2}{4\sigma_s^2}\right) \times \exp\left(-\frac{|\rho_{s1} - \rho_{s2}|^2}{2\sigma_g^2}\right) \quad (9.50)$$

where ρ_{s1}, ρ_{s2} is the coordinate vector of two points in the source plane; parameters A , σ_s , and σ_g represent the light intensity, beam waist width, and coherence length of the light source, respectively.

9.3 Beam Propagation in Atmospheric Turbulence

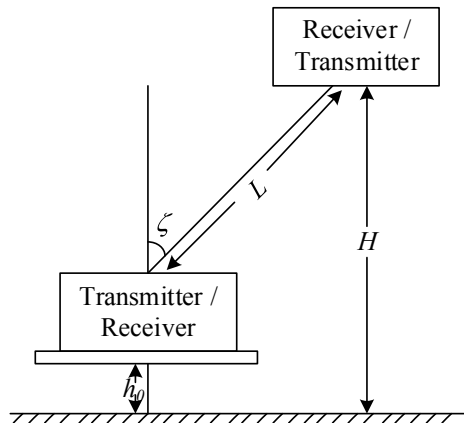
9.3.1 Beam Spread and Beam Drift

9.3.1.1 Beam Spread

As shown in Fig. 9.7, L is the beam transmission distance, and ζ is the zenith angle during the beam transmission. When the vertex angle is $\zeta = 90^\circ$, the beam is transmitted horizontally. When the vertex angle is $0^\circ \leq \zeta < 90^\circ$, the beam propagates obliquely. Slant path transmission can be divided into uplink and downlink transmission. When the beam is transmitted in the uplink, h_0 represents the height of the transmitter from the ground, and H represents the height of the receiver from the ground. When the beam is in downlink transmission, h_0 represents the height of the receiver from the ground, and H represents the height of the transmitter from the ground.

The atmospheric structure constant along the slant path varies with the height, and is affected by wind speed, temperature, humidity, and air pressure. For horizontal transmission, the typical value of the atmospheric structure constant is generally considered, but the study of slant propagation is more complex.

Fig. 9.7 Beam propagation model in the atmosphere



The cross-spectral density of partially coherent GSM beams at $z = 0$ is

$$W^{(0)}(\rho_{s1}, \rho_{s2}, 0) = A \exp\left(-\frac{|\rho_{s1}|^2 + |\rho_{s2}|^2}{4\sigma_s^2}\right) \times \exp\left(-\frac{|\rho_{s1} - \rho_{s2}|^2}{2\sigma_g^2}\right), \quad (9.51)$$

where ρ_{s1}, ρ_{s2} are the coordinate vectors of two points in the source plane. Parameters A, σ_s , and σ_g represent the light intensity, beam waist width, and coherence length of the light source, respectively. Based on the generalized Huygens–Fresnel principle, the cross-spectral density of a GSM beam at z is as follows [5]:

$$\begin{aligned} W(\rho_1, \rho_2, z) &= \left(\frac{k}{2\pi z}\right)^2 \iint d^2\rho_{s1} \iint d^2\rho_{s2} W^{(0)}(\rho_{s1}, \rho_{s2}, 0) \\ &\times \exp\left\{-\frac{ik}{2z}[(\rho_1 - \rho_{s1})^2 - (\rho_2 - \rho_{s2})^2]\right\} \\ &\times \langle \exp[\phi * (\rho_1, \rho_{s1}) + \phi(\rho_2, \rho_{s2})] \rangle, \end{aligned} \quad (9.52)$$

where ρ_1, ρ_2 are the coordinate vectors of two points in the Z plane, and $\langle \cdot \rangle$ is the operation of ensemble average.

$$\langle \exp[\phi * (\rho_1, \rho_{s1}) + \phi(\rho_2, \rho_{s2})] \rangle = \exp\left[-\frac{1}{2}D_\phi(\rho_{s1} - \rho_{s2}, \rho_1 - \rho_2)\right] \quad (9.53)$$

where $D_\phi(\rho_{s1} - \rho_{s2}, \rho_1 - \rho_2)$ is the wave structure function and can be expressed as Eq. (9.54) for oblique transmission [6, 7].

$$\begin{aligned} D_\phi(\mathbf{p}, \mathbf{Q}, z) &= 8\pi^2 k^2 \sec \zeta \int_{h_0}^H \int_0^\infty \kappa \phi_n(\kappa, h) \\ &\{1 - J_0[(1 - \xi)\mathbf{p} + \xi\mathbf{Q}|\kappa]\} d\kappa dh, \end{aligned} \quad (9.54)$$

where ζ represents the zenith angle of the beam emission, which is the angle between the transmission path and the vertical direction; and J_0 is the first kind of zero order Bessel function, which can be expanded as

$$J_0(x) = 1 - \left(\frac{x}{2}\right)^2 + \frac{1}{(2!)^2} \left(\frac{x}{2}\right)^4 - \dots, \quad |x| < \infty \quad (9.55)$$

Let $\mathbf{p} = \rho_1 - \rho_2$, $\mathbf{Q} = \rho_{s1} - \rho_{s2}$, and $\phi_n(\kappa, h) = \phi_n(\kappa)C_n^2(h)$, where $\phi_n(\kappa)$ represents the refractive index power spectrum, and $C_n^2(h)$ represents the atmospheric structure constant model. When the beam propagates upward, $\xi = 1 - \frac{h-h_0}{H-h_0}$; when the beam propagates downward, $\xi = \frac{h-h_0}{H-h_0}$. Under the condition of strong turbulence, Bessel function from Eq. (9.55) can be approximated by the first two terms [8]:

$$J_0[|(1 - \xi)\mathbf{p} + \xi \mathbf{Q}| \kappa] = 1 - \frac{1}{4} \times [(1 - \xi)\mathbf{p} + \xi \mathbf{Q}]^2 \kappa^2. \quad (9.56)$$

By substituting Eq. (9.56) into Eq. (9.54), the following can be obtained by product differentiation and simplification:

$$D_\phi(\mathbf{p}, \mathbf{Q}, z) = B_0(B_1 \mathbf{p}^2 + B_2 \mathbf{p} \cdot \mathbf{Q} + B_3 \mathbf{Q}^2), \quad (9.57)$$

where

$$B_0 = 2\pi^2 k^2 \sec \zeta \int_0^\infty 0.033 \kappa^3 (-\kappa^2 / \kappa_m^2) (\kappa^2 + \kappa_0^2)^{-11/6} d\kappa, \quad (9.58a)$$

$$B_1 = \int_{h_0}^H C_n^2(h) (1 - \xi)^2 dh, \quad (9.58b)$$

$$B_2 = 2 \int_{h_0}^H C_n^2(h) (1 - \xi) \xi dh, \quad (9.58c)$$

$$B_3 = \int_{h_0}^H C_n^2(h) \xi^2 dh. \quad (9.58d)$$

By substituting Eqs. (9.56) and (9.57) into Eq. (9.52)

$$\begin{aligned} W(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, z) &= \left(\frac{k}{2\pi z} \right)^2 \iint d^2 \boldsymbol{\rho}_{s1} \iint d^2 \boldsymbol{\rho}_{s2} A \exp \left(- \frac{|\boldsymbol{\rho}_{s1}|^2 + |\boldsymbol{\rho}_{s2}|^2}{4\sigma_s^2} \right) \\ &\times \exp \left(- \frac{|\boldsymbol{\rho}_{s1} - \boldsymbol{\rho}_{s2}|^2}{2\sigma_g^2} \right) \\ &\times \exp \left\{ - \frac{ik}{2z} [(\boldsymbol{\rho}_1 - \boldsymbol{\rho}_{s1})^2 - (\boldsymbol{\rho}_2 - \boldsymbol{\rho}_{s2})^2] \right\} \\ &\times \exp \left[- \frac{1}{2} B_0 (B_1 \mathbf{p}^2 + B_2 \mathbf{p} \cdot \mathbf{Q} + B_3 \mathbf{Q}^2) \right]. \end{aligned} \quad (9.59)$$

Let $\boldsymbol{\rho}_1 = \boldsymbol{\rho}_2 = \boldsymbol{\rho}$ in Eq. (9.59). Then the intensity distribution of a GSM beam in slant path transmission is as follows.

$$I(\boldsymbol{\rho}, z) = W(\boldsymbol{\rho}, \boldsymbol{\rho}, z) = \frac{A}{\Delta^2(z)} \exp \left(- \frac{|\boldsymbol{\rho}|^2}{2\sigma_s^2 \Delta^2(z)} \right), \quad (9.60)$$

where

$$\Delta^2(z) = 1 + \left(\frac{z}{k\sigma_s\delta} \right)^2 + \frac{2Mz^2}{k^2\sigma_s^2}, \quad (9.61a)$$

$$\frac{1}{\delta^2} = \frac{1}{4\sigma_s^2} + \frac{1}{\sigma_g^2}, \quad (9.61b)$$

$$M = \frac{1}{2}B_0B_3. \quad (9.61c)$$

When the atmospheric refractive index structure constant $C_n^2(h)$ is constant, $M = \frac{1}{3}\pi^2k^2z \int_0^\infty \kappa^3\phi_n(\kappa)d\kappa$. When propagating in free space, $M = 0$.

It can be seen that when the beam propagates from the plane $z = 0$ to z , the axial intensity of the beam decreases from A to $A/\Delta^2(z)$. The intensity of the beam is related to the refractive index structure constant, coherent length, and waist width of the beam. For the turbulence power spectrum model [9] in turbulence term M , the modified von Kármán spectrum taking both internal and external scales into account is adopted:

$$\phi_n(\kappa, h) = 0.033C_n^2(h) \exp(-\kappa^2/\kappa_m^2) \times (\kappa^2 + \kappa_0^2)^{-11/6}, \quad (9.62)$$

where $\kappa_m = 5.92/l_0$, $\kappa_0 \approx 2\pi/L_0$, l_0 and L_0 represent the internal and external scales of turbulence, respectively, and $C_n^2(h)$ represents the atmospheric refractive index structure model. Here, the height dependent model proposed by ITU-R in 2001 is used as the following expression:

$$C_n^2(h) = 8.148 \times 10^{-56} v_{\text{RMS}}^2 h^{10} e^{-h/1000} + 2.7 \times 10^{-16} e^{-h/1500} + C_0 e^{-h/100}, \quad (9.63)$$

where $v_{\text{RMS}}^2 = \sqrt{v_g^2 + 30.69v_g + 348.91}$ is the vertical path wind speed, v_g is the near-surface wind speed, and C_0 is the near-surface atmospheric refractive index structure constant. C_0 is ranged 10^{-17} – $10^{-13} \text{ m}^{-2/3}$ at medium turbulence. It is assumed that the refractive index structure constant of the near-surface atmosphere is $C_0 = 1.7 \times 10^{-14} \text{ m}^{-2/3}$, and the intensity and propagation of the beam are discussed. For horizontal transmission, h in Eq. (9.63) becomes transmitter height. The normalized intensity of the beam is defined as [10]

$$I^N(\rho, z) = \frac{I_{ts}(\rho, z)}{I_{fs}(\rho, z)|_{\rho=0, \sigma_g \rightarrow \infty, M=0}}, \quad (9.64)$$

where $I_{ts}(\rho, z)$ is the intensity distribution of the beam propagating obliquely in atmospheric turbulence, and $I_{fs}(\rho, z)|_{\rho=0, \sigma_g \rightarrow \infty, M=0}$ is the axial intensity of the

fully coherent beam propagating obliquely in free space. The normalized intensity of the beam is [11]

$$I^N(\boldsymbol{\rho}, z) = \frac{1 + (z/2k\sigma_s^2)^2}{\Delta^2(z)} \exp\left[-\frac{\rho^2}{2\sigma_s^2 \Delta^2(z)}\right]. \quad (9.65)$$

The expression of partially coherent light propagation in free space is as follows.

$$I_N(\boldsymbol{\rho}, z) = \frac{1 + \left(\frac{z}{2k\sigma_s^3}\right)^2}{1 + \left(\frac{z}{k\sigma_s^2\delta}\right)^2} \exp\left\{-\frac{\rho^2}{2\sigma_s^2 \left[1 + \left(\frac{z}{k\sigma_s^2\delta}\right)^2\right]}\right\} \quad (9.66)$$

The expression of partially coherent light propagation in atmospheric turbulence is

$$I_N(\boldsymbol{\rho}, z) = \frac{1 + \left(\frac{z}{2k\sigma_s^3}\right)^2}{1 + \left(\frac{z}{k\sigma_s^2\delta}\right)^2 + \frac{2M_3 z^2}{k^2 \sigma_s^2}} \exp\left\{-\frac{\rho^2}{2\sigma_s^2 \left[1 + \left(\frac{z}{k\sigma_s^2\delta}\right)^2 + \frac{2M_3 z^2}{k^2 \sigma_s^2}\right]}\right\}. \quad (9.67)$$

The expression of fully coherent light propagation in free space is

$$I_N(\boldsymbol{\rho}, z) = \exp\left\{-\frac{\rho^2}{2\sigma_s^2 \left[1 + \left(\frac{z}{2k\sigma_s^3}\right)^2\right]}\right\}. \quad (9.68)$$

The expression of fully coherent light propagation in atmospheric turbulence is

$$I_N(\boldsymbol{\rho}, L) = \frac{1 + \left(\frac{z}{2k\sigma_s^3}\right)^2}{1 + \left(\frac{z}{2k\sigma_s^3}\right)^2 + \frac{2M_3 z^2}{k^2 \sigma_s^2}} \exp\left\{-\frac{\rho^2}{2\sigma_s^2 \left[1 + \left(\frac{z}{2k\sigma_s^3}\right)^2 + \frac{2M_3 z^2}{k^2 \sigma_s^2}\right]}\right\}. \quad (9.69)$$

According to the definition of the root mean square beam width of the beam spread [12]:

$$\omega(z) = \sqrt{\frac{\int I(\boldsymbol{\rho}, z) |\boldsymbol{\rho}|^2 d^2 \boldsymbol{\rho}}{\int I(\boldsymbol{\rho}, z) d^2 \boldsymbol{\rho}}}. \quad (9.70)$$

Substituting Eq. (9.61) into Eq. (9.70), we obtain [13]

$$\omega(z) = [2\sigma_s^2 \Delta^2(z)]^{\frac{1}{2}} = \left[2\sigma_s^2 + \frac{2}{k^2 \delta^2} z^2 + \frac{4M}{k^2} z^2 \right]^{\frac{1}{2}}. \quad (9.71)$$

The first two terms represent the beam spread caused by free space diffraction, and the third term represents the beam propagation caused by turbulence. It can be seen that the beam width $\omega(z)$ increases from $\sqrt{2\sigma_s^2}$ to $\sqrt{2\sigma_s^2 \Delta(z)}$ when the beam propagates from $z = 0$ plane to z . To intuitively analyze the influence of turbulence on beam propagation, the relative beam width is defined as the ratio of beam width in turbulence to beam width in free space, and the expression is [14]

$$\omega_r(z) = \left(1 + \frac{\frac{4M}{k^2} z^2}{2\sigma_s^2 + \frac{2}{k^2 \delta^2} z^2} \right)^{\frac{1}{2}}. \quad (9.72)$$

At the same time, the angular spread of the beam can be calculated according to the definition of angular spread [15]:

$$\theta = \lim_{z \rightarrow \infty} \frac{\omega(z)}{z} = \left(\frac{2}{k^2 \delta^2} + \frac{4M}{k^2} \right)^{\frac{1}{2}} \quad (9.73)$$

The first term of Eq. (9.73) represents the angular spread induced in free space, which is related to the beam parameter term δ , and the second term represents the angular spread caused by turbulence, which is related to the turbulence term M . The expression of the relative angular spread of the beam can be obtained as [16]

$$\theta_r = \frac{\theta_{\text{turb}}}{\theta_{\text{free}}} = (1 + 2M\delta^2)^{1/2} \quad (9.74)$$

where θ_{turb} represents the angular spread of the beam in turbulence, and θ_{free} represents the angular spread of the beam in free space.

9.3.1.2 Beam Drift

Beam drift is primarily due to the refraction of large-scale turbulent vortices. When the diameter of the beam is much smaller than the turbulence scale, the main effect of the turbulence is that the beam deflects randomly and deviates from the original propagation direction. At this time, the propagation direction of the beam or the projection position on the receiving surface is random; that is, the beam drift occurs as shown in Fig. 9.8. The drift of partially coherent GSM beams in atmospheric turbulence is analyzed as follows [17].

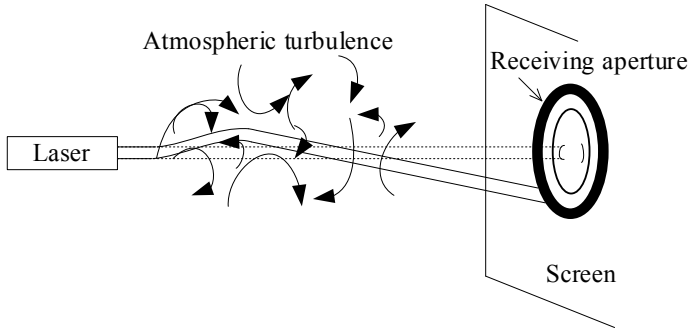


Fig. 9.8 Diagram of beam drift [17]

Beam drift is usually described by the change of the centroid position of the spot. The centroid of the spot is defined as [18]

$$\boldsymbol{\rho} = \frac{\iint \boldsymbol{\rho} I(\boldsymbol{\rho}) d\boldsymbol{\rho}}{\iint I(\boldsymbol{\rho}) d\boldsymbol{\rho}}. \quad (9.75)$$

Namely,

$$x_c = \frac{\iint x I(x, y) dx dy}{\iint I(x, y) dx dy}, \quad y_c = \frac{\iint y I(x, y) dx dy}{\iint I(x, y) dx dy}. \quad (9.76)$$

Then, the drift variance of the centroid is

$$\sigma_\rho^2 = \langle \rho_c^2 \rangle = \frac{\iint \iint (\boldsymbol{\rho}_1 \cdot \boldsymbol{\rho}_2) I(\boldsymbol{\rho}_1) I(\boldsymbol{\rho}_2) d^2 \boldsymbol{\rho}_1 d^2 \boldsymbol{\rho}_2}{[\iint I(\boldsymbol{\rho}) d\boldsymbol{\rho}]^2}. \quad (9.77)$$

9.3.2 Drift and Spread of a Horizontally Propagating Beam

9.3.2.1 Drift of a Horizontally Propagating Beam

Andrews uses Rytov approximation to obtain the average intensity distribution on the axis [19]:

$$\langle I(0, L) \rangle = \frac{w_0^2}{W_{LT}^2} = \frac{w_0^2}{W^2(1 + T)}, \quad (9.78)$$

where $\langle \cdot \rangle$ represents ensemble average, W_{LT} represents long-term beam spread, and W_{LT}^2 is written as [19]

$$W_{LT}^2 = \underbrace{W^2}_{\text{diffract}} + \underbrace{W^2 T_{ss}}_{\text{sm-scale spread}} + \underbrace{W^2 T_{LS}}_{\text{large-scale beam wander}}. \quad (9.79)$$

Here, we decompose T in Eq. (9.78) into $T = T_{ss} + T_{LS}$: T_{ss} and T_{LS} are the beam spread caused by small scale and beam drift caused by large scale, respectively. Therefore, the last term of Eq. (9.79) can be interpreted as beam drift ρ_c , and the sum of the first and second terms is the short-term beam spread.

