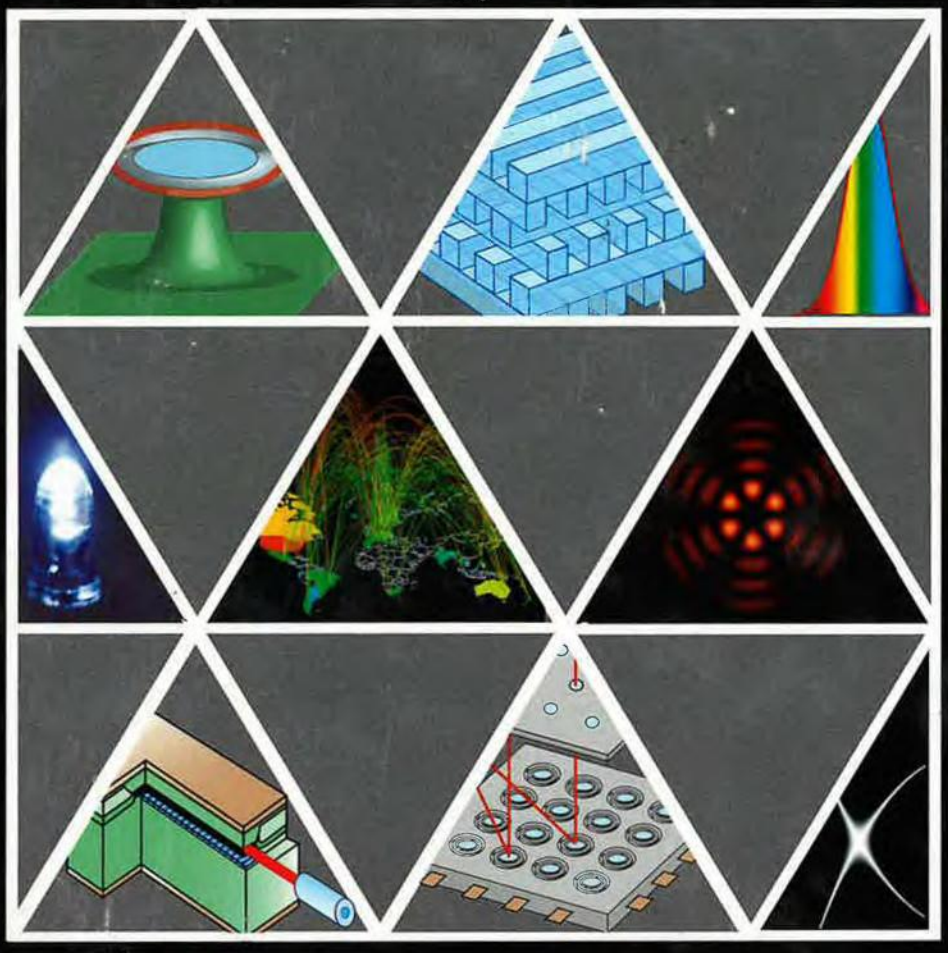
WILEY

FUNDAMENTALS OF PHOTONICS

Second Edition



B. E. A. Saleh
M. C. Teich

Wiley Series in Pure and Applied Optics • Bahaa E. A. Saleh, Series Editor

CYTEK V. BECKMAN
PGR2025-00084
BECKMAN 2036

FUNDAMENTALS OF PHOTONICS

FUNDAMENTALS OF PHOTONICS

SECOND EDITION

BAHAA E. A. SALEH
Boston University

MALVIN CARL TEICH
Boston University
Columbia University



WILEY-INTERSCIENCE
A John Wiley & Sons, Inc., Publication

Copyright © 2007 by John Wiley & Sons, Inc. All rights reserved.

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

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic format. For information about Wiley products, visit our web site at www.wiley.com.

Wiley Bicentennial Logo: Richard J. Pacifico

Library of Congress Cataloging-in-Publication Data is available.

ISBN: 978-0-471-35832-9

Printed in the United States of America

10 9 8 7 6 5 4 3 2

CONTENTS

PREFACE TO THE SECOND EDITION v

PREFACE TO THE FIRST EDITION xi

1 RAY OPTICS 1

1.1 Postulates of Ray Optics 3
1.2 Simple