Andrews and Philips used Rytov approximation to process the last term in Eq. (9.79) to obtain a classical drift variance model [19]:

$$\begin{aligned} \langle \rho_c^2 \rangle &= W^2 T_{LS} = 4\pi^2 k^2 w^2(L) \int_0^L \int_0^\infty \kappa \Phi_n(\kappa) H_{LS}(\kappa, z) \\ &\times \left[1 - e^{-\Lambda_p L \kappa^2 (1-z/L)^2 / k} \right] d\kappa dz \end{aligned} \quad (9.80)$$

where Λ_p is the beam output parameter, $\Lambda_p = 2L/kw^2(L)$, and $w(L)$ is the beam radius at the receiving end in free space. Since the main factor causing the beam drift is the large-scale turbulence, a large-scale filtering function is introduced. The function only contributes to the beam spreading when the size of the random nonuniform turbulence vortex is equal to or greater than the beam size:

$$H_{LS}(\kappa, z) = \exp[-\kappa^2 w^2(z)]. \quad (9.81)$$

To emphasize the natural characteristics of beam drift, the last term in Eq. (9.80) is approximated by geometrical optics.

$$1 - e^{-\Lambda_p L \kappa^2 (1-z/L)^2 / k} \cong \frac{\Lambda_p L \kappa^2 (1-z/L)^2}{k}, \quad L \kappa^2 / k \ll 1. \quad (9.82)$$

Using Eqs. (9.81) and (9.82) to simplify Eq. (9.80), we obtain

$$\langle \rho_c^2 \rangle = 8\pi^2 L^2 \int_0^L \int_0^\infty \kappa^3 \Phi_n(\kappa) \exp[-\kappa^2 w^2(z)] \times \left(1 - \frac{z}{L}\right)^2 d\kappa dz. \quad (9.83)$$

Since the beam drift is primarily caused by the scale of the turbulence, the large-scale exponential spectrum is considered:

$$\Phi_n(\kappa) = 0.033 C_n^2 \kappa^{-11/3} [1 - \exp(-\kappa^2 / \kappa_0^2)]. \quad (9.84)$$

where κ_0 is the spatial wave number corresponding to the outer scale L_0 and $\kappa_0 = 1/L_0$. C_n^2 is the refractive index structure constants, which represents the strength of the turbulence and is constants in horizontal transmission.

Equations (9.81)–(9.84) are substituted into Eq. (9.83), and after the integral calculation, we obtain

$$\langle \rho_c^2 \rangle = 7.25 C_n^2 L^2 \int_0^L \left(1 - \frac{z}{L}\right)^2 w^{-1/3}(z) \left[1 - \left(\frac{\kappa_0^2 w^2(z)}{1 + \kappa_0^2 w^2(z)}\right)\right]^{1/6} dz, \quad (9.85)$$

where $w(z)$ is the spreading radius of partially coherent GSM beams in atmospheric turbulence.

9.3.2.2 Beam Spreading Radius

At $z = 0$, the complex amplitude distribution of a Gaussian beam in free space is [21]

$$U(\mathbf{r}, 0) = \exp\left[-\left(\frac{1}{w_0^2} + \frac{jk}{2F_0}\right)\mathbf{r}^2\right]. \quad (9.86)$$

where $r = |\mathbf{r}| = (x^2 + y^2)^{1/2}$ represents a two-dimensional radial vector perpendicular to the transmission direction, w_0 and F_0 are the initial beam radius and the curvature radius of the phase front, respectively, and $k = 2\pi/\lambda$ is the number of light waves in the free space.

According to the generalized Huygens–Fresnel principle [22], the cross-spectral density function of partially coherent GSM beams propagating in atmospheric turbulence is obtained as [23]

$$\begin{aligned} W(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2; L) &= \langle U(\boldsymbol{\rho}_1, L) U^*(\boldsymbol{\rho}_2, L) \rangle \\ &= \frac{1}{(\lambda L)^2} \iint d^2\mathbf{r}_1 d^2\mathbf{r}_2 W(\mathbf{r}_1, \mathbf{r}_2, 0) \times \langle \exp[\psi(\mathbf{r}_1, \boldsymbol{\rho}_1) + \psi^*(\mathbf{r}_2, \boldsymbol{\rho}_2)] \rangle \\ &\quad \times \exp\left\{\frac{ik}{2L}[(\boldsymbol{\rho}_1 - \mathbf{r}_1)^2 - (\boldsymbol{\rho}_2 - \mathbf{r}_2)^2]\right\}, \end{aligned} \quad (9.87)$$

where $W(\mathbf{r}_1, \mathbf{r}_2, 0)$ is the cross-spectral density function of the transmitter. Considering the phase diffusion at the emission aperture, the emission field can be expressed as

$$\tilde{U}(\mathbf{r}, 0) = U(\mathbf{r}, 0) \exp[i\varphi_d(\mathbf{r})], \quad (9.88)$$

where $\exp[i\varphi_d(\mathbf{r})]$ describes the random disturbance caused by phase diffusion. Assuming that the spatial random phase is independent and that it is a Gaussian

function, the cross-spectral density function of the GSM beam at the transmitter can be expressed as

$$\begin{aligned}
 W_0(\mathbf{r}_1, \mathbf{r}_2, 0) &= \langle \tilde{U}(\mathbf{r}_1, 0) \tilde{U}^*(\mathbf{r}_2, 0) \rangle \\
 &= U(\mathbf{r}_1, 0) U^*(\mathbf{r}_2, 0) \\
 &\quad \times \langle \exp[i\psi_{d1}(\mathbf{r}_1)] \exp[i\psi_{d2}(\mathbf{r}_2)] \rangle \\
 &= U(\mathbf{r}_1, 0) U^*(\mathbf{r}_2, 0) \exp[-(\mathbf{r}_1 - \mathbf{r}_2)^2 / 2\sigma_g^2], \quad (9.89)
 \end{aligned}$$

where σ_g^2 is the Gaussian variance of the ensemble average describing the random phase, which represents the spatial coherence of the emission source. Let

$$\begin{aligned}
 \mathbf{r}_c &= (\mathbf{r}_1 + \mathbf{r}_2)/2, \quad \mathbf{r}_d = \mathbf{r}_1 - \mathbf{r}_2, \\
 \boldsymbol{\rho}_c &= (\boldsymbol{\rho}_1 + \boldsymbol{\rho}_2)/2, \quad \boldsymbol{\rho}_d = \boldsymbol{\rho}_1 - \boldsymbol{\rho}_2, \quad (9.90)
 \end{aligned}$$

and use Eqs. (9.88) and (9.90) into Eq. (9.90), we obtain

$$W_0(\mathbf{r}_c, \mathbf{r}_d, 0) = \exp \left\{ -\frac{1}{w_0^2} \left[\frac{1}{2} (\mathbf{r}_d^2 + 4\mathbf{r}_c^2) \right] - \frac{ik}{2F_0} (2\mathbf{r}_d \cdot \mathbf{r}_c) - \frac{\mathbf{r}_d^2}{2\sigma_g^2} \right\}. \quad (9.91)$$

According to Eq. (9.90), we can also write Eq. (9.91) as

$$\exp \left\{ \frac{jk}{2L} [(\boldsymbol{\rho}_1 - \mathbf{r}_1)^2 - (\boldsymbol{\rho}_2 - \mathbf{r}_2)^2] \right\} = \exp \left\{ \frac{jk}{L} [(\mathbf{r}_c - \boldsymbol{\rho}_c) \cdot (\mathbf{r}_d - \boldsymbol{\rho}_d)^2] \right\}. \quad (9.92)$$

The complex phase cross-correlation function of spherical wave can be expressed as [22]

$$\langle \exp(\psi(\mathbf{r}_1, \boldsymbol{\rho}_1) + \psi^*(\mathbf{r}_2, \boldsymbol{\rho}_2)) \rangle = \exp[-D_\psi/2], \quad (9.93)$$

where D_ψ is the phase structure function, which can be expressed as

$$D_\psi(\mathbf{r}_c, \boldsymbol{\rho}_c) \cong 8\pi^2 k^2 L \int_0^1 \int_0^\infty \kappa \Phi_n(\kappa) \times [1 - J_0(|(1-\xi)\mathbf{r}_d + \xi\boldsymbol{\rho}_d|\kappa)] d\kappa d\xi, \quad (9.94)$$

where $\Phi_n(\kappa)$ is the power spectrum model of refractive index fluctuation. By substituting Kolmogorov spectrum into Eq. (9.94),

$$D_\psi(\mathbf{r}_d, \boldsymbol{\rho}_d) \cong 2.92k^2L \int_0^1 dt C_n^2 |\mathbf{r}_d t + (1-t)\boldsymbol{\rho}_d|^{5/3} \quad (9.95)$$

$$\cong 2\{[\mathbf{r}_d^2 + \mathbf{r}_d \cdot \boldsymbol{\rho}_d + \boldsymbol{\rho}_d^2]/\rho_0^2\}.$$

From Eqs. (9.93) and (9.95), the following can be obtained:

$$\langle \exp(\psi(\mathbf{r}_1, \boldsymbol{\rho}_1) + \psi^*(\mathbf{r}_2, \boldsymbol{\rho}_2)) \rangle = \exp\left[\frac{-1}{\rho_0^2}(\mathbf{r}_d^2 + \mathbf{r}_d \cdot \boldsymbol{\rho}_d + \rho_d^2)\right], \quad (9.96)$$

where $\rho_0 = (0.55C_n^2k^2z)^{-3/5}$ is the coherent length of a spherical wave in atmospheric turbulence, and C_n^2 is the structure constant of the atmospheric refractive index. When the transmission distance $z = L$, the cross-spectral density function at the receiver can be written as

$$W(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2; L) = \frac{1}{(\lambda L)^2} \iint d^2\mathbf{r}_d \iint d^2\mathbf{r}_c W_0(\mathbf{r}_1, \mathbf{r}_2; 0)$$

$$\exp\langle \exp[\psi(\mathbf{r}_1, \boldsymbol{\rho}_1) + \psi^*(\mathbf{r}_2, \boldsymbol{\rho}_2)] \rangle \quad (9.97)$$

$$\times \exp\left\{\frac{ik}{2L}[(\boldsymbol{\rho}_1 - \mathbf{r}_1)^2 - (\boldsymbol{\rho}_2 - \mathbf{r}_2)^2]\right\}.$$

Using Eqs. (9.92)–(9.96) into Eq. (9.97), we obtain

$$W(\boldsymbol{\rho}_c, \boldsymbol{\rho}_d; L) = \frac{1}{(\lambda L)^2} \iint d^2\mathbf{r}_d \iint d^2\mathbf{r}_c \exp\left(\frac{-2\mathbf{r}_c^2}{w_0^2}\right)$$

$$\times \exp\left[\frac{-ik\mathbf{r}_c \cdot \mathbf{r}_d}{F_0} + \frac{ik\mathbf{r}_c \cdot (\mathbf{r}_d - \boldsymbol{\rho}_d)}{L}\right] \quad (9.98)$$

$$\times \exp\left[-\frac{\mathbf{r}_d^2}{2w_0^2} - \frac{\mathbf{r}_d^2}{2\sigma_g^2} - \frac{\mathbf{r}_d^2 + \mathbf{r}_d \cdot \boldsymbol{\rho}_d + \rho_d^2}{\rho_0^2} - \frac{ik\boldsymbol{\rho}_c \cdot (\mathbf{r}_d - \boldsymbol{\rho}_d)}{L}\right].$$

By integrating Eq. (9.98):

$$W(\boldsymbol{\rho}_c, \boldsymbol{\rho}_d, L) = \langle I(\boldsymbol{\rho}_c, \boldsymbol{\rho}_d, L) \rangle$$

$$= \frac{w_0^2}{w_\zeta^2(L)} \exp\left\{-\rho_d^2\left(\frac{1}{\rho_0^2} + \frac{1}{2w_0^2\Lambda_0^2}\right) + \frac{2j\boldsymbol{\rho}_c \cdot \boldsymbol{\rho}_d}{w_0^2\Lambda_0}\right\} \exp\left[\frac{-2\rho_c^2}{w_\zeta^2(L)}\right] \quad (9.99)$$

$$\times \exp\left[\frac{-(j\phi)^2\rho_d^2}{2w_\zeta^2(L)}\right] \exp\left[\frac{-2j\phi\boldsymbol{\rho}_c \cdot \boldsymbol{\rho}_d}{w_\zeta^2(L)}\right],$$

where $\phi \equiv \frac{\Theta_0}{\Lambda_0} - \Lambda_0 \frac{w_0^2}{\rho_0^2}$, $w_\zeta(z)$ is the effective radius of the beam in atmospheric turbulence, that is, the spreading radius, and $F_\zeta(z)$ is the radius of curvature in the wavefront. These can be expressed as

$$w_\zeta(L) = w_0(\Theta_0^2 + \zeta \Lambda_0^2)^{1/2}, \quad \zeta = \zeta_s + \frac{2w_0^2}{\rho_0^2}, \quad (9.100)$$

$$F_\zeta(L) = \frac{L(\Theta_0^2 + \zeta \Lambda_0^2)}{\phi \Lambda_0 - \zeta \Lambda_0^2 - \Theta_0^2}, \quad \phi = \frac{\Theta_0}{\Lambda_0} - \Lambda_0 \frac{w_0^2}{\rho_0^2}. \quad (9.101)$$

In Eqs. (9.100) and (9.101), $\Theta_0 = (F_0 - L)/F_0$, $\Lambda_0 = 2L/kw_0^2$ is the effective beam parameter, ζ is the global coherence parameter, and $\zeta_s = 1 + w_0^2/\sigma_g^2$ is the coherence parameter of the light source, where $2\sigma_g^2 = l_c^2$ [23]. When the transmission distance $z = L$, Eq. (9.101) is the expression of $w(z)$.

9.3.3 Drift and Spread of a Slant Propagating Beam

The oblique propagation diagram of beam in atmospheric turbulence is shown in Fig. 9.9. According to the geometric relationship in the figure, it can be seen that [17]

$$\cos \theta = H/L = h/z, \quad (h \leq H, z \leq L), \quad (9.102)$$

where θ is the zenith angle, H is the vertical height of the receiver from the ground, h_0 is the vertical height from the transmitter to the ground, L is the distance between the transmitter and the receiver, z is the transmission distance with a size of 0 to L , and h is the vertical height of transmission with a size of 0 to H .

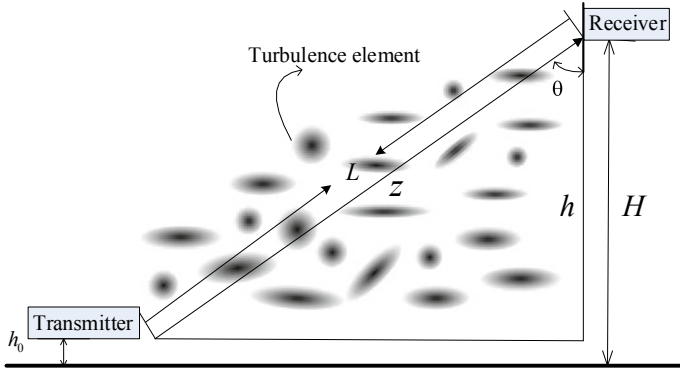


Fig. 9.9 Schematic diagram of ramp transmission [17]

9.3.3.1 Drift of a Slant Propagating Beam

In the slant propagation, the classical drift variance model in horizontal propagation continues to be used [9]:

$$\begin{aligned} \langle \rho_c^2 \rangle = W^2 T_{LS} = & 4\pi^2 k^2 w^2(L) \int_0^L \int_0^\infty \kappa \phi_n(\kappa) H_{LS}(\kappa, z) \\ & \times \left[1 - e^{-\Lambda_p L \kappa^2 \xi^2 / k} \right] d\kappa dz. \end{aligned} \quad (9.103)$$

The atmospheric structure constant $C_n^2(h)$ and refractive index fluctuation power spectral density function $\phi_n(\kappa, h)$ vary with altitude. Introducing a normalized variable $\xi = h/H = z/L$, the beam drift in slant propagation is written as

$$\langle \rho_c^2 \rangle = 8\pi^2 L^2 \int_0^L \int_0^\infty \kappa^3 \phi_n(\kappa, z) \exp[-k^2 w^2(z)] \times \left(1 - \frac{z}{L}\right)^2 d\kappa dz. \quad (9.104)$$

Equation (9.104) is the classical drift variance model of slant propagation, which is consistent with horizontal propagation. As long as the atmospheric refractive index fluctuation power spectrum and the beam propagation radius are known, the beam drift characteristics can be studied. A large-scale exponential spectrum [20] is selected

$$\Phi_n(\kappa) = 0.033 C_n^2(z) \kappa^{-11/3} [1 - \exp(-\kappa^2 / \kappa_0^2)], \quad (9.105)$$

where $C_n^2(z)$ is used to describe the structure constant of atmospheric refractive index in turbulent atmosphere. Here, the atmospheric structure constant model H-V with altitude change published by ITU-R is adopted. According to the distance normalization variable $\xi = h/H = z/L$, it can be written as

$$\begin{aligned} C_n^2(z) &= C_n^2(\xi L) = C_n^2(\xi H / \cos \theta) \\ &= 8.148 \times 10^{-56} v_{\text{RMS}}^2 \left(\xi H / \cos \theta^{10} \right) \times \exp(-\xi H / \cos \theta / 1000) + 2.7 \times 10^{-16} \\ &\quad \times \exp(-\xi H / \cos \theta / 1500) + C_{n0}^2 \exp(-\xi H / \cos \theta / 100), \end{aligned} \quad (9.106)$$

where $v_{\text{RMS}}^2 = \sqrt{v_g^2 + 30.69 v_g + 348.91}$ is the vertical path wind speed, v_g is the near-surface wind speed (generally $v_g = 21 \text{ m/s}$), and C_{n0}^2 is the atmospheric structure constant near the ground (generally taken as $C_{n0}^2 = 1.7 \times 10^{-14}$). By introducing Eq. (9.106) into Eq. (9.105),

$$\langle \rho_c^2 \rangle = 7.25 L^3 \int_0^1 C_n^2(\xi L) (1 - \xi)^2 w^{-1/3}(\xi L) \left[1 - \left(\frac{\kappa_0^2 w^2(z)}{1 + \kappa_0^2 w^2(z)} \right)^{1/6} \right] d\xi. \quad (9.107)$$

9.3.3.2 Spread of a Slant Propagating Beam

In slant propagation, the phase structure function D_ψ in Eq. (9.93) changes to

$$D_\psi(\boldsymbol{\rho}_c, \boldsymbol{\rho}_d) \cong 8\pi^2 k^2 L \int_0^1 \int_0^\infty \kappa \Phi_n(\kappa, z) \times [1 - J_0(|(1 - \xi)\mathbf{r}_d + \xi\boldsymbol{\rho}_d|\kappa)] d\kappa d\xi. \quad (9.108)$$

By substituting Kolmogorov spectrum into Eq. (9.108),

$$\begin{aligned} D_\psi(\mathbf{r}_d, \boldsymbol{\rho}_d) &\cong 2.92 k^2 L \int_0^1 dt C_n^2(\xi L) |\mathbf{r}_d t + (1 - t)\boldsymbol{\rho}_d|^{5/3} \\ &\cong 2 \{ [\mathbf{r}_d^2 + \mathbf{r}_d \cdot \boldsymbol{\rho}_d + \boldsymbol{\rho}_d^2] / \rho_0^2 \}. \end{aligned} \quad (9.109)$$

Similarly, from Eqs. (9.107) and (9.109)

$$\langle \exp(\psi(\mathbf{r}_1, \boldsymbol{\rho}_1) + \psi^*(\mathbf{r}_2, \boldsymbol{\rho}_2)) \rangle = \exp \left[\frac{-1}{\rho_0^2} (r_d^2 + \mathbf{r}_d \cdot \boldsymbol{\rho}_d + \rho_d^2) \right]. \quad (9.110)$$

The coherent length of a spherical wave in atmospheric turbulence during the slant propagation is as follows:

$$\rho'_0 = [1.46 k^2 \int_0^L C_n^2(z) (1 - z/L)^{5/3} dz]^{-3/5}, \quad (9.111)$$

where $C_n^2(z)$ is the atmospheric refractive index structure constant describing turbulence intensity. Equations (9.109)–(9.111) are integrated into Eq. (9.98):

$$\begin{aligned} W(\boldsymbol{\rho}_c, \boldsymbol{\rho}_d, L) &= \langle I(\boldsymbol{\rho}_c, \boldsymbol{\rho}_d, L) \rangle \\ &= \frac{w_0^2}{w_\xi^2(L)} \exp \left\{ -\boldsymbol{\rho}_d^2 \left(\frac{1}{\rho_0^2} + \frac{1}{2w_0^2 \Lambda_0^2} \right) + \frac{2j\boldsymbol{\rho}_c \cdot \boldsymbol{\rho}_d}{w_0^2 \Lambda_0} \right\} \exp \left[\frac{-2\boldsymbol{\rho}_c^2}{w_\xi^2(L)} \right] \\ &\quad \times \exp \left[\frac{-(j\phi)^2 \boldsymbol{\rho}_d^2}{2w_\xi^2(L)} \right] \exp \left[\frac{-2j\phi \boldsymbol{\rho}_c \cdot \boldsymbol{\rho}_d}{w_\xi^2(L)} \right], \end{aligned} \quad (9.112)$$

where $\phi \equiv \frac{\Theta_0}{\Lambda_0} - \Lambda_0 \frac{w_0^2}{\rho_0'^2}$, and $w_\zeta(z)$ and $F_\zeta(z)$ are the beam propagation radius and phase front curvature radius in oblique atmospheric turbulence, respectively, which are expressed as

$$w_\zeta(L) = w_0(\Theta_0^2 + \zeta \Lambda_0^2)^{1/2}, \quad \zeta = \zeta_s + \frac{2w_0^2}{\rho_0'^2}, \quad (9.113)$$

$$F_\zeta(L) = \frac{L(\Theta_0^2 + \zeta \Lambda_0^2)}{\phi \Lambda_0 - \zeta \Lambda_0^2 - \Theta_0^2}, \quad \phi = \frac{\Theta_0}{\Lambda_0} - \Lambda_0 \frac{w_0^2}{\rho_0'^2}. \quad (9.114)$$

Equation (9.113) is consistent with horizontal propagation, and $\Theta_0 = (F_0 - L)/F_0$. $\Lambda_0 = 2L/kw_0^2$ is the effective beam parameter; ζ is the global coherence parameter, $\zeta_s = 1 + w_0^2/\sigma_g^2$ is the source coherence parameter [22], where $2\sigma_g^2 = l_c^2$. When the transmission distance $z = L$, Eq. (9.113) can be simplified as

$$\begin{aligned} w(z) &= w_0 \left[\left(1 - \frac{z}{F_0}\right)^2 + \left(\zeta_s + \frac{2w_0^2}{\rho_0'^2}\right) \left(\frac{2z}{kw_0^2}\right)^2 \right]^{1/2} \\ &= w_0 \left[\left(1 - \frac{z}{F_0}\right)^2 + \zeta_s \left(\frac{2z}{kw_0^2}\right)^2 + \frac{2w_0^2}{\rho_0'^2} \left(\frac{2z}{kw_0^2}\right)^2 \right]^{1/2}. \end{aligned} \quad (9.115)$$

Equation (9.115) is the beam spreading radius of a partially coherent GSM beam in atmospheric turbulence in slant propagation. In the formula, the first two terms are the free space diffraction spreading, and the last term is the spreading of turbulence effect. F_0 is the phase curvature of wavefront, z is the transmission distance variable, and ζ_s is the coherent parameter of the light source, which always satisfies $\zeta_s \geq 1$ (the beam at $\zeta_s = 1$ is fully coherent, $\zeta_s > 1$ is partially coherent, $\zeta_s \rightarrow \infty$ is completely incoherent); l_c is the spatial coherence length: the smaller l_c is, the worse the coherence is. When $l_c \rightarrow \infty$, the light is completely coherent, and when $l_c \rightarrow 0$, the light is completely incoherent.

9.3.4 Fluctuation of Angle of Arrival

For the quasi-monochromatic field $U(\vec{r})$ with frequency ω propagating from plane $z = 0$ to half space $z > 0$ in turbulent atmosphere, a random function of the position of refractive index $n(\vec{r})$ in half space is assumed, where $n(\vec{r}) = 1 + n_1(\vec{r})$ and $n_1 \ll 1$. Then the field can be described by the following scalar wave equation [24]:

$$\{\nabla^2 + k^2[1 + 2n_1(\vec{r})]\}U(\vec{r}) = 0 \quad (9.116)$$

Assuming that the incident field $U(\vec{r}, 0)$ is close to the z axis, the solution of Eq. (9.116) is expressed as

$$U(\vec{\rho}, 0) = \iint U(\vec{r}, 0) G_0(\vec{r}, \vec{r}, z) d^2\vec{r}, \quad (9.117)$$

where $G_0(\vec{r}, \vec{\rho}, z) = \frac{-ik}{2\pi z} \exp\left[ikz + \frac{ik}{2z}|\vec{\rho} - \vec{r}|^2 + \psi(\vec{r}, \vec{\rho})\right]$ is Green's function term, and $\psi(\vec{r}, \vec{\rho})$ is the random perturbation of a complex phase of a spherical wave in turbulent atmosphere, which depends on medium characteristics. Then Eq. (9.116) is expressed as

$$U(\vec{\rho}, z) = \frac{-ik}{2\pi z} \exp(ikz) \iint d^2r U(\vec{r}, 0) \exp\left[\frac{ik}{2z}|\vec{\rho} - \vec{r}|^2 + \psi(\vec{r}, \vec{\rho})\right]. \quad (9.118)$$

According to Huygens–Fresnel principle, spherical waves can be generated at every point of the wavefront, and the envelope of these spherical waves constitutes a new wavefront. Equation (9.118) reflects the influence of atmospheric turbulence on the transmission through $\psi(\vec{r}, \vec{\rho})$, which is called the generalized Huygens–Fresnel principle expression and is suitable for weak and strong turbulence fluctuation regions.

The propagation characteristics of light waves in turbulent atmosphere are reflected in the complex cross-correlation function $\Gamma(\vec{\rho}_1, \vec{\rho}_2; \tau) = \langle E(\vec{\rho}_1; \tau) E^*(\vec{\rho}_2; t + \tau) \rangle$, where $\langle \cdot \rangle$ represents the overall average, and E is the optical frequency electric field. The cross-spectral density function $W(\vec{\rho}_1, \vec{\rho}_2; \omega) = \langle U(\vec{\rho}_1; \omega) U^*(\vec{\rho}_2; \omega) \rangle$ can be used to study the propagation of partially coherent fields in turbulent atmosphere [25]; this is the time Fourier transform of $\Gamma(\vec{\rho}_1, \vec{\rho}_2; \tau)$, and ω is the frequency of light wave. The cross-spectral density function obeys the Helmholtz equation and is a measure of the correlation between the fluctuations of two field components at the same frequency [26]. If the field is strictly monochromatic or the band is sufficiently narrow, then

$$\frac{|\vec{\rho}_2 - \vec{\rho}_1|}{c} \ll \frac{1}{\Delta\omega}. \quad (9.119)$$

Then, the two characteristic quantities obtain the same result. In the laser communication system, if the typical value $|\vec{\rho}_2 - \vec{\rho}_1|$ is of the order of 10 cm, $\Delta\omega < 1$ GHz is required.

Here, we consider the change of $W(\vec{\rho}_1, \vec{\rho}_2; \omega = \omega_0)$, and ω_0 is the center frequency of the monochromatic Gaussian beam. Based on the generalized Huygens–Fresnel principle, the cross-spectral density function [27, 28] is as follows:

$$\begin{aligned} W(\vec{\rho}_1, \vec{\rho}_2; z) &= \langle U(\vec{\rho}_1; z) U^*(\vec{\rho}_2; z) \rangle \\ &= \frac{1}{(\lambda z)^2} \iiint \int d^2\vec{r}_1 d^2\vec{r}_2 W(\vec{r}_1, \vec{r}_2) \langle \exp[\psi(\vec{r}_1, \vec{\rho}_1) + \psi^*(\vec{r}_2, \vec{\rho}_2)] \rangle \times \\ &\quad \exp\left\{\frac{ik}{2z}[(\vec{\rho}_1 - \vec{r}_1)^2 - (\vec{\rho}_2 - \vec{r}_2)^2]\right\}, \end{aligned} \quad (9.120)$$

where $W_0(\vec{r}_1, \vec{r}_2)$ is the cross-spectral density function at the transmitter. For partially coherent beams, a phase scatterer is placed at the aperture of the laser transmitter:

$$\tilde{U}(\vec{r}, 0) = U(\vec{r}, 0) \exp[i\varphi_d(\vec{r})], \quad (9.121)$$

where $U(\vec{r}, 0) = \exp\left[-\left(\frac{1}{w_0^2} + \frac{ik}{2F_0}\right)r^2\right]$, and $\exp[i\varphi_d(\vec{r})]$ are small random perturbation terms caused by phase diffusion.

Assuming that the global mean value of a part of the independent random phases caused by diffusion is Gaussian and only depends on the propagation distance of the beam rather than the actual diffusion path, the cross-spectral density function of the transmitter can be expressed as

$$\begin{aligned} W_0(\vec{r}_1, \vec{r}_2) &= \langle \tilde{U}(\vec{r}_1; z) \tilde{U}^*(\vec{r}_2; z) \rangle \\ &= U(\vec{r}_1, 0) U^*(\vec{r}_2, 0) \times \langle \exp[i\varphi_{d1}(\vec{r}_1)] \exp[i\varphi_{d2}(\vec{r}_2)] \rangle \\ &= U(\vec{r}_1, 0) U^*(\vec{r}_2, 0) \exp\left[-(\vec{r}_1 - \vec{r}_2)^2 / 2l_c^2\right], \dots \end{aligned} \quad (9.122)$$

where l_c is the partial coherence length, which describes the partial coherence characteristics of the transmitter. The cross-spectral density function $W(\vec{\rho}_1, \vec{\rho}_2; z)$ on the receiving plane is expressed as follows:

$$\begin{aligned} W(\vec{\rho}_1, \vec{\rho}_2; z) &= \frac{1}{(\lambda z)^2} \iiint \int d^2\vec{r}_1 d^2\vec{r}_2 W_0(\vec{r}_1, \vec{r}_2) \\ &\langle \exp[\psi(\vec{r}_1, \vec{\rho}_1) + \psi^*(\vec{r}_2, \vec{\rho}_2)] \rangle \times \\ &\exp\left\{\frac{ik}{2z}[(\vec{\rho}_1 - \vec{r}_1)^2 - (\vec{\rho}_2 - \vec{r}_2)^2]\right\}. \end{aligned} \quad (9.123)$$

The cross-correlation function of the complex phase of a spherical wave is expressed as

$$\langle \exp[\psi(\vec{r}_1, \vec{\rho}_1) + \psi^*(\vec{r}_2, \vec{\rho}_2)] \rangle = \exp[-D_\psi/2], \quad (9.124)$$

where D_ψ is the phase structure function, which represents the mean square deviation of the phase between two points [4]. The expression is

$$\begin{aligned} D_\psi(\vec{r}_d, \vec{\rho}_d) &\cong 8\pi^2 k^2 L \int_0^1 \int_0^\infty \kappa \Phi_n(\kappa, L_0, l_0) \\ &\times \left\{ [1 - J_0(|(1 - \xi)(\vec{r}_1 - \vec{r}_2) + \xi(\vec{\rho}_1 - \vec{\rho}_2)|\kappa)] d\kappa d\xi, \right. \end{aligned} \quad (9.125)$$

where $\Phi_n(\kappa, L_0, l_0)$ is the refractive index turbulence spectrum model. Under the Kolmogorov refractive index spectrum model, the structure constant is

$$\begin{aligned}
D_\psi &= 2.92k^2z \int_0^1 dt C_n^2(tz) |(\vec{r}_1 - \vec{r}_2)t + (1-t)(\vec{\rho}_1 - \vec{\rho}_2)|^{5/3} \\
&\cong 2 \left\{ [(\vec{r}_1 - \vec{r}_2)^2 + (\vec{r}_1 - \vec{r}_2)(\vec{\rho}_1 - \vec{\rho}_2) + (\vec{\rho}_1 - \vec{\rho}_2)^2] / \rho_T^2 \right\}.
\end{aligned} \tag{9.126}$$

According to Eqs. (9.124) and (9.126)

$$\begin{aligned}
&\exp[\psi(\vec{r}_1, \vec{\rho}_1) \psi^*(\vec{r}_2, \vec{\rho}_2)] \\
&\cong \exp \left\{ - [(\vec{r}_1 - \vec{r}_2)^2 + (\vec{r}_1 - \vec{r}_2)(\vec{\rho}_1 - \vec{\rho}_2) + (\vec{\rho}_1 - \vec{\rho}_2)^2] / \rho_T^2 \right\},
\end{aligned} \tag{9.127}$$

where $\rho_T^2 = \left[1.46k^2z \int_0^1 C_n^2(\xi z) (1-\xi)^{5/3} d\xi \right]^{-3/5}$ is the coherent length of spherical wave in oblique turbulent atmosphere. Equation (9.127) can be written as

$$\begin{aligned}
W(\vec{\rho}_1, \vec{\rho}_2; z) &= \frac{1}{(\lambda z)^2} \iiint \int d^2\vec{r}_1 d^2\vec{r}_2 W_0(\vec{r}_1, \vec{r}_2) \\
&\exp \left\{ \frac{ik}{2z} [(\vec{\rho}_1 - \vec{r}_1)^2 - (\vec{\rho}_2 - \vec{r}_2)^2] \right\} \\
&\times \exp \left\{ - [(\vec{r}_1 - \vec{r}_2)^2 + (\vec{r}_1 - \vec{r}_2)(\vec{\rho}_1 - \vec{\rho}_2) + (\vec{\rho}_1 - \vec{\rho}_2)^2] / \rho_T^2 \right\}.
\end{aligned} \tag{9.128}$$

Equation (9.128) is the cross-spectral density function of partially coherent beams propagating in slant path turbulent atmosphere. For different types of partially coherent beams at the receiving plane, the cross-spectral density function depends on $W_0(\vec{r}_1, \vec{r}_2)$.

For narrow-band light source, the cross-correlation function can be approximately replaced by cross-spectral density $W(\vec{\rho}_1, \vec{\rho}_2, z)$ [28]:

$$\Gamma(\vec{\rho}_1, \vec{\rho}_2; z) = \langle U(\vec{\rho}_1, z) U^*(\vec{\rho}_2, z) \rangle \approx W(\vec{\rho}_1, \vec{\rho}_2, z). \tag{9.129}$$

Then, the relationship between the wave structure function and the cross-correlation function is [29, 30]

$$\exp \left[-\frac{1}{2} D(\vec{\rho}_1, \vec{\rho}_2, z) \right] = \frac{|\Gamma(\vec{\rho}_1, \vec{\rho}_2, z)|}{[\Gamma(\vec{\rho}_1, \vec{\rho}_1, z) \Gamma(\vec{\rho}_2, \vec{\rho}_2, z)]^{1/2}}. \tag{9.130}$$

The cross-correlation function of partially coherent GSM beams arriving at the receiving surface in turbulent atmosphere is expressed as [2]

$$\begin{aligned}
W(\vec{\rho}_c, \vec{\rho}_d; z) &= \langle I(\vec{\rho}_c, \vec{\rho}_d; z) \rangle \\
&= \frac{w_0^2}{W_\xi^2(z)} \exp \left\{ -\rho_d^2 \left(\frac{1}{\rho_T^2} + \frac{1}{2w_0^2 \Lambda_0^2} \right) - \frac{4\vec{\rho}_c^2 - \varphi^2 \vec{\rho}_d}{2W_\xi^2(z)} - \frac{ik\vec{\rho}_c \cdot \vec{\rho}_d}{F_\xi^2(z)} \right\},
\end{aligned} \tag{9.131}$$

where $\varphi = \Theta_0/\Lambda_0 - \Lambda_0 w_0^2/\rho_T^2 \cdot W_\zeta(z)$ and $F_\zeta(z)$ are the effective beam radius and phase wavefront curvature radius of GSM beam in turbulent atmosphere, respectively.

$$W_\zeta(z) = w_0(\Theta_0^2 + \zeta \Lambda_0^2)^{1/2}, \quad \zeta = \zeta_s + \frac{2w_0^2}{\rho_T^2}, \quad (9.132)$$

$$F_\zeta(z) = \frac{z(\Theta_0^2 + \zeta \Lambda_0^2)}{\varphi \Lambda_0 - \zeta \Lambda_0^2 - \Theta_0^2}, \quad \Lambda_0 = 2z/kw_0^2, \quad (9.133)$$

where the effective parameters of Gaussian beam at Θ_0 and Λ_0 transmitters are as follows: when $\Theta_0 = 1$, the beam is collimated. When $\Theta_0 < 1$, the beam is convergent, and when $\Theta_0 > 1$, the beam is divergent [2]. ζ is the global coherence parameter, and $\zeta_s = 1 + w_0^2/l_c^2$ is the source coherence parameter of the beam at the transmitter. If $\zeta_s = 1$, the beam is completely coherent, and if $\zeta_s > 1$, the beam is partially coherent.

When propagating horizontally, the structure constant C_n^2 of atmospheric refractive index does not change with altitude and remains constant, and ρ_T degenerates to $\rho_0 = [0.545C_n^2 k^2 L]^{-3/5}$. The phase configuration function of partially coherent GSM beam in horizontal path turbulent atmosphere can be obtained by synthesizing Eqs. (9.110) and (9.111).

$$D(\vec{\rho}_1, \vec{\rho}_2, z) = 2\rho^2 \left\{ \left(\frac{1}{\rho_0^2} + \frac{1}{2w_0^2 \Lambda_0^2} \right) - \frac{\Theta_0^2}{2W_\zeta^2(z) \Lambda_0^2} + \frac{\Lambda_0 w_0^2}{W_\zeta^2(z) \rho_0^2} - \frac{\Lambda_0^2 w_0^4}{2W_\zeta^2(z) \rho_0^4} \right\} \quad (9.134)$$

The phase structure function represents the mean square deviation of the phase between two points. The phase fluctuation of partially coherent GSM beams can be measured by Eq. (9.114). The wave structure function $D = D_\chi + D_s$ can be approximately processed as $D \approx D_s$ because the phase structure function D_s plays a leading role; therefore, the influence of the logarithmic amplitude structure function D_χ can be ignored [2]. From Eq. (9.134), the variance of the fluctuation of the angle of arrival for a Gaussian and Schell beam that is horizontally propagating through atmospheric turbulence can be obtained:

$$\langle \alpha^2 \rangle = \frac{D(\vec{\rho}_1, \vec{\rho}_2, L)}{(k\rho)^2} = \frac{2}{k^2} \left\{ \left(\frac{1}{\rho_0^2} + \frac{1}{2w_0^2 \Lambda_0^2} \right) - \frac{\varphi^2}{2W_\zeta^2} \right\}. \quad (9.135)$$

9.3.5 Influence of Beam Drift and Spread on a Communication System

The equation of laser propagation in atmospheric turbulence is as follows.

$$\nabla^2 \psi(r, t) - \frac{n^2}{c^2} \cdot \frac{\partial^2 \psi(r, t)}{\partial t^2} = 0 \quad (9.136)$$

The solution is

$$\psi(r, t) = A_0(r) \exp[\chi + i\varphi(r)] \exp(-i\omega t), \quad (9.137)$$

where $A_0(r)$ is the amplitude of light wave, $\varphi(r)$ is the phase of light wave, ω is the circular frequency, and $\chi = \ln[A(r)/A_0(r)]$ is the logarithmic amplitude fluctuation caused by the atmospheric turbulence. The logarithmic amplitude fluctuation can be expressed as intensity.

$$\ln \frac{I(r, t)}{I_0} = \ln \left[\frac{A(r)}{A_0(r)} \right]^2 = 2 \ln \frac{A(r)}{A_0(r)} = 2\chi. \quad (9.138)$$

When only considering the influence of atmospheric turbulence on the bit error rate, the amplitude change can be approximately regarded as the noise caused by the atmospheric turbulence. It can be obtained from Eq. (9.138).

$$\ln \frac{I(r, t)}{I_0} = 2 \ln \frac{A_0(r) + A_i(r)}{A_0(r)} = 2 \ln(1 + \varepsilon), \quad (9.139)$$

where $A_i(r)$ is the amplitude of noise, and $\varepsilon = \frac{A_i(r)}{A_0(r)}$ is the ratio of noise to signal amplitude. When the signal strength is I_0 , and the noise intensity is $\langle I_n \rangle$, the SNR caused by the atmospheric turbulence is

$$\text{SNR} = \frac{I_0}{\langle I_n \rangle} = \frac{\langle A_0^2(r) \rangle}{\langle A_i^2(r) \rangle} = \frac{1}{\langle \varepsilon^2 \rangle} = \frac{1}{\langle \chi^2 \rangle}. \quad (9.140)$$

In the case of strong turbulence, the SNR can be approximately expressed by Taylor series:

$$\text{SNR} = \frac{1}{\langle \chi^2 + \chi^3 + \dots \rangle} \approx \frac{1}{\alpha \langle \chi^2 \rangle} = \frac{4}{\alpha \sigma_{\ln}^2}, \quad (9.141)$$

where α is the scintillation intensity factor, and $1 \leq \alpha \leq 2$ and σ_{\ln}^2 are the scintillation indices. When the beam spreading effect is considered, the effective signal-to-noise ratio SNR_{eff} is obtained as follows:

$$\text{SNR}_{\text{eff}} = \frac{\text{SNR}}{\left[1 + 1.33\sigma_{\ln I}^2 \left(\frac{2L}{kw_L^2} \right)^{5/6} \right] + F \cdot \sigma_{\ln I}^2 \cdot \text{SNR}} \quad (9.142)$$

where w_L is the beam width of the propagation plane at the receiving point, $w_L = w_0(\Theta_0^2 + \Lambda_0^2)^{1/2}$, and is the beam width at the transmission point. $\Theta_0 = 1 - L/F_0$, $\Lambda_0 = 2L/(kw_0^2)$ are transmission distance, $F = \sigma_{\ln I}^2(D)/\sigma_{\ln I}^2(D=0)$.

If the surface area of the optical receiver is sufficiently large, f is very small and negligible. Then, the effective signal-to-noise ratio SNR_{eff} can be simplified as

$$\text{SNR}_{\text{eff}} = \frac{\text{SNR}}{\left[1 + 1.33\sigma_{\ln I}^2 \left(\frac{2L}{kw_L^2} \right)^{5/6} \right]}. \quad (9.143)$$

For plane waves, the scintillation index is

$$\sigma_{\ln I}^2 = 1.23C_n^2 k^{7/6} L^{11/6}. \quad (9.144)$$

When the optical receiver receives the optical signal in a digital laser communication system, the bit error rate is

$$\text{BER} = \frac{1}{2} \left[\text{erfc} \left(\frac{\text{SNR}}{\sqrt{2}} \right) \right]. \quad (9.145)$$

The bit error rate can be expressed as

$$\text{BER} = \frac{1}{2} \left[\text{erfc} \left(\frac{\text{SNR}}{\sqrt{2}} \right) \right] = \frac{1}{2} \left[\text{erfc} \left(\frac{4}{\sqrt{2\alpha\sigma_{\ln}^2}} \right) \right]. \quad (9.146)$$

When the beam spreading effect is considered, the bit error rate can be expressed as

$$\text{BER} = \frac{1}{2} \left[\text{erfc} \left(\frac{\text{SNR}_{\text{eff}}}{\sqrt{2}} \right) \right], \quad (9.147)$$

where the effective signal-to-noise ratio is

$$\text{SNR}_{\text{eff}} = \frac{\text{SNR}}{\left[1 + 1.33\sigma_{\ln I}^2 \left(\frac{2L}{kw_L^2} \right)^{5/6} \right]} = \frac{\text{SNR}}{\left[1 + 1.33 \times 1.23C_n^2 k^{7/6} L^{11/6} \left(\frac{2L}{kw_L^2} \right)^{5/6} \right]}. \quad (9.148)$$

9.4 Summary and Prospects

Atmospheric turbulence is an unavoidable problem in atmospheric laser communication, and its suppression must be achieved. Partially coherent light has a good ability to suppress atmospheric turbulence and provides an effective way to address atmospheric turbulence. Effective generation of partially coherent beams is one of the future directions in atmospheric laser communication.

9.5 Questions

- 9.1. What is the fully coherent light? What is the partially coherent light?
- 9.2. Briefly describe the Huygens Fresnel principle.
- 9.3. Briefly describe the phenomenon of beam drift and intensity flicker.
- 9.4. Briefly describe the beam spreading phenomenon.
- 9.5. How does beam spreading affect the bit error rate of a communication system?
- 9.6. What are the impacts of angle of arrival fluctuation on a communication system?

9.6 Exercises

- 9.1. Using the generalized Huygens Fresnel principle, prove that the integral of

$$\begin{aligned} \Gamma_{pp,diff}(\mathbf{r}_1, \mathbf{r}_2, L) &= \frac{k^2}{4\pi^2} \int \int_{-\infty}^{\infty} d^2s_1 \int \int_{-\infty}^{\infty} d^2s_2 \\ &\exp\left(-\frac{s_1^2 + s_2^2}{w_0^2}\right) \exp\left(-\frac{|s_1 - s_2|^2}{l_c^2}\right) \\ &\times \exp\left(\frac{ik}{2L}|s_1 - \mathbf{r}_1|^2 - \frac{ik}{2L}|s_2 - \mathbf{r}_2|^2\right) \end{aligned}$$

is

$$\begin{aligned} \Gamma_{pp,diff}(\mathbf{r}_1, \mathbf{r}_2, L) &= \frac{W_0^2}{W_1^2(1 + 4\Lambda_1 q_c)} \exp\left[\frac{ik}{L} \left(\frac{1 - \Theta_1 + 4\Lambda_1 q_c}{1 + 4\Lambda_1 q_c}\right) \mathbf{r} \cdot \mathbf{p}\right] \\ &\times \exp\left[\frac{2r^2 + \rho^2/2}{W_1^2(1 + 4\Lambda_1 q_c)}\right] \exp\left[-\left(\frac{\Theta_1^2 + \Lambda_1^2}{1 + 4\Lambda_1 q_c}\right) \left(\frac{\rho^2}{l_c^2}\right)\right] \end{aligned}$$

where $\mathbf{p} = \mathbf{r}_1 - \mathbf{r}_2$, $\rho = |\mathbf{p}|$ and dimensionless coherent parameter $q_c = L/kl_c^2$, and $l_c = \sqrt{2}\sigma_c$ is the correlation length.

- 9.2. How to get the complex coherence coefficient from Eq. (9.13) as follow:

$$\begin{aligned}\text{DOC}_{\text{pp,diff}}(\rho, L) &= \frac{|\Gamma_{\text{pp,diff}}(\mathbf{r}_1, \mathbf{r}_2, L)|}{\sqrt{\Gamma_{\text{pp,diff}}(\mathbf{r}_1, \mathbf{r}_1, L)\Gamma_{\text{pp,diff}}(\mathbf{r}_2, \mathbf{r}_2, L)}} \\ &= \exp\left[-\left(\frac{\Theta_1^2 + \Lambda_1^2}{1 + 4\Lambda_1 q_c}\right)\left(\frac{\rho^2}{l_c^2}\right)\right]\end{aligned}$$

- 9.3. For partially coherent light, prove that the ratio of the number of speckles on the input plane and the output plane is a constant, that is

$$\frac{W_0^2}{l_c^2} = \frac{W_{pp,dif}^2}{\rho_{pp,dif}^2},$$

where $W_{pp,diff} = W_1 \sqrt{1 + 4\Lambda_1 q_c}$ is the size of the partially coherent spot and W_0 is the spot diameter.

- 9.4. Under a strong diffusion, prove that the average speckle radius can be expressed as:

$$\rho_{\text{pp,speckle}} = \lim_{l_c \rightarrow 0} \sqrt{\frac{l_c^2(1 + 4\Lambda_1 q_c)}{\Theta_1^2 + \Lambda_1^2}} = \frac{\sqrt{2}\lambda L}{\pi W_0}$$

- 9.5. Consider a Gaussian spectrum model

$$\Phi_S(\kappa) = \frac{\langle n_1^2 \rangle l_c^3}{8\pi\sqrt{\pi}} \exp\left(-\frac{l_c^2 \kappa^2}{4}\right)$$

with normalization

$$\frac{\sqrt{\pi} \langle n_1^2 \rangle \kappa^2 l_c \Delta z}{1 + 4\Lambda_1 q_c} = 1$$

Explain how to solve the following equations:

$$\begin{aligned}(1) \quad \sigma_{r,\text{diff}}^2(\mathbf{r}, L) &= 2\pi^2 k^2 \Delta z \int_0^\infty \kappa \Phi_S(\kappa) e^{-\Lambda_1 L \kappa^2 / k} [I_0(2\Lambda_1 r k) - 1] d\kappa \\ &= \exp\left[\frac{4\Lambda_1^2 r^2}{(1 + 4\Lambda_1 q_c) l_c^2}\right] - 1\end{aligned}$$

$$(2) \quad T_{\text{diff}}(L) = 4\pi^2 k^2 \Delta z \int_0^\infty \kappa \Phi_S(\kappa) \left(1 - e^{-\Lambda_1 L \kappa^2 / k}\right) d\kappa = 4\Lambda_1 q_c$$

$$\begin{aligned}(3) \quad \Delta_{\text{diff}}^2(\mathbf{r}_1, \mathbf{r}_2, L) &= 4\pi^2 k^2 \Delta z \int_0^\infty \kappa \Phi_S(\kappa) e^{-\Lambda_1 L \kappa^2 / k} \\ &\times [I_0(2\Lambda_1 r_1 k) + I_0(2\Lambda_1 r_2 k) - 2J_0(\kappa |\Theta_1 \mathbf{p} - 2i\Lambda_1 \mathbf{r}|)] d\kappa \\ &= 2\left(\frac{\Theta_1 + \Lambda_1^2}{1 + 4\Lambda_1 q_c}\right) \frac{\rho^2}{l_c^2} - \frac{4i\Theta_1 \mathbf{p} \cdot \mathbf{r}}{(1 + 4\Lambda_1 q_c) l_c^2}\end{aligned}$$

- 9.6. Consider the ratio of spot radius between image plane and pupil surface

$$\frac{\rho_{\text{ip,speckle}}}{\rho_{\text{pp,speckle}}} = \left(\frac{W_1}{W_G} \right) \left(\frac{W_{\text{ip,diff}}}{W_{\text{pp,diff}}} \right),$$

Explain that under the strong diffusion limit, the average speckle radius is reduced to

$$\frac{\rho_{\text{ip,speckle}}}{\rho_{\text{pp,speckle}}} = \frac{W_2}{W_G} \sqrt{\frac{\Theta_1^2 + \Lambda_1 \Omega_G}{\Lambda_1 (\Lambda_1 + \Omega_G)}}$$

- 9.7. Supposed that the collimated beam diameter is 50 mm, the laser wavelength is 1550 nm, the correlation length is 20 mm, the atmospheric structure constant is $C_n^2 = 2 \times 10^{-14} \text{ m}^{-2/3}$, and the beam propagation distance is 1 km. Calculate the beam radius after propagation in atmospheric turbulence.
- 9.8. Given the beam parameters are as follows:

$$\Theta_{ed} = \frac{\Theta_1}{1 + 4\Lambda_1 q_c}, \quad \Lambda_{ed} = \frac{\Lambda_1 N_S}{1 + 4\Lambda_1 q_c},$$

Prove that the above beam parameters can also be defined by the propagation parameters and the number of speckles as follows:

$$\Theta_{ed} = \frac{\Theta_0}{\Theta_0^2 + \Lambda_0^2 N_S}, \quad \Lambda_{ed} = \frac{\Lambda_0 N_S}{\Theta_0^2 + \Lambda_0^2 N_S}.$$

- 9.9. Consider a collimated beam with wavelength $\lambda = 10.6 \mu\text{m}$ and diameter 3 cm at the transmitter, calculate the scintillation index at the point on the axis in distance of 3 km away from the transmitter after beam expansion with correlation length of $l_c = 1 \text{ cm}$. Suppose $C_n^2 = 3 \times 10^{-13} \text{ m}^{-2/3}$ and ignore the influence of internal and external scales, what is the scintillation index in the case of strong diffusion ($l_c \rightarrow 0$)?
- 9.10. Assuming that the beam and atmosphere characteristics are the same as above. When the source coherence time (τ_s) and the detector (τ_D) response time meet the following conditions, calculate the longitudinal component of the scintillation index:
- (1) $\tau_s/\tau_D = 1$;
 - (2) $\tau_s/\tau_D = 10$;
 - (3) $\tau_s/\tau_D = 0.1$.
- 9.11. What is the scintillation index of off-axis position $r/W_1 = 0.8$ when considering question 9.9 and correlation length $l_c = 1 \text{ cm}$ and $l_c \rightarrow 0$ respectively?

Consider a Gaussian spectral function $\Phi_s(k) = \frac{\langle n_1^2 \rangle l_c^3}{8\pi\sqrt{\pi}} \exp\left(-\frac{l_c^2 k^2}{4}\right)$, where l_c is transverse correlation radius ($l_c = \sqrt{2}\sigma_c$), and the normalization function is $\frac{2\sqrt{\pi}\langle n_1^2 \rangle k^2 l_c \Delta z}{1+4\Lambda_1 q_c} = 1$, and $\langle n_1^2 \rangle$ is the index fluctuation of refractive index.

9.12. Prove that

$$\begin{aligned} \sigma_{\text{r,diff}}^2(0, L) &= 8\pi^2 k^2 \Delta z \int_0^\infty \kappa \Phi_S(\kappa) e^{-\Lambda_1 L \kappa^2 / k} \\ &\quad \left[1 - \cos\left(\frac{\Theta_1 L \kappa^2}{k}\right) \right] d\kappa \\ &= 1 - \frac{1 + 4\Lambda_1 q_c}{(1 + 4\Lambda_1 q_c)^2 + 16\Theta_1^2 q_c^2}. \end{aligned}$$

- 9.13. Consider a collimated beam with wavelength $\lambda = 10.6 \mu\text{m}$ and diameter 3 cm at the transmitter, which is reflected by a smooth target with size of 4 cm at 1 km from the transmitter. Calculate the free space spot radius of W_2 when the reflected wave arrives at the receiving plane. If the target is a rough surface of $l_c = 1 \text{ cm}$, what is the free space spot radius $W_{2,\text{diff}}$?
- 9.14. Consider a collimated beam with wavelength $\lambda = 10.6 \mu\text{m}$ and diameter 3 cm at the transmitter, which is reflected by a target of size 7 cm and $l_c = 1 \text{ cm}$ at 800 m from the transmitter. Calculate the spatial coherence radius of the echo beam assuming that, on the transmitting plane, $\rho_0 = 2 / (0.55 C_n^2 k^2 L)^{3/5}$ and $C_n^2 = 2.6 \times 10^{-14} \text{ m}^{-2/3}$. In addition, calculate the average number of speckles when the receiving lens size is 4 cm.

References

1. Wang L (2006) Study on transmission characteristics and spectral changes of partially coherent light. Southwest Jiaotong University, Chengdu, pp 10–11
2. Li Y (2014) Propagation characteristics of partially coherent beams in oblique turbulent atmosphere. Xi'an: Xi'an University of Electronic Science and technology, pp 38–40
3. Zhang X, Dan Y, Zhang B (2012) Beam propagation of partially coherent light propagating obliquely in turbulent atmosphere. Acta Optica Sinica 32(12):120, 1001–1–7
4. Rao R (2005) Light propagation in turbulent atmosphere. Anhui Science and Technology Press, Hefei, pp 136–139
5. Dan YQ, Zhang B (2008) Beam propagation factor of partially coherent flat-topped beams in a turbulent atmosphere. Opt Express 16(20):15563–15575
6. Andrews LC, Phillips RL (2005) Laser beam propagation through random media. SPIE optical Engineering Press, Bellingham, p 195
7. Shirai T, Dogariu A, Wolf E (2003) Directionality of Gaussian Schell-model beams propagating in atmospheric turbulence. Opt Lett 28(8):080610–080612
8. Wei L, Liu LR, Sun JF, Yang QG, Zhu YJ (2007) Change in degree of coherence of partially coherent electromagnetic beams propagating through atmosphere turbulence. Opt Commun 27(1):1–8

9. Wang H, Wang X, Zeng A, Yang K (2007) Influence of atmospheric turbulence on the temporal coherence of oblique propagation quasi monochromatic Gaussian Schell beams. *Acta Opt Sinica* 27(9)
10. Xiang J, Wu Z, Wang M (2013) Broadening and drift of partially coherent Gaussian Schell beams in atmospheric turbulence. *Infrared Laser Eng* 42(3):658–662
11. Ke X, Wang W (2015) Intensity and propagation of partially coherent light in oblique and horizontal atmospheric turbulence. *Acta Appl Sci* 33(2):142–154
12. Duan M, Li J, Wei J (2013) Propagation of partially coherent Hermite Gaussian beams in oblique atmospheric turbulence. *High Power Lasers Part Beams* 25(9):2252–2256
13. Wang W (2015) Intensity distribution and beam propagation of partially coherent light in atmospheric turbulence. Xi'an: Xi'an University of technology, pp 33–35
14. Zhang X, Dan Y, Zhang B (2012) Beam propagation of partially coherent light propagating obliquely in turbulent atmosphere. *Acta Opt Sinica* 32(12):120, 1001–1–7
15. Yang AL, Zhang ET, Ji XL et al (2008) Angular spread of partially coherent Hermite-cosh-Gaussian beams propagating through atmospheric turbulence. *Opt Express* 16(12):8366–8380
16. Ke X, Wang W (2015) Beam propagation and angular spread of partially coherent light in atmospheric turbulence. *J Infrared Laser Eng* 44(7)
17. Han M (2015) Beam drift of partially coherent light in atmospheric turbulence. Xi'an: Xi'an University of technology
18. Rao R (2005) Light propagation in turbulent atmosphere. Anhui Science and Technology Press, Hefei
19. Andrews LC and Phillips RL (2005) Laser beam propagation through random media. SPIE Press, pp 83–99
20. Tofsted H (1992) Outer-scale effects in beam-wander and angle-of-arrival variances. *Appl Opt* 31(27):5865–5870
21. Ishimaur A (1978) Wave propagation and scattering in random media. Academic Press, New York
22. Yura HT (1972) Mutual coherence function of a finite cross section optical beam propagating in a turbulent medium. *Appl Opt* 11(6):1399–1406
23. Xiao X, Voelz DG. Beam wander analysis for focused partially coherent beams propagating in turbulence[J]. *Optical Engineering*, 2012, 51(2):026001–1–026001–7.
24. Lutomirski R, Yura HT (1971) Propagation of a finite optical beam in an inhomogeneous medium[J]. *Appl Opt* 10:1652–1658
25. Mandel L, Wolf E (1995) Optical coherence and quantum optics. Cambridge University Press, Cambridge, pp 98–123
26. Ricklin JC, Davidson FM (2002) Atmospheric turbulence effects on a partially coherent Gaussian beam: implications for free-space laser communication. *J Opt Soc Am A* 19:1794–1802
27. Ke X, Zhang Y (2015) Intensity scintillation effect of partially coherent light in atmospheric turbulence. *Acta optica Sinica* 35(1):0106001–0106007
28. Zhang Y (2015) Study on scintillation effect and angle of arrival fluctuation effect of partially coherent light. Xi'an, Xi'an University of Technology
29. Wang SCH, Plonus MA (1979) Optical beam propagation for a partially coherent source in the turbulent atmosphere. *J Opt Soc Am* 69:1297–1304
30. Andrews LC, Phillips RL (1998) Laser beam propagation through random media. SPIE Press, Washington, pp 1–19

Chapter 10

Optical Communication in the Future



With the increasing scarcity of radio spectrum and the unprecedented activity of human space activities, optical communication will be the primary means of human space communication in the future. This chapter introduces X-ray space communication, terahertz wave communication, quantum communication, neutrino communication, and gravitational wave communication.

10.1 X-ray Space Optical Communication

Compared with other communication methods such as microwave and laser, X-ray space optical communication has the advantages of low power consumption, good directivity, strong security, long transmission distance, strong anti-interference ability, and wide communication frequency band.

10.1.1 Backgrounds

In the 1950s, human beings began to explore the outer space using space technology. Since the beginning of the twenty-first century, *Courage* and *Opportunity* Mars probes have successfully landed, and *Cassini* spacecraft arrived on Saturn, successfully released *Huygens* probe, and landed on Titan, the largest satellite of Saturn. Deep space exploration has increasingly become the focus of attention. At present, the electromagnetic wave band used in deep space communication is primarily focused on microwave.

The wavelength of X-rays is approximately $(20\text{--}0.06) \times 10^{-8}$ cm. This type of ray, with short wavelength and high penetrability, can make fluorescence of many invisible solid materials visible. Therefore, X-ray is widely used in medical fluoroscopy and nondestructive testing. Furthermore, X-ray can be used as an information carrier in communication.

In 2007, an astrophysicist at NASA Goddard Space Flight Center proposed the concept of using X-rays as the transmission medium to realize space communication. Additionally, Zhao Baosheng, a researcher at Xi'an Institute of Optics and Mechanics of Chinese Academy of Sciences independently proposed the framework of X-ray communication.

When the photon energy of X-ray is greater than 10 keV ($\lambda < 0.1$ nm) and the atmospheric pressure is lower than 10^{-1} Pa, the transmission rate of X-rays is almost 100%, which means that the transmission of X-rays in space environment has almost no attenuation. The refractive index of an X-ray in any medium is approximately 1, so there is almost no dispersion problem. X-ray communication has the following potential advantages. (1) X-ray photon energy is large, propagation attenuation in vacuum is very small, and dispersion is colorless. Therefore, it is expected that X-ray can realize long-distance space transmission. (2) The frequency of X-rays is very high (up to 10^{18} Hz). If X-ray modulation technology can be proposed and developed, the very large transmission bandwidth of X-ray communication can be utilized; this solves the problem of extremely low data transmission rate in deep space communication. (3) Space X-ray communication technology can be used in some special circumstances in which microwave and laser cannot penetrate.

10.1.2 X-ray Communication System

10.1.2.1 System Structure

As shown in Fig. 10.1 [1], the X-ray communication system is primarily composed of a signal source module, modulation circuit module, X-ray grating controlled modulation source, X-ray detector, demodulation circuit module, and sink module. The source output is a digital signal, and the sink is used to receive the digital signal

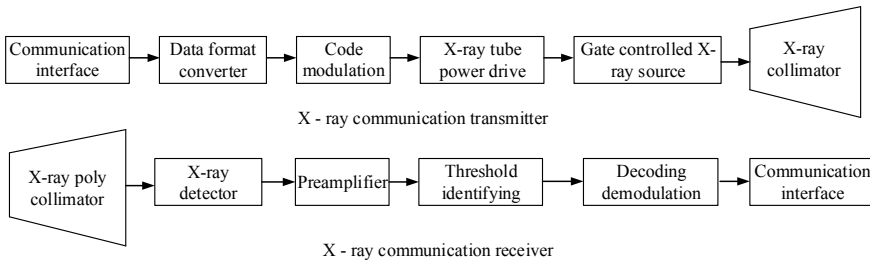


Fig. 10.1 System block diagram of X-ray space communication [1]

output by the demodulation circuit. The main function of the modulation circuit is to convert the complex metal oxide semiconductor level of the digital signal into $-10-+2$ V level and add it to the gate-controlled electrode to control the X-ray emission. The demodulation circuit restores the output signal of the detector into a digital signal. The working principle of X-ray communication system is similar to that of laser communication, but because the band of loading signal is X-ray, the specific implementation scheme and device selection are different from traditional optical communication.

10.1.2.2 X-ray Transmitter

The main function of X-ray transmitter is to transmit an X-ray signal with the communication signal. There are many ways to generate X-rays, including X-ray tube, X-ray free-electron laser, and synchrotron radiation.

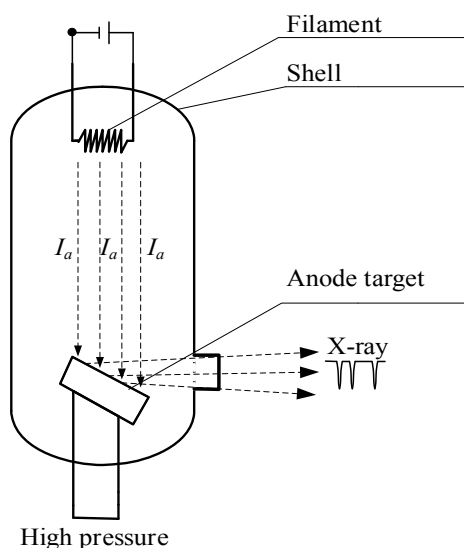
(1) X-ray tube [2]

X-ray tube is produced by the impact of high-speed electrons on a metal target. It can be divided into two types: inflatable tube and vacuum tube. In 1895, Röntgen discovered X-ray by using crux tube, which is the earliest gas filled X-ray tube. When the tube is connected with high-voltage power supply, the gas in the tube is ionized. Under the bombardment of positive ions, electrons escape from the cathode and impact on the target surface to generate X-rays. Gas filled X-ray tube is difficult to control and has a low power and short life, so it is rarely used. In 1913, W. D. Coolidge invented the vacuum X-ray tube, in which the vacuum degree was not less than 10^{-4} Pa. The cathode of the X-ray tube is a directly heated spiral tungsten wire, and the anode is a metal target embedded in the copper block end face. The working principle of the X-ray tube is to apply high voltage (V_a) on the anode to accelerate the electron beam (I_a) emitted from the cathode and to accelerate the electrons to bombard the anode target with eV_a energy to generate X-rays. In 2002, a new method of X-ray generation was developed [3, 4]. This method uses carbon nanotubes to make “field emission cathode-ray tube” to emit high-energy electrons. X-rays can be produced without using high-energy electron beams generated at high temperature. At room temperature, a thin layer of carbon nanotubes can generate high-energy electron beams, which can emit X-rays when the power is turned on. Figure 10.2 shows the schematic diagram of the X-ray tube producing X-rays.

(2) X-ray free-electron laser

A free-electron laser (FEL) works by moving a relativistic beam of high-quality electrons through a vibrator. Under the action of a magnetic field, transverse acceleration motion is produced, and the velocity component consistent with the transverse field strength component of the electromagnetic field is obtained. The transverse velocity component generated by the magnetic field is coupled with the radiation field of the electron. The energy extracted from the electron beam is finally converted into light field radiation. The FEL is designed based on this basic idea.

Fig. 10.2 Structure of an X-ray tube [2]



An X-ray FEL can produce femtosecond coherent light with adjustable wavelength and high intensity. The FEL has the advantages of wide wavelength coverage, easy adjustment of wavelength, high brightness, narrow pulse, and good coherence.

(3) Synchrotron radiation

The electromagnetic radiation emitted by relativistic charged particles moving along an arc in a magnetic field is called synchrotron radiation. High-energy physicists have always regarded that synchrotron radiation consumes the energy of the accelerator and hinders the improvement of the particle energy as a disadvantage of synchrotron. With the deepening of scientific research, it has been found that synchrotron radiation can be used in high-performance light sources ranging from infrared to hard X-ray wavelengths. Generally, X-rays produced by synchrotron radiation can only be produced by high-speed electron cyclotron. This type of equipment is too large to be used in a spacecraft, so it is not suitable for space communication applications.

10.1.2.3 Cathode of an X-ray Source

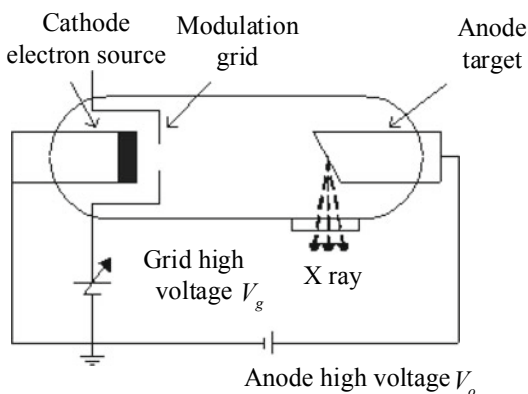
As the electron emission source of vacuum electronic devices, the material selection of a cathode, which is called the “heart” of vacuum electronic devices, plays an important role in the performance of an X-ray source. The essence of electron emission from a hot electron cathode is to heat the cathode material to high temperature. Then, the electron overcomes the surface barrier of the cathode material to activate a free electron.

10.1.2.4 X-ray Grid-Controlled Modulation Source

Figure 10.3 shows the schematic diagram of the X-ray grid-controlled modulation source. Based on the traditional X-ray tube, the modulation grid and the electron focusing electrode are added. The advantages of the X-ray source are small volume, easy modulation, and simple structure. The principle of modulation is as follows. When the digital signal input from the grid is high level 1, the electrons generated by the filament cathode will move towards the grid under the action of the electric field, bombard the anode target, and generate X-rays after passing through the grid. When the digital signal input by the grid is low-level 0, the voltage between the loaded filament and the grid will hinder the electron from moving towards the anode; thus, no X-rays can be generated. The electron focusing electrode controls the size of the electron beam spot between the grid and anode target to realize the function of electron focusing. In addition, the time dispersion of electrons is reduced, and the time resolution is improved.

Figure 10.4 shows the relationship between the grid-controlled voltage of the tube and the X-ray photon counting rate measured by the detector. It can be seen that when the grid voltage is 0 V, the energy of the X-ray is the maximum; that is, at this time, the number of X-ray pulses emitted by the bulb tube is the largest, and the corresponding input digital signal is high level 1. When the grid voltage is -7 V, the number of X-ray photons received by the detector is basically 0, which indicates that the electrons in the X-ray source are completely intercepted by the grid, and the corresponding input digital signal is low-level 0. The function of the modulation circuit is to convert the complex metal oxide semiconductor level of digital signal into $-10 \sim +2$ V level and add it to the gate-controlled electrode to control the X-ray emission.

Fig. 10.3 X-ray grid-controlled modulation source



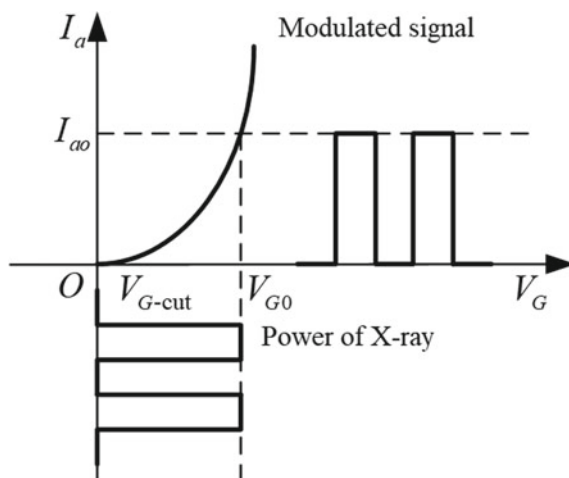


Fig. 10.4 Relationship between photon count rate measured by detector and grid voltage V_G [5, 6]

10.1.2.5 X-ray Detector

The main function of an X-ray detector is to convert the received X-ray signal into electrical signal. In an optical communication system, the X-ray detector plays the role of an antenna, which has large receiving area and high gain, conversion rate, and sensitivity. This is conducive to reduction of the bit error rate and improvement of the quality of communication. The wavelength of X-rays is primarily between 0.01 nm and 10 nm. To improve the receiving efficiency of the detector, the detector must have a higher response in this band [5, 6].

Figure 10.5 shows the structure of an X-ray photoelectric detector based on the external photoelectric effect. It is primarily composed of a high quantum efficiency

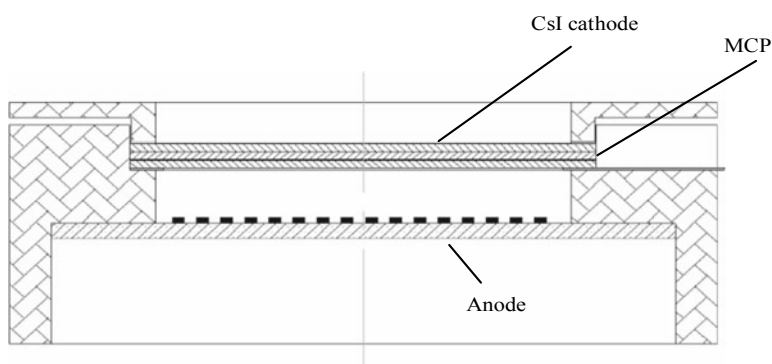


Fig. 10.5 Structure of an X-ray detector [5, 6]

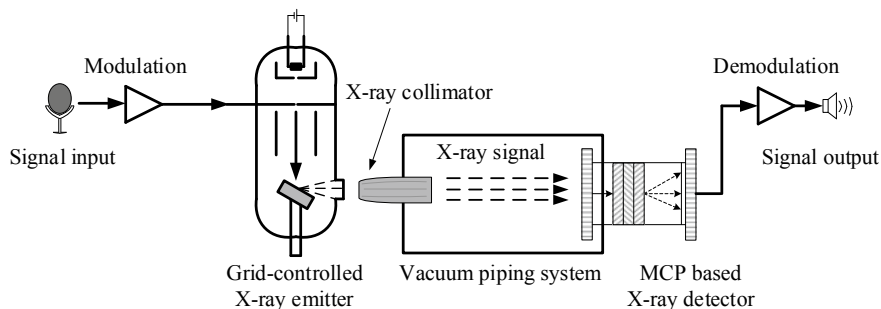


Fig. 10.6 X-ray communication ground simulator [5, 6]

photocathode, input window, high gain electron multiplier, and impedance matching charge collection anode. The gain of microchannel plates (MCP) can reach 10^3 dB, which is primarily used to realize the avalanche multiplication of electrons. Polyimide is one of the organic polymer materials with the best comprehensive properties. It can withstand temperatures of above 400°C , has no obvious melting point, and has high insulation performance. It can be used as the material of the input window. The main function of the photocathode is to convert the light signal into electrical signal. As the alkali halide compound CsI is sensitive to X-rays, it can be used as the material of the photocathode. The main function of the impedance matching charge collection anode is to collect and output electronic charge. Figure 10.6 is a schematic diagram of an X-ray communication ground simulator.

10.1.3 Development Directions and Prospects

The refractive index of an X-ray in any medium is 1, and there is no dispersion problem. Therefore, the X-ray is expected to realize long-distance space transmission with small volume, weight, and power consumption and to solve the challenge of transmission distance in deep space communication. The key points of X-ray deep space wireless communication include four aspects: (1) transmission theory of space communication (including studying the temporal characteristics of an X-ray passing through interstellar space and medium, pulse broadening, channel delay, and the influence of solar and celestial gravity on X-ray propagation); (2) high power, broadband X-ray pulse modulation emission technology, focused on solving the theoretical and technical bottlenecks of high-power X-ray emitters, the quantitative relationship between output power density and structure parameters, and the relationship between electron beam time dispersion and structure and modulation parameters; (3) research of extremely weak X-ray detection technology focused on solving the problem of high quantum efficiency photoelectric detection, which primarily involves X-ray detectors and large area array integration technology, electronic readout system with

ultrafast time response, and signal extraction and demodulation method in single photon state; and (4) deep space channel model, channel capacity, and influencing factors.

10.2 Orbital Angular Momentum Multiplexing Communication

10.2.1 *Vortex Beam*

A vortex is one of the most common phenomena in nature. It is not only found in classical macroscopic systems, such as water, clouds and cyclones, but also in quantum microscopic systems, such as superfluids, superconductors and Bose–Einstein condensation. The vortex is considered as an intrinsic morphological feature of wave [7].

When tidal motion is studied, a type of special point in the vortex of the tide is found. The tidal peak disappears when the tide is in contact with the isotropic line. From this phenomenon, it can be seen that there is a singular point in the tidal wave, namely, the optical vortex [8]. Richards and Boivin et al. [9, 10] found that a singular ring is formed at the focal plane of an achromatic lens, and experimentally found an optical vortex at the focal plane due to line rotation. This confirms the existence of optical vertices in the optical wave field. In 1973, Carter et al. [11] used a computer to simulate the characteristics of the singular ring and found that the singular ring can be created or disappeared when the beam is slightly perturbed. In 1974, Nye et al. [12] found the existence of phase singularities in seawater acoustic waves in the study of scattered fields, and extended the concept of singularities to the field of electromagnetic waves for the first time. In 1981, Baranova et al. [13, 14] discovered the existence of randomly distributed optical vortices on a laser spot, and experimentally found that the probability of generating optical vortices in a scattered optical field can be determined under certain conditions, but no optical vortex field of high order topological charge is generated. In 1992, through theoretical and experimental studies, Swartzlander et al. [15] discovered the existence of optical vortex solitons in self-focusing media, and found that the optical vortex solitons interact with the nonlinear media during transportation. This finding is a great contribution to the propagation of optical vortices. In 1998, Voitsekhovich et al. [16] investigated the properties of the phase singularity number density in detail under certain fluctuation conditions, and the results showed that the phase singularity number density is not a specific value but obeys certain statistical distribution, which is related to the probability distribution of the spatial derivatives of the amplitudes.

In the twenty-first century, as the further expansion of the research fields of optical vortex, the understanding of optical vortex has reached a new height. As a form of wave, the vortex light has not only spin angular momentum, but also orbital angular momentum (OAM) caused by the phase structure of the spiral. The beams with OAM

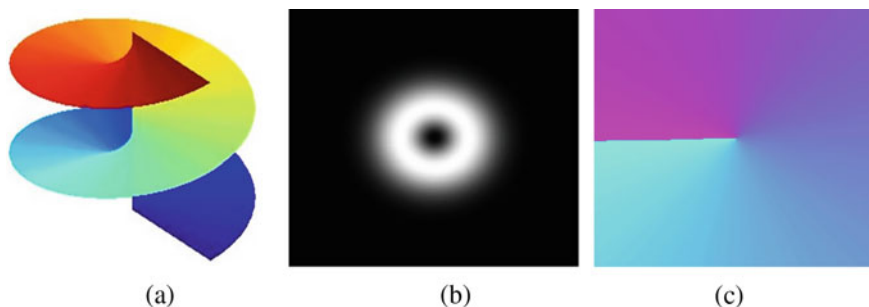


Fig. 10.7 Optical vortex field: **a** spiral wave front **b** light intensity distribution **c** phase distribution

are called “optical vortices”. Figure 10.7 shows the spiral wavefront, light intensity and phase distribution of the optical vortex field. Sub-contact and non-damaging manipulation of microscopic particles can be achieved through the transfer of photon OAM in the optical vortex field to atoms, molecules, colloidal particles, and other substances. In addition, the vortex beam also has important potential applications in the fields of radio frequency as well as the quantum secure communication due to its topological charge [17].

10.2.2 Generation of a Vortex Beam

In order to realize OAM multiplexing communication, the first problem is to generate a vortex beam with OAM. The methods for generating a vortex beam can be divided into two categories: the spatial generation method and the fiber generation method.

10.2.2.1 Spatial Generation Method

The methods for generating a vortex beam using spatial structures include: direct generation method, mode conversion method, spiral phase plate method, and computational holography method.

(1) Direct generation method

A vortex beam is generated directly by a laser resonant cavity [18]. This method has strict requirements on the axisymmetric of the resonant cavity in the experiment, and it is difficult to obtain a stable beam output.

(2) Mode conversion method

A Hermite-Gaussian (HG) beam without OAM can be converted into a Laguerre-Gaussian (LG) beam by a mode converter composed of two cylindrical lenses in a non-axisymmetric optical system, as shown in Fig. 10.8. This method was first

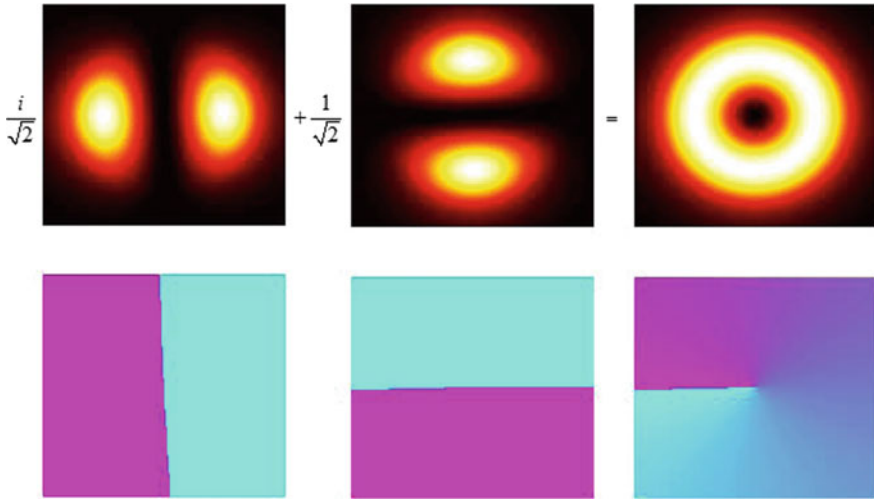


Fig. 10.8 Mode conversion of HG (HG01 and HG10) beams to LG01 beams [20]

proposed by Allen et al. in 1993, and it is also applicable for converting a LG beam to a HG beam [19]. A HG beam can be transformed into a vortex beam with OAM by simply adding a phase factor $\exp(i/l\theta)$ that varies with azimuthal angle to the HG beam [20].

The conversion efficiency of the mode conversion method is high, but the structure of the optical system in the conversion process is complex. In addition, the key optical devices used in the system are difficult to process and prepare, and it is not easy to control the type and parameters of the resulting vortex beam. These drawbacks remarkably limit the application of this method.

(3) Spiral phase plate method

A spiral phase plate [21] is a transparent plate whose thickness is proportional to the azimuth of rotation with respect to the center of the plate, and the surface structure is similar to a rotating table. When a light beam passes through the spiral phase plate, the spiral surface of the phase plate causes a different change in the light's optical path, which results in a different change in the light's phase. Subsequently, a phase factor with spiral characteristics is produced, as shown in Fig. 10.9.

The spiral phase plate method generates a vortex beam with high conversion efficiency, but the topological charge of the generated optical vortex is not unique. For a phase plate, using the laser in a specific mode only generates a specific output beam. Flexible control of the type and parameters of the vortex beam is not supported, and the preparation of high-quality phase plates is difficult.

(4) Computer-generated hologram method

In 1991, Hackenberg proposed the computer-generated holograms (CGH) method to generate the required diffraction grating pattern for vortex beam generation [22].

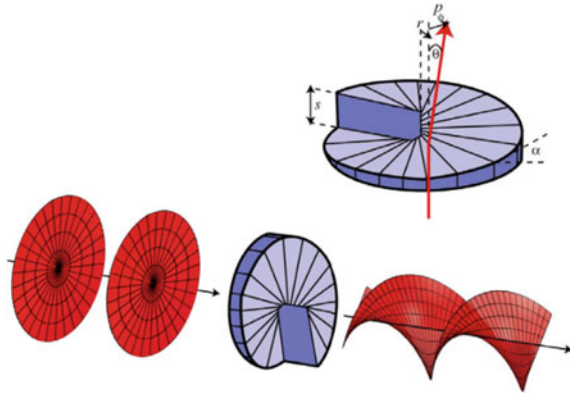


Fig. 10.9 Spiral phase plate method to generate a beam with spiral phase [21]

This method is based on the principle of interference and diffraction of light, and uses a computer program to generate the interference pattern of the target and reference light to generate a vortex beam. The computer-generated hologram is a fast and flexible method with a wide range of applications, and can either be implemented by a computer-generated hologram or spatial light modulator. To generate a vortex beam using a computer-generated hologram, the negative of a forked grating is calculated, then a Gaussian plane wave is guided to directly pass through the forked grating, as shown in Fig. 10.10a.

The spatial light modulator (SLM) method is to load the forked grating onto the SLM and let the Gaussian plane wave be directly incident on the SLM, as shown in Fig. 10.10b. Bo Bin et al. [23] used a reflective SLM to generate a beam, which was experimentally investigated by interfering with a plane wave light. The results verified that a vortex beam with different topological charges was generated, and the energy conversion efficiency generated by the vortex beam was high.

Using the SLM method, the position, size, and topological charge of the generated optical vortex can be controlled by controlling the hologram displayed on the SLM through computer. In addition, the position of the optical vortex can be dynamically adjusted in real time.

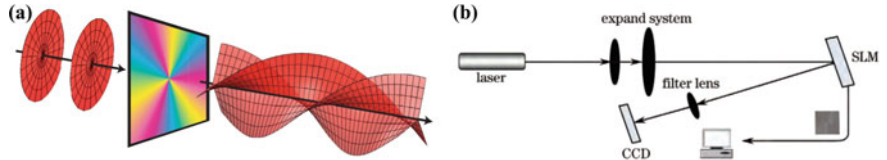


Fig. 10.10 Vortex beam generation using the computer-generated hologram method: **a** Computer-generated hologram **b** spatial light modulator

10.2.2.2 Fiber Optics Method

To meet the development and application requirements of OAM optical communication systems, several methods have been proposed to generate vortex beams using optical fibers, which primarily include: (a) fiber coupler conversion method [24], (b) photonic crystal fiber conversion method [26], and (c) optical waveguide device conversion method [27].

In 2011, Yan et al. [25] generated an OAM beam by mode superposition through an Hermite-Gaussian beam with four microfiber inputs. Later, this research group improved the microfiber by replacing it with a core of square fibers placed inside the ring fiber. This improved OAM generation coupler reduces the processing complexity since it requires only one input fiber in structure. With a simpler structure compared with the traditional OAM generation method, this method is of great significance for the promotion of the transmission technology of OAM information in optical fiber in future. However, its shortcomings are the large waveguide dispersion, low purity of the generated OAM modes, and instability and high sensitive to wavelength changes for higher-order OAM modes.

In 2012, Willner designed a new set of converters for OAM generation using a photonic crystal fiber (PCF) [26]. The basic principle is that an input Hermite-Gaussian beam is mode converted to generate a series of vortex eigenmodes, and the desired OAM modes can be generated by simply selecting an appropriate combination of the vortex eigenmodes. In the same year, Wong et al. [31] reported in *SCIENCE* a spiral PCF mode converter, which generates more OAM modes. When a laser feeds a linearly polarized super continuous light into the PCF, this converter azimuthally modulates the input light and changes the phase of the input light. Then, a vortex light with OAM is generated. In addition, the spiral PCF has the advantage that the generated OAM topological charge varies regularly with the structural parameters of the fiber, which is of a great significance for generating more OAM modes.

In 2012, Cai X L et al. reported in *SCIENCE* that a silicon-integrated OAM vortex beam transmitter has been implemented, where a Hermite-Gaussian beam was transmitted in a silicon waveguide, and then coupled into a ring waveguide. An echo wall mode was thus generated [27]. Due to the periodic sawtooth protrusions on the inner wall of the waveguide, the Hermite-Gaussian beam produced a phase difference when transmitted in the ring waveguide, which caused the optical wave vector to change. Finally, the vortex light with different OAM modes was generated above the ring waveguide. This converter is not only small in size, but also generate a vortex beam with low phase sensitivity and stable OAM patterns. In addition, it can be integrated on a large scale to generate multiple topological charge controllable OAM beams simultaneously.

10.2.2.3 Comparison of Vortex Beam Generation Methods

The unique phase structure and the unique OAM characteristics of the vortex beam make it valuable for applications in quantum information transmission, particle manipulation, and molecular optics. However, these applications rely on the generation of a high quality vortex beam. The generation methods mentioned above have their own advantages and disadvantages. Table 10.1 shows a comparative analysis of these methods for generating a vortex beam. Therefore, based on the existing conditions and technologies, searching for the effective methods to generate high quality vortex beams is an urgent problem in this field.

10.2.3 OAM Multiplexing Communication System

Compared with the traditional optical communication, the beam with OAM has new degrees of freedom, which makes the OAM multiplexing technology has unique advantages in improving the channel capacity and spectrum utilization of the system.

Table 10.1 Comparison of different vortex beam generation methods

Methods	Advantages	Disadvantages
Direct generation	Generated directly in the laser resonant cavity	Difficulty in obtaining a stable vortex beam Difficult to achieve high-order vortex beam generation
Mode conversion	High conversion efficiency	Relatively complex optical structure and difficult device preparation Not easy to control the type and parameters of the vortex beam
Spiral phase plate	High conversion efficiency	Not easy to control the type and parameters of the vortex beam Difficulty in preparation of high quality phase plates
Computer-generated hologram	Control over the position, size and parameters of the vortex beam	The optical path is strictly required because it needs to be incident to the center of the hologram
Fiber optics	Facilitating the diffusion in optical communication systems The generated vortex beam is more stable	Experimentally, only low-order vortex beams are currently possible, and high-order is currently difficult to achieve

10.2.3.1 Backgrounds

Optical wireless communication, or free-space optical (FSO) communication, is a technology that can transmit information such as data, voice and images using laser as a carrier. Due to the absorption and scattering of the light signal by the atmosphere, the beam transmitted in space is attenuated. The atmospheric turbulence effect causes drift, flicker and expansion of the laser spot, resulting in large bit error rate and even communication interruption [28]. Although the conventional channel coding method can suppress turbulence, the traditional communication method does not meet the demand of multiplexed communication in the case of strong turbulence, dense fog, etc. A new technique is needed to improve the channel capacity and spectrum utilization. In the existing multiplexing technology, the utilization of frequency, time, code, space and other dimensions have been maximized. Restricted by the information modulation format of wave in free space and optical fiber, the information is not interoperable between free space and multimode optical fiber networks. Therefore, the network capacity and communication security are difficult to be fully satisfied. To increase the communication capacity, improve the spectrum efficiency, and establish a highly reliable and secure communication network, OAM multiplexing technology is widely concerned.

OAM multiplexing communication has the following advantages [29].

- (1) **Security.** Due to the uncertainty relationship between the topological charge l and azimuth angle θ of the OAM, only when the OAM beam is received orthogonally and completely, the OAM state can be detected accurately. The angular tilt and partial reception cause the power of the transmitting mode to spread to other modes, and the probability of correct detection of the transmitting OAM state is reduced. Therefore, OAM optical communication can be effectively anti-eavesdropping.
- (2) **Orthogonality.** The vortex beams of different OAM modes have inherent orthogonality, which provides the possibility of modulating information on different vortex beams, and the information transmitted on different OAM channels does not interfere with each other, which improves the reliability of information transmission.
- (3) **Multi-dimensionality.** With the infinite number of OAM vortex beam eigenstates, it is possible to realize multiple information transmission on the same spatial path, thus improving the dimensionality of multiplexed communication.
- (4) **High spectrum utilization.** The spectrum utilization of vortex beam multiplexing communication is much higher than that of LTE (Long Term Evolution), 802.11n and DVB-T (Digital Video Broadcasting—Terrestrial) because OAM is used for multiplexing information transmission. Terrestrial).
- (5) **High transmission rate.** The transmission rate of OAM multiplexed communication is higher than LTE, 802.11n and DVB-T, and experimental studies show that it can reach Tbit order of magnitude.

With the continuous research on OAM, the vortex beam multiplexing technology with OAM, as a new multiplexing dimension, has attracted wide attentions in the field of information transmission. In order to improve the information transmission rate and guarantee the security, the multiplexing technology with OAM vortex beam is a viable option. This multiplexing technology adopts the infinity of OAM quantum number (or mode number) to transmit information in multiple channels, and adopts the orthogonality between different OAM modes to realize the modulation of information, and finally loads the information into two or more vortex beams with orbital momentum to realize a multiplexing transmission of information.

The development of OAM optical wireless communication is rapid in recent years. In 2012, Wang Jian et al. [30] proposed and demonstrated a new high-speed communication model using spatial optical modulator to realize OAM multiplexing, which achieved a free-space optical communication system capacity of 1369.6 Gbit/s. The breakthrough of Wang Jian et al. lies in using the spatial state of photons to increase the spectral efficiency of transmission, which greatly improves the transmission rate of the system. In the same year, Tamburini et al. [31] conducted a 442 m transmission experiment using wireless optical link OAM mode multiplexing in the city of Venice. In 2014, Krenn [32] achieved a 3 km optical wireless communication link using OAM beam in the city center of Vienna with a strong air interference environment. In the same year, Xu et al. [33] adopted a MIMO adaptive equalization to reduce the crosstalk between signals of OAM multiplexing system due to atmospheric turbulence. Huang et al. [34] implemented a free-space 4-way OAM mode multiplexing technique using 4×4 MIMO technique and outlier detection, where each OAM beam carries information at a rate of 20 Gbit/s, which effectively reduced the BER of the system. In 2016, Ren et al. [35] experimentally investigated the application of MIMO techniques in OAM multiplexing systems and found that the effect of atmospheric turbulence on OAM optical communication can be effectively mitigated using spatial diversity and MIMO equalization. In 2017, Xiang Zhang [36] proposed an acoustic OAM communication technique based on an active transducer array. The principle is to generate an acoustic vortex field containing 8 topological charges by a phased array consisting of 64 acoustic sources radiating signals encoded with composite vortex states, and to receive and demodulate them at the receiving end with another acoustic phased array.

10.2.3.2 Principle of OAM Multiplexing

The existing multiplexing techniques include: frequency division multiplexing, time division multiplexing, code division multiplexing and space division multiplexing. These multiplexing technologies have made breakthroughs in their research fields: The first generation (1G) technology is inseparable from the frequency division multiplexing. In the second generation (2G) technology, the time division multiplexing and code division multiplexing technology are introduced, which opens the era of digital communication. The third generation (3G) technology applies to the space division multiplexing technology, which enables the same carrier frequency to be reused in

Table 10.2 Comparison of transmission rate and spectrum utilization

Communication system	OAM	LTE	802.11n	DVB-T
Spectrum utilization	$95.5 \text{ bits}^{-1} \text{ Hz}^{-1}$	$16.32 \text{ bits}^{-1} \text{ Hz}^{-1}$	$2.4 \text{ bits}^{-1} \text{ Hz}^{-1}$	$0.55 \text{ bits}^{-1} \text{ Hz}^{-1}$
Transmission rate	2.56 Tbits^{-1}	326.4 Mbits^{-1}	144.4 Mbits^{-1}	$31.668 \text{ Mbits}^{-1}$

different directions. The fourth generation (4G) combines the OFDM and MIMO technologies, which have greatly improved the capacity and spectrum utilization of communication systems.

The OAM multiplexing technology uses the orthogonal characteristics of OAM beams to load multiple signals to be transmitted on OAM beams with different number of topological charge for data transmission. This multiplexing method achieves multiple independent OAM beam channels on the same carrier frequency at the same time by using the different topological charge numbers to distinguish different transmission channels at the receiver side. It has been found that, since the vortex beam with OAM can be tensed into an infinite-dimensional Hilbert space, the OAM multiplexing at the same carrier frequency results in a better transmission performance [37]. This feature provides a new degree of freedom for the efficient use of the spectrum.

Table 10.2 represents the transmission rates and spectrum utilization for communication system of LTE, 802.11, DVB-T and OAM multiplexing. From Table 10.2, the spectrum utilization and system transmission rate of OAM multiplexing is significantly better than the other three communication systems. This is because, compared to the conventional multiplexing techniques, OAM multiplexing is performed by using the OAM mode carried by the carrier as a modulation parameter in the multiplexing process.

10.2.3.3 System Model

The model of OAM multiplexing communication system in atmospheric turbulence is shown in Fig. 10.11. Taking four-channel multiplexing for example, the QPSK modulation is applied to the input bit stream, and the modulated signal is loaded on a Gaussian beam generated by the solid-state laser by optical modulation. Then, the electrical signal is converted into an optical signal. The Gaussian beam with the modulated information is then converted into OAM beams with different topological charge corresponding to the spatial phase masks. The generated OAM multiplexed state is transmitted through atmospheric turbulence after multiplexing the four vortex beams with different topological charge. At the receiver, the OAM beam in multiplexed state is de-multiplexed to obtain four OAM beams. Each beam is then de-multiplexed and converted into a Gaussian beam. Finally, the QPSK signal loaded on

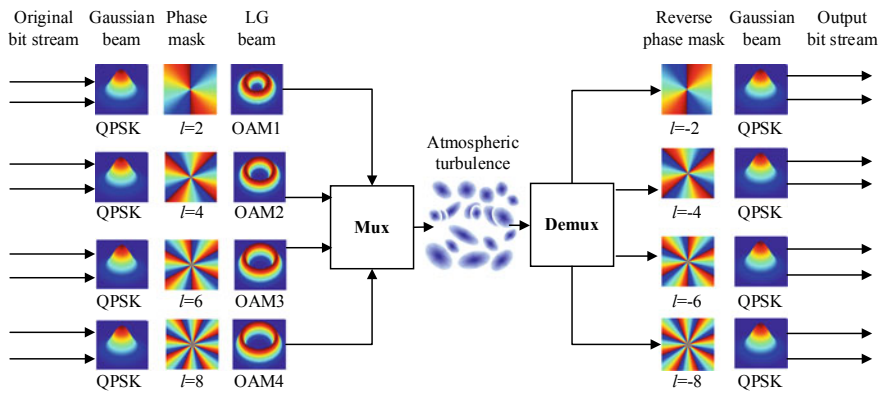


Fig. 10.11 OAM multiplexing communication system model

the Gaussian beam is extracted for demodulation to recover the original bit stream, that is, the optical signal is converted to the original electrical signal.

10.3 Neutrino Communication

Neutrino is a type of neutral elementary particle with minimal mass and no charge. It can travel near the speed of light and penetrate any substance with little energy loss [38, 39]. Neutrino communication is a communication method that uses neutrinos to deliver information.

10.3.1 Neutrino

Neutrino is one of the basic particles that make up an atom. It is very light (even less than one thousandth of an electron) and neutral. The interaction between a neutrino and other basic particles of matter is very weak, so the energy loss of the neutrino is very small. It is estimated that the energy loss is only one over 10 billion even when crossing the earth along its diameter.

There are three types of neutrinos: electron neutrino, μ -neutrino and τ -neutrino. The sun produces only the electron neutrino, and only this type of neutrino can be detected. Electron neutrinos release electrons, μ -neutrinos release μ , and τ -neutrinos release τ .

Electron neutrinos interact with atoms to release energy and illuminate a nearly spherical region. Unlike electrons, μ neutrinos are not good at interacting with each other. They must travel through at least 1 km of ice to produce a light cone. τ -neutrinos

decay rapidly, and their appearance and disappearance produces two photospheres, which are called “double explosions”.

10.3.2 Neutrino Communication

Neutrino communication is a communication method that uses neutrino beam to carry information. All wireless communication (including electronic and optical communication) must extend the transmission distance through various relay equipment (e.g., communication satellite and ground station). Because neutrinos hardly interact with any matter, the distance of neutrino communication is further than people can imagine.

Neutrinos travel in a straight line near the speed of light and can penetrate steel, sea water, and even the entire earth. They are attractive and ideal information carriers. Since neutrinos hardly interact with any substance, neutrino communication does not pose any harm to human beings.

10.3.3 Neutrino Communication System

Neutrino communication includes a transmitter and receiver. The function of the transmitter is divided into three parts:

- (1) Generator of neutrino streams as information carriers;
- (2) Modulator that loads information onto the neutrino stream for transmission;
- (3) Transmitter of the neutrino stream for transmitting the modulated neutrino stream to the transmission channel.

The functions of the receiver is also divided into three parts:

- (1) The front-end receiving device that receives the modulated neutrino stream from the neutrino communication channel;
- (2) Demodulator that demodulates the signal loaded on the neutrino stream;
- (3) Signal recovery device that reshapes and amplifies the demodulated signal to recovers the transmitted original signal.

The concept of neutrino communication is shown in Fig. 10.12.

10.3.4 Key Technologies of Neutrino Communication

A neutrino communication system includes the generation method and device of the neutrino beam, modulation and demodulation of the neutrino beam, and the receiving

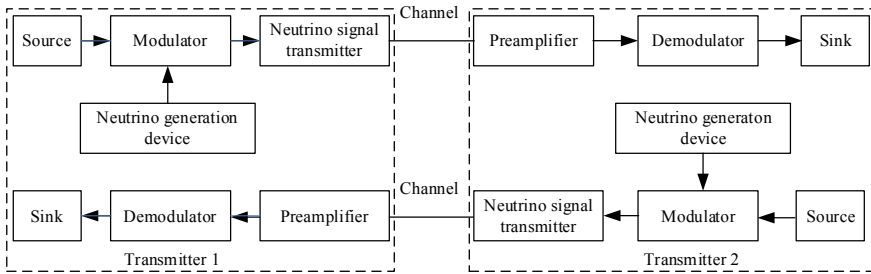


Fig. 10.12 Concept of neutrino communication

method and device. The key technologies involved in neutrino communication are as follows [40–43]:

(1) Generation of a neutrino source

Nuclear fusion reaction can produce neutrino beams, but this type of equipment is bulky and not easy to move. Moreover, it is necessary to prevent nuclear leakage due to damage, and the equipment is expensive. When the energy reaches 5×10^{11} eV, the speed of neutrino beam reaches the speed of light. By controlling the energy flow density of the neutrino beam to load up the information (e.g., video, audio, and data signals), the communication between locations at any distance can be realized.

(2) Modulation in neutrino communication

Similar to optical communication, in the transmitter, the modulator modulates the neutrino beam with the data information and then, through a magnetic field, controls the information bearing neutrino beam to transmit towards the desired target. In the receiver, the signal carried by the neutrino beam is demodulated to recover the original signal.

(3) Reception of a neutrino beam

The demodulation of neutrino communication is based on the Cherenkov effect. No matter how far away the neutrino beam propagates, as long as it passes through water with depth of 400 m above the receiving end, it will react with the neutrons of the water atoms to produce high-energy negative μ . In water, the negative μ can propagates at a speed close to the speed of light. When it travels a distance of 60–70 m, the “Cherenkov effect” (i.e., 0.4–0.7 μm Cherenkov light) is produced with an angle of 41° between the direction of the light and negative μ . By direct detection of visible light using a photomultiplier tube in water, the transmitted signals can be demodulated.

The functions of the neutrino communication receiver can be divided into three parts [40–43]: (1) Preamplifier that removes the interference and fading in the transmission and recovers the modulated neutrino stream sent by the transmitter to the channel. (2) Demodulator that demodulates useful information from the received neutrino signal. (3) Baseband signal recovery device that further reshapes and amplifies the demodulated signal to recover the baseband signal.

10.3.5 Characteristics of Neutrino Communication

Neutrino communication is generally considered to be the most suitable means for deep space communication. The main advantages are as follows.

- (1) Neutrino communication has a wide frequency band and large channel capacity. It can propagate in a straight line at nearly the speed of light. It can easily penetrate steel, sea water, and even the whole earth with little energy loss.
- (2) It has strong anti-interference ability and is not impeded by other communication modes. Because the neutrino beam propagates in a straight line using a different mechanism, it will not interfere with other communication modes.
- (3) It is safe, reliable, and secure.
- (4) It is suitable for deep space communication. Because neutrinos are electrically neutral, the neutrino beam is not affected by the ionosphere, sunspots, and other external factors. Thus, when the neutrino beam is used for communication in outer space, it can reach the information destination at the speed of light. Therefore, neutrino communication will be a powerful way of communication in universe.

10.4 Gravitational Wave Communication

The movement of electric charge produces an electromagnetic wave, and the movement of matter produces a gravitational wave. Gravitational wave can be compared with electromagnetic wave in some respects. Gravitation, like electromagnetic force, is a force that can work long distance. According to the modern theory of gravity, the universal gravitation between any two objects is transmitted by the gravitational field. When the object is accelerated, it will cause a change of the gravitational field. This change has the nature of a wave, which propagates at a limited speed. This wave is called the gravitational wave. Since light waves are transmitted by photons, Einstein assumed that gravitational waves are transmitted by gravitons. Gravitational wave communication refers to the use of gravitational waves to transmit signals. Gravitational wave is produced by the vibration of matter and is a shear wave propagating at the speed of light. It has a strong penetrating force, and no matter can stop the propagation of the gravitational wave [44–47].

10.4.1 Detection of Gravitational Waves

Gravity described by Einstein's general relativity theory is a phenomenon produced by the curvature of space–time, and mass can cause this curvature. As matter moves through space and time, the curvature of the neighborhood changes. The curvature

change of a mass object will propagate outward at the speed of light like a wave. This propagation phenomenon is the gravitational wave. Any object with mass or violent acceleration will produce gravitational waves.

A gravitational wave is a shear wave that travels at the speed of light. Because of its high nonlinearity, it does not have the reflection, interference, and diffraction properties of electromagnetic and mechanical waves and does not meet the superposition principle. Its action cross-section is very small, and its penetrating ability is very strong. It can travel through time and space and penetrate the earth.

10.4.1.1 Resonant Rod Gravitational Wave Detector

The pioneering work of gravitational wave detection is attributed to an American physicist Joseph Weber, whose original gravitational wave detector was an aluminum alloy cylinder with a high Q value. When the gravitational wave comes, the space–time changes, and the distance between the two ends of the cylinder rod in the space–time also changes slightly. Generally, the strength of the gravitational wave is expressed by a dimensionless amplitude $h = \Delta l / l$. If the frequency of the gravitational wave is consistent with the intrinsic frequency of the aluminum rod, the aluminum rod will oscillate under the excitation of the gravitational wave, and the signal of the gravitational wave will be amplified Q times. This type of metal cylinder gravitational wave detector is often called a resonant rod detector or resonant mass detector and is shown in Fig. 10.13.

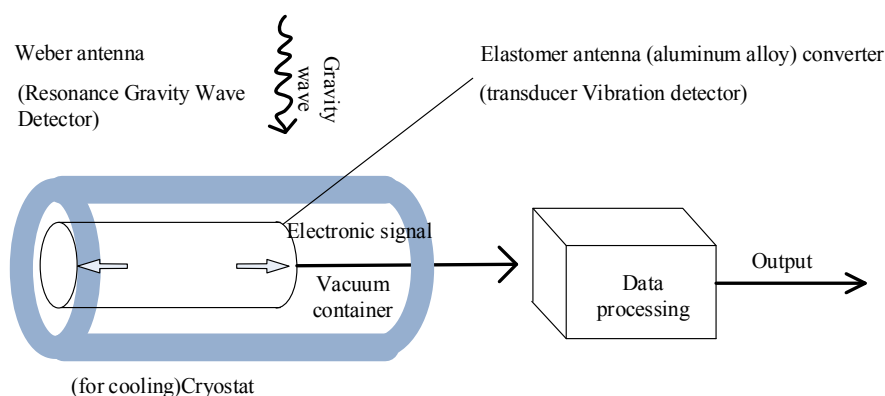


Fig. 10.13 Weber resonant rod detector [48, 49]

10.4.1.2 Laser Interference Gravitational Wave Detector

The influence of a gravitational wave on the relative position change of an object is very weak, and the length change caused by a gravitational wave of typical intensity is only one billionth.

As shown in Fig. 10.14, the interferometer consists of a light source (far left side of the above figure), a beam splitter (center), two mirrors (upper and rightmost), and a photo detector (lower side). The light from the light source is divided into two beams after passing through the beam splitter. One beam of the light penetrates directly and still propagates in the original direction, whereas the other beam is reflected in a direction perpendicular to the original propagation. It is called the two arms of the interferometer. When the two beams meet the total mirror in their respective light paths, they return to the spectroscope in the original light path.

When the two beams return to the beam splitter, the beam directly passing through the returning beam and the beam returning from the vertical reflection are again split into two by the splitter, and one beam is generated on the left and the other at the bottom. At this time, a part of the light is directly passing through and part of the light is reflected by the beam splitter in the two light paths on the left and below. These two parts of the light are superimposed together, and the two light paths interfere with each other due to the fluctuation of the light. If the wave crest of one wave coincides with that of the other, the interference will be long, and the peak height will be higher after superposition. In contrast, if the crest of one wave overlaps with the trough of the other, interference cancellation occurs. In general, the shape of the superposition wave formed by the two waves depends on the relative position of the wave crests.

This interference occurs when the light reflected from the two arms meets in the spectroscope. The position between the two reflected light peaks depends on the

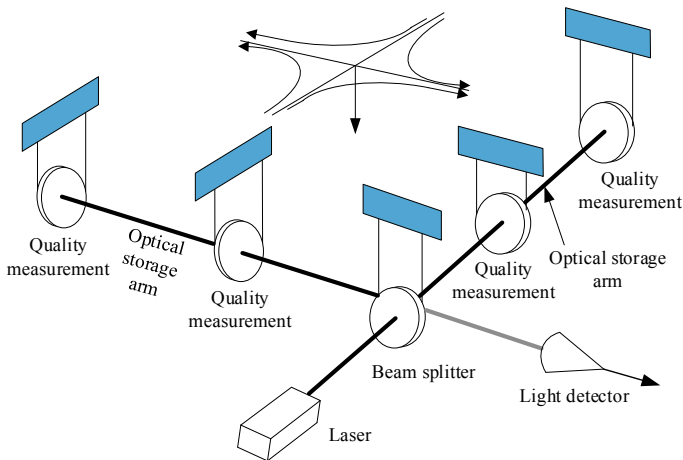


Fig. 10.14 Concept map of laser interference gravitational wave detection [49, 50]

length of the two arms. When the arm lengths are different, the position between the wave peaks is different, and the intensity of the superimposed light changes. It is the intensity change of the superimposed light that is used to detect the change of arm length caused by gravitational waves. Since the wavelength of the light used in interferometers is very short (approximately several thousand nanometers), this interference method can be used to measure very small arm length changes.

10.4.1.3 Electromagnetic Coupling Detection

After the 1970s, the electromagnetic coupling detection scheme was proposed [51, 52]. Through the interaction between an electromagnetic field and a gravitational wave, the change of electromagnetic field can be observed, and the existence of the gravitational wave can be found. Typical experimental facilities are the spherical resonator scheme in Italy and the circular waveguide scheme in Britain. In Italy, two spherical cavities are connected by a disc. The natural frequency of the cavity can be adjusted. The middle disc-shaped unit implements the function of frequency modulation. Electromagnetic fields of two resonant modes are stored in the cavity. The interaction of a gravitational wave and the superconducting cavity wall will produce a movement, and the electromagnetic field in the cavity will register this movement and generate energy conversion. When the frequency of the gravitational wave is equal to the difference frequency of the two modes of electromagnetic field, the energy conversion rate is the highest. The probability of detecting high-frequency gravitational waves can be increased by forming an array of such differential frequency double spherical resonators. The British circular waveguide scheme utilizes the interaction between a high-frequency gravitational wave and electromagnetic wave polarization vector in which the polarization vector rotates around the propagation direction of the electromagnetic wave. When the resonance condition of the electromagnetic wave is established, that is, the phases of the electromagnetic wave and the high-frequency gravitational wave are always the same, the effect will accumulate and can be improved linearly by increasing the number of waveguides [51, 52].

It has been claimed that gravitational waves have been detected in the laboratory, but this has not been supported. From an experimental point of view, considerable achievements have been made in the detection of gravitational waves. Researchers predict that direct detection of gravitational waves may be realized in the near future.

10.4.2 Generation of Gravitational Waves

Gravitational wave is a gravitational field which propagates in the form of a wave with a finite velocity. According to general relativity, the acceleration of the mass will produce gravitational waves. The main properties of a gravitational wave are as follows: it is a shear wave and plane wave at far source, it has two independent polarization states, it carries energy, and it propagates at the speed of light in vacuum.

Gravitational waves carry energy and should be detectable. But the intensity of the gravitational wave is very weak, and the efficiency of matter of gravitational wave absorption is very low, so it is very difficult to detect gravitational waves directly.

Theoretically, a spring oscillator can produce gravitational waves. A certain mass object is connected at both ends of a spring. If it is vibrated, gravitational waves will be generated; thus, the device is also called the “gravitational oscillator”. Moreover, a heavy rod that rotates about its central vertical axis generates gravitational waves.

If a steel bar weighing 500 t and measuring 20 m rotates at a speed of 5 rad/s (which is the maximum rotation speed within its strength limit), the gravitational wave energy generated is only 10–29 W. A 10 cm long spring with a weight of 1 kg at both ends oscillates at a frequency of 100 Hz and an amplitude of 1 cm. To turn its gravitational energy into electric energy to light a 50 W bulb, it needs more oscillators than all the basic particles that make up the earth.

10.4.3 Difficulties in Gravitational Wave Detection

- (1) Since the intensity of natural gravitational wave is very small and weak, it is not sufficient to produce enough gravitational wave for an experiment.
- (2) There are limitations of natural and technical conditions: space noise, testing equipment, thermal motion noise, signal conversion loss, ground vibration, etc.
- (3) The interaction cross-section of gravitational wave with mass is very small and decays with the square of distance.

10.5 Terahertz Wave Communication

10.5.1 Terahertz Wave and Its Advantages

Terahertz wave refers to the electromagnetic wave whose wavelength is between 3 mm and 3 μm (100 GHz–10 THz). This band is located between microwave and infrared radiation, which is the edge of the electromagnetic and optical research. The early research on terahertz radiation can be traced back to the 1980s. Due to the lack of effective methods to generate and detect terahertz electromagnetic waves, scientists have limited understanding of the electromagnetic radiation properties in this band. In the past ten years, the rapid development of ultrafast optoelectronic technology has provided a stable excitation source for terahertz pulse generation. With the generation of terahertz radiation, its application has also been rapidly developed. Compared with microwave and optical communication, terahertz wave has the following characteristics.

10.5.1.1 Comparison with Microwave Communication

- (1) The transmission capacity of terahertz communication is higher than that of microwave communication by 1–4 orders of magnitude. The wireless transmission rate can be as high as 10 Gbps, hundreds or even thousands of times faster than the current ultra-wideband technology;
- (2) Terahertz wave beam is narrower, has better directionality, can detect smaller targets, and locate more accurately;
- (3) Terahertz wave has higher security and stronger anti-interference ability;
- (4) Terahertz wavelength is shorter than microwave wavelength; thus, a smaller terahertz antenna can realize the same function.

10.5.1.2 Comparison with Optical Communication

- (1) The energy of terahertz wave is low in the order of meV, which is only 1/40 of photon energy;
- (2) Terahertz wave communication is working in harsh environment of plasma, dust, and smoke. There is a strong demand for terahertz wave communication in plasma in aerospace and military fields.

10.5.1.3 Characteristics of a Terahertz Wave

- (1) High transmittance: Terahertz wave has good penetrability in many dielectric materials and nonpolar materials and can be used for perspective imaging of opaque objects. It is an effective complementary of X-ray and ultrasonic imaging technologies and can be used for nondestructive testing in the process of security or quality inspection.
- (2) Low energy: Terahertz photon energy is 4.1 meV. Terahertz radiation does not cause photoionization and does not destroy the tested material, so it is very suitable for examination of human body or other biological samples to extract their refractive index and absorption coefficient conveniently.
- (3) Water absorption: Water is strongly absorbent of terahertz radiation; because the water content in tumor tissue is significantly different from that in normal tissue, the location of tumor can be determined by analyzing the water content in the tissue.
- (4) Transient: The typical pulse width of a terahertz pulse is in the order of picoseconds. It is convenient to study the time-resolved spectroscopy of various materials including liquid, gas, semiconductor, high-temperature superconductor, ferromagnetic, etc., and the interference of background radiation noise can be effectively suppressed by sampling measurement technology.
- (5) Coherence. The coherence of terahertz comes from its coherence generation mechanism. Terahertz coherent measurement technology can directly measure the amplitude and phase of the electric field to conveniently extract

optical parameters, such as refractive index, absorption coefficient, extinction coefficient, and dielectric constant.

- (6) Fingerprint spectrum: Terahertz band contains abundant physical and chemical information. Most of the polar molecules and biological macromolecules are in the terahertz band. According to these fingerprint spectra, terahertz spectral imaging technology can distinguish the shape of an object, analyze its physical and chemical properties, and provide the theoretical basis and detection technology for antidrug and anti-terrorism systems, disposal of explosives, and so on.

Terahertz wave communication can achieve a transmission rate of 10 Gbps. If the influence of water vapor is not considered for satellite communication, terahertz communication can perform high security satellite communication with extremely high bandwidth.

10.5.2 Terahertz Wave Transmitting Antenna

10.5.2.1 Generation of a Terahertz Wave

- (1) The method of generating broadband pulse radiation in terahertz by photoconductivity: the metal electrode with bias voltage and photoelectric semiconductor material are formed into antenna, and the electron–hole pair is generated in the photoconductive material by an ultrafast laser beam. The free carrier is accelerated in the bias electric field, and the photocurrent will radiate THz wave.
- (2) The method of generating broadband pulse radiation in terahertz by optical rectification: the low-frequency electrode field is generated by the interaction of a laser pulse (pulse width in sub-picosecond order) and nonlinear medium (e.g., LiNbO_3 , LiTaO_3 and ZnTe), which radiates a terahertz wave.
- (3) Narrow band continuous terahertz pulse generation technology: a narrow band light source is characterized by a single bulge at the center frequency of the spectrum, and its bandwidth is very narrow. At present, research primarily focuses on two directions: one is to extend low-frequency microwave to high-frequency by means of electronics; this is characterized by high efficiency, and a high-power terahertz wave can be generated, but the frequency of the generated terahertz wave is low. The other direction is to extend optics, especially laser technology to low frequencies; this can produce terahertz wave with good directivity and coherence. The output power is small.

As a type of terahertz source, GaAs photoconductive antenna can generate broadband terahertz wave. The basic structure of GaAs photoconductive antenna is to prepare metal electrode on semiconductor surface. Its structure is shown in Fig. 10.15. Semiconductor substrate materials are usually low-temperature GaAs (LT- GaAs),

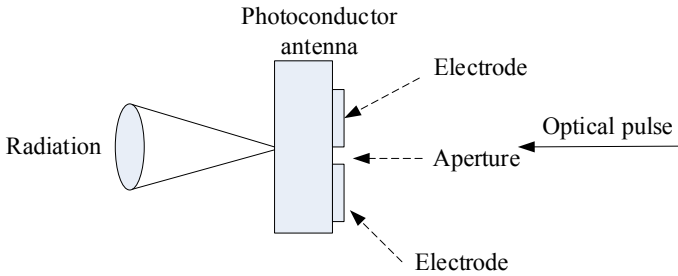


Fig. 10.15 Terahertz wave generated by photoconductive antenna [53]

or semi-insulating GaAs (Si- GaAs) and InP, which have ultrafast carrier characteristics. The models to analyze the terahertz radiation mechanism of the photoconductive antenna are the Drude–Lorentz model and large aperture photoconductive antenna model (current instantaneous model) [53].

10.5.2.2 Drude–Lorentz Model

The Drude–Lorentz model regards photo-generated carriers as electron gas and ignores Einstein’s diffusion effect in photoconductive antenna materials. When the concentration of the photo-generated carriers is approximately 10^{16} – 10^{18} cm, the photo-generated carriers can be well maintained in the thermal equilibrium state under the effect of scattering mechanism. The radiation effect of charged particles caused by the acceleration of the charged particles in the electric field can be analyzed. Starting from Maxwell equation, the electromagnetic quantities related to the radiation field can be solved by introducing vector potential A and scalar potential Φ [49].

10.5.2.3 Current Impulse Model

The current impulse model is constructed as follows. A large aperture semi-insulating GaAs photoconductive antenna with a bias electric field is used as a terahertz radiation source. The GaAs photoconductive antenna is in high resistance state when there is no light. When the antenna is irradiated by femtosecond laser, a large number of electron–hole pairs are generated in the photoconductive material between the two electrodes. When the electrodes move in two different directions, the electric field changes, and the electron–hole pair movement in the photoconductor generates transient current, which can generate terahertz band electromagnetic wave. Because the mobility of the hole is much smaller than that of the electron, the current generated by the entire motion is very small and can be ignored.

10.5.3 Terahertz Detector

There are two methods for terahertz wave detection: photoconductive sampling and electro-optic sampling, as shown in Figs. 10.16 and 10.17, respectively. Figure 10.16 is the device diagram of photoconductive sampling detecting a terahertz pulse. The device of the photoconductive sampling detecting a terahertz pulse is very similar to that of the photoconductive antenna generating a terahertz wave. The difference is that there is no bias voltage on the photoconductive antenna in the detection device, and the pump and detection pulses have adjustable time delay through the time delay line. When the probe pulse irradiates the photoconductive antenna, a free carrier is

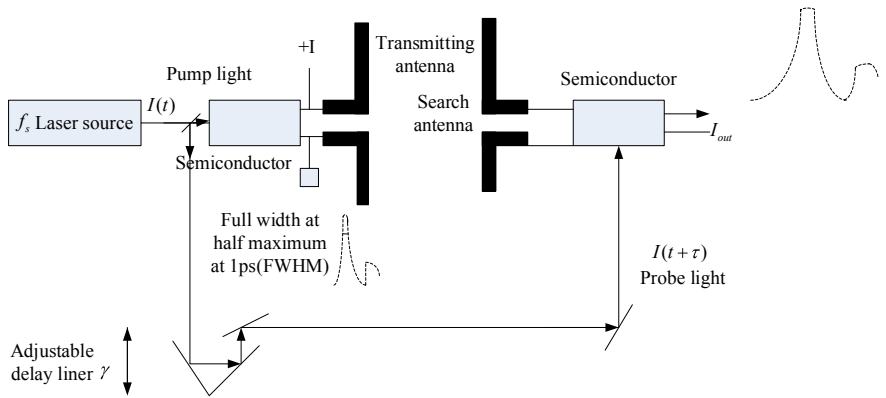
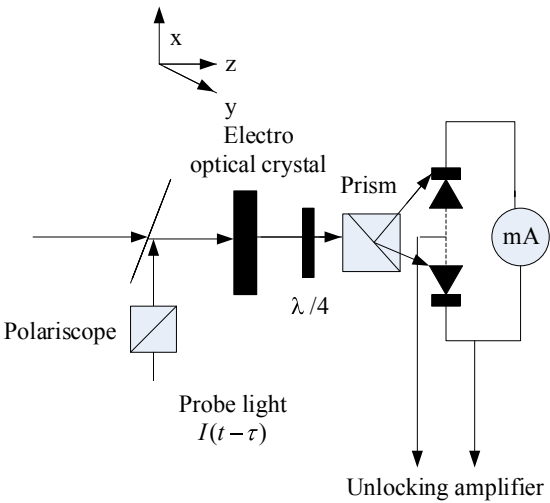


Fig. 10.16 Diagram of photoconductive sampling

Fig. 10.17 Diagram of electro-optical sampling



generated in the photoconductive conductor. The terahertz pulse irradiates the photoconductive antenna as a bias electric field to accelerate the carrier to generate current. Therefore, only when the probe pulse and terahertz pulse act on the photoconductive antenna at the same time, can current be generated. Since the pulse width of the detection light pulse is far less than that of the terahertz pulse, the measured current reflects the terahertz electric field at the moment of the arrival of the detection light pulse. The entire information of terahertz wave electric field can be obtained by adjusting the current measured by the delay line.

Figure 10.17 shows the structure of an electro-optic sampling and detecting device for terahertz pulse. When the terahertz pulse and the detection pulse pass through the electro-optic crystal at the same time, because the terahertz pulse will change the refractive index of the electro-optic crystal and cause transient birefringence, which will affect the propagation of the detection light pulse in the crystal the linear polarization detection light pulse can be changed into the elliptical polarization optical pulse. When the detection pulse propagates in the crystal, the degree of polarization changes. The information of the terahertz wave electric field can be obtained by measuring the change in the polarization degree. Similarly, changing the time delay of the terahertz and detection pulses can obtain the entire time domain waveform of the terahertz pulse.

10.5.4 Terahertz Wave Modulator

Encoding the information into the terahertz carrier is the primary problem faced by terahertz communication technology. The modulation of a terahertz wave is similar to that of an optical signal. According to the relationship between the terahertz wave and terahertz source, the modulation of the terahertz wave can be divided into internal and external modulation.

Internal modulation refers to the loading of modulation information during the formation of the terahertz wave; this means changing the parameters of the laser oscillation according to the law of the modulation signal to change the output characteristics of the terahertz wave. Examples include the voltage modulation scheme based on the photoconductive antenna and the high-speed phase modulation scheme based on the terahertz optical mixer.

External modulation means that after the formation of the terahertz wave, its parameters, such as frequency, intensity, and phase, are changed through the nonlinear effect of terahertz wave propagation in nonlinear materials. At present, the reported terahertz nonlinear materials for terahertz wave modulation include semiconductor materials, photonic crystals, meta-materials, ferroelectric materials, and liquid crystal materials. For the terahertz wave in wireless transmission, the internal modulation scheme has more advantages in the application of terahertz modulation.

Figure 10.18 shows a terahertz wave modulator model; LD1 and LD2 are two single-mode lasers, and two MZ isolators are used to prevent the influence of reflected waves on the system to ensure its stable operation. A half-wave plate and polarizer

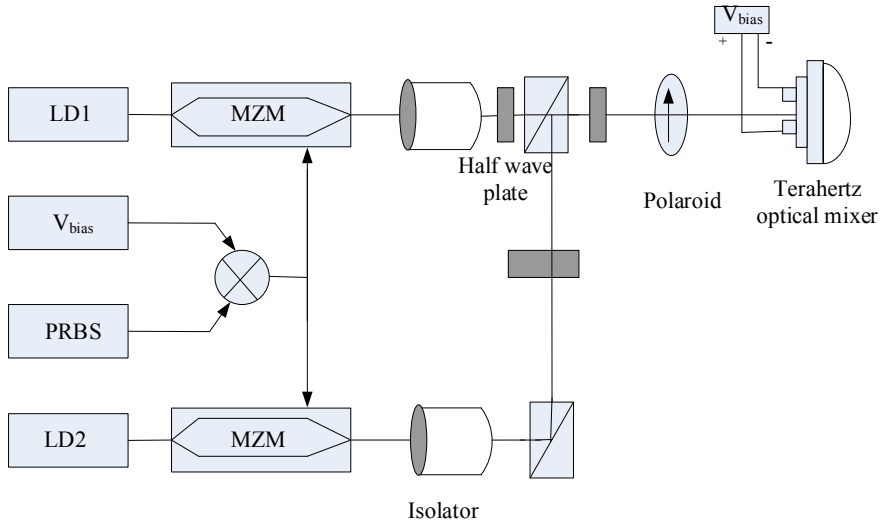


Fig. 10.18 Terahertz wave modulator model [54, 55]

are used to modulate the light intensity of two beams and make them equal. After passing through the polarizer, the light of two wavelengths is focused to the active region of the terahertz optical mixing device by focusing lens. The bias voltage of the modulator is the half-wave voltage of each wavelength of the two optical channels, and V_{bias} is a voltage pulse signal with a voltage amplitude equal to the half-wave voltage with a frequency of Ω ($\Omega = 20$ GHz), which is used to control the MZ modulator. When a pseudo-random binary sequence is 1, two beams of light are modulated into optical pulse signal by MZ modulator at frequency Ω respectively and mix in the active region of the optical mixer. At this time, the radiated terahertz carrier contains information sideband. Therefore, we can detect the information sideband with bolometer detection, indicating that the information code is 1. When the pseudo-random binary sequence is 0, the clock voltage signal is turned off, and the MZ modulator is no longer modulated; the intensity of the two beams remains unchanged, and a common terahertz wave mixing occurs. The terahertz wave radiated does not contain modulation information. Therefore, the detector bolometer can detect the information sideband, indicating that the information code is 0 [54, 55].

10.5.5 Transmission of Terahertz Waves in the Atmosphere

The absorption effect of the atmosphere causes the refractive index of the atmosphere to become a complex number. Its value depends on the pressure, temperature, and humidity in the atmosphere. It is a function of space–time and frequency. Because the resonance frequencies of oxygen and water vapor molecules in the atmosphere are

in the range of 0.01–1 THz, the complex refractive index characteristics of the atmosphere are complicated. Since the 1940s, many scientists have carried out theoretical and experimental studies on this issue. Van Vleck first deduced that there are many absorption lines near 60 GHz, with isolated oxygen absorption line at 112 GHz [56]. He also discovered the pressure broadening effect in atmospheric absorption. Liebe provided the parameters and empirical formulas of absorption spectra of oxygen and water vapor [57]. After that, Gibbins proposed a simplified formula for atmospheric absorption at sea level [58].

Terahertz wireless communication technology is one of the most promising application fields of terahertz, with a very high application value, especially suitable for interstellar communication, short-range atmospheric communication, and indoor broadband wireless communication [59–61]. In the atmosphere, terahertz communication technology can communicate and telemetry reentry vehicles such as missiles, satellites, and spaceships. Due to its ability to penetrate plasma, terahertz wave is expected to become the only effective communication tool to overcome the “black-out” phenomenon. In ground short-range wireless communication, THz can achieve wireless transmission rate of above 10 Gbps, which is significantly better than the current ultra-wideband technology, and can meet the needs of wireless communication technology in the next 10–20 years. The new generation communication system based on terahertz wave has the advantages of large capacity, high transmission rate, low eavesdropping rate, high anti-interference, and all-weather functionality.

10.6 Summary and Prospects

With the rapid development of wireless communication, new technologies and standards emerge constantly for digitalization, integration, broadband, and standards. Modern communication technology characterized by humanization and personalization will develop rapidly. In the next few years, the X-ray space communication, terahertz wave communication, quantum communication, neutrino communication, and gravitational wave communication will be developed in a large span.

10.7 Questions

- 10.1 What are the characteristics of X-ray communication?
- 10.2 What is a vortex? What are some examples of vortices in reality?
- 10.3 What is a vortex beam? What are the main characteristics?
- 10.4 What are the methods for generating a vortex beam?
- 10.5 Using the spatial light modulator to generate a vortex beam, what is the pattern loaded on the spatial light modulator? Explain the working principle.
- 10.6 What is the basis to realize the OAM multiplexing system? What is the system model?

- 10.7 What is gravitational wave? What are the difficulties in gravitational wave detection?
- 10.8 How many types are the neutrinos? What is the principle of neutrino communication?
- 10.9 What is quantum communication? What is entanglement?
- 10.10 Describe the characteristics of a terahertz wave.
- 10.11 Is the refractive index of a terahertz wave in the atmosphere a real number? Why?
- 10.12 Describe the principles of the internal and external modulation of terahertz wave modulator.
- 10.13 Please describe how a terahertz wave is generated.
- 10.14 Briefly describe the characteristics of the terahertz wave communication.
- 10.15 Briefly describe the teleportation of quantum.

10.8 Exercises

- 10.1 Consider a steel bar with weight $m = 4.9 \times 10^8$ g, radius 1 m, length 20 m, and ultimate strength 3×10^9 dyne / cm². How big is the gravitational wave radiated if the steel bar rotates around its centroid until it breaks?
- 10.2 If the incident wavelength is 0.1–1 THz and the refractive index of dust is 1.53–0.008i. Try to discuss the scattering of terahertz waves if the diameter of the sand is 0.01, 0.05 and 0.1 mm.

References

1. Zhao B, Wu C, Sheng L et al (2013) New generation deep space wireless communication based on X-ray. *Acta Photonica Sinica* 07:801–804
2. Deng N, Zhao B, Sheng L et al (2013) Space voice communication system based on X-ray. *Acta physica Sinica* 06:106–112
3. Zhang J, Li H, Li J et al (2015) Carbon nanotube field emission microfocus high speed X-ray tube. *Vacuum Electron Technol* 1(8–11):23
4. Shi W, Gui J, Wang H et al (2015) Design of measurement and control system for carbon nanotube field emission X-ray source. *Nuclear Electron Detect Technol* 2:172–175
5. Wang LQ, Su T, Zhao B et al (2015) Bit error rate analysis of X-ray communication system. *Acta Physica Sinica* 12:119–123
6. Ma X, Zhao B, Sheng L et al (2014) Study on grid controlled pulse generator for space X-ray Communication. *Acta Physica Sinica* 16:81–87
7. Yuan SC, Jia P, Lei T et al (2014) Optical vortex and orbital angular momentum optical communication. *Shenzhen Univ J Sci Technol* 31(4):331–346
8. Whewell W (1833) Essay towards a first approximation to a map of cotidal lines. *Proc R Soc Lond* 3(1):188–190
9. Billy R, Wolf E (1959) Electromagnetic diffraction in optical systems II. Structure of the image field in an aplanatic system. *Proc Royal Soc Lond A* 253(1274):358–379

10. Boivin A, Wolf E (1965) Electromagnetic field in the neighborhood of the focus of a coherent beam. *Phys Rev Lett* 138(6B):1561–1565
11. Carter WH (1973) Anomalies in the field of a Gaussian beam near focus. *Optics Commun* 7(3):211–218
12. Nye JF, Berry MV (1974) Dislocations in wave trains. *Proc Royal Soc Lond A* 336(1605):165–190
13. Baranova NB, Zeldovich BY, Mameev AV et al (1981) Dislocations of the wavefront of a speckle inhomogeneous field. *JETP Lett* 33(4):206–210
14. Baranova NB, Zeldovich BY, Mameev AV et al (1982) An investigation of the dislocation density of a wave front in light fields having a speckle structure. *Zhurnal Eksperimental'noi i Teroreticheskoi Fiziki* 83(52):1702–1710
15. Schwartzlander GA, Law CT (1992) Optical vortex solitons observed in Kerr nonlinear media. *Phys Rev Lett* 69(17):2503–2506
16. Voitsekovich VV, Kouznetsov D, Moronov DK (1998) Density of turbulence induced phase dislocations. *Appl Opt* 37(21):4525–4535
17. Bouchal Z, Celechovsky R (2004) Mixed vortex states of light as information carriers. *New J Phys* 6(6):1–15
18. Couillet P, Gil L, Rocca F (1989) Optical vortices. *Opt Commun* 73(89):403–408
19. Beijersbergen MW, Allen L, Veen VD et al (1993) Astigmatic laser mode converters and transfer of orbital angular momentum. *Opt Commun* 96(1–3):123–132
20. McGloin D, Simpson NB, Padgett MJ (1998) Transfer of orbital angular momentum from a stressed fiber-optic waveguide to a light beam. *Appl Opt* 37(3):469–472
21. Yao AM, Padgett MJ (2011) Orbital angular momentum: Origins, behavior and applications. *Adv Opt Photon* 3(2):161–204
22. Heckenberg NR, Mchuff R, Smith CP et al (1992) Generation of optical phase singularities by computer-generated holograms. *Opt Lett* 17(3):221–223
23. Bo B, MK, Zhao JL et al (2012) Vortex beam generation with a reflective pure phase liquid crystal spatial light modulator. *Optoelectron Laser* 23(1):74–78
24. Wang J, Yang JY, Fazal I M et al (2011) 25.6-bit/s/Hz spectral efficiency using 16-QAM signal over pol-muxed multiple orbital-angular-momentum modes. *IEEE photonics conference (PHO)*. Arlington: IEEE, pp 587–588
25. Yan Y, Wang J, Zhang L et al (2011) Fiber coupler for generating orbital angular momentum modes. *Opt Lett* 36(21):4269–4271
26. Yue A, Zhang L, Yan Y et al (2012) Octave-spanning supercontinuum generation of vortices in an As₂S₃ ring photonic crystal fiber. *Opt Lett* 37(11):1889–1891
27. Cai XL, Wang JW, Cai XL, Strain MJ et al (2012) Integrated compact optical vortex beam emitters. *Science* 338(6105):363–366
28. Lv H (2011) Vortex optical field orbital angular momentum for spatial optical quantum communication. Xi'an: Xi'an University of Technology, pp 1–2
29. Lu XH, Huang HQ, Zhao CL et al (2008) Vortex beams and optical vortices. *Adv Laser Optoelectron* 45(1):50–56
30. Wang J, Yang JY, Fazal IM et al (2012) Terabit free-space data transmission employing orbital angular momentum multiplexing. *Nat Photon* 6(7):488–496
31. Tamburini F, Mari E, Sponselli A et al (2012) Encoding many channels in the same frequency through radio vorticity: first experimental test. *New J Phys* 14(11):78001–78004
32. Krenn M, Fickler R, Fink M et al (2014) Twisted light communication through turbulent air across Vienna. *New J Phys* 16(11):1–9
33. Xu ZD, Gui CC, Li SH et al (2014) Fractional orbital angular momentum (OAM) free-space optical communications with atmospheric turbulence assisted by MIMO equalization. *Integr Photon Res, Silicon and Nanophotonics*
34. Huang H, Cao YW, Xie GD et al (2014) Crosstalk mitigation in a free-space orbital angular momentum multiplexed communication link using 4×4 MIMO equalization. *Opt Lett* 39(15):4360–4363

35. Ren YX, Wang Z, Xie GD et al (2016) Demonstration of OAM-based MIMO FSO link using spatial diversity and MIMO equalization for turbulence mitigation. 2016 optical fiber communications conference and exhibition (OFC)
36. Shi CZ, Dubois M, Wang Y et al (2017) High-speed acoustic communication by multiplexing orbital angular momentum. *Proc Natl Acad Sci USA* 114(28):7250–7253
37. Leach J, Jack B, Romero J et al (2010) Quantum correlations in optical angle-orbital angular momentum variables. *Science* 329(5992):662–665
38. Zhang Y, Li M (2004) Discussion on the mechanism of neutrino communication. *Proceedings of the 10th symposium of the chinese high tech cross research association of mathematical and mechanical physics*
39. Fang L, Liu Y (1982) Neutrino static mass and early universe clustering process. *Sci Bull* 04:201–203
40. Wu F (1995) Technology of neutrino communication principle. *Physics* 10:637
41. Li D (1994) On neutrino communication. *Modern Commun* 07:209
42. Lin Y (1996) Neutrino communication: a new revolution in the history of communication technology. *Digital Commun* 01(8–9):13
43. Xie H, Gao J, Liu C et al (2006) Neutrino communication to submarine. *Acta Instrumenta Sinica* S3:2071–2074
44. Hua W (1986) Experimental results of gravitational wave communication, the third kind of wave communication. *World Sci* 11:13–14
45. Wang D (1987) Gravitational wave positioning and communication technology. *Telecommun Technol* 05:1–5
46. Ke W (2005) Status quo of the world gravitational wave detection network. *Astronomical research and technology. J Natl Astron Observ* 3:199–203
47. Peng T (2013) Theory and design of gravity communication with Cavendish torsion balance. Shanghai Normal University, Shanghai
48. Tang M, Li F, Zhao P et al (2002) Gravitational wave, gravitational wave source and gravitational wave detection experiment. *J Yunnan Observ* 03:71–87
49. Wang Y, Zhu X, Liu J et al (2014) Laser interferometer gravitational wave detector. *Progress Astron* 03:348–382
50. Deng X (2014) The future of gravitational wave detection. *World Sci* 09(28–30):39
51. Long F, Zhou Z, Luo J (1998) Bending spring vibration isolation system for gravitational wave detector. *J Huazhong Univ Sci Technol* 07:14–16
52. Wang D (1990) Key problems of gravitational wave detection and positioning. *Telecommun Technol* 01:30–34
53. Huang Y, Wang Y, Tang K, Kang F et al (2007) New progress in gravitational wave theory and experiment. *Progress Astron* 01:58–73
54. Xiao J (2010) Research on terahertz wave generated by photoconductive antenna. Northwest University, Xi ' an
55. Chen S (2015) study on transmission characteristics of broadband terahertz wave in plasma., Xi ' an : Xi'an University of Technology
56. Jiang Y (2011) Research on terahertz wave modulation technology and terahertz wave filtering technology based on optical mixing device. wuhan:Huazhong University of Science and Technology
57. Sun D (2013) Research on terahertz wave control technology and related functional devices. University of Electronic Science and technology, Chengdu
58. Van Vleck JH, Weisskopf VF (1945) *Rev Modern Phys* 17(1):227–236
59. Leibe JH (1985) An updated model for millimeter wave propagation in moist air. *Radio Sci* 20(5):1069–1089
60. Gibbins CJ (1986) Improved algorithms for the determination of specific attenuation at sea level by dry air and water vapor, in the frequency range 1–359 GHz. *Radio Sci* 21(6):949–954
61. Huang S (2010) Study on transmission characteristics of terahertz wave. University of Electronic Science and technology, Xi'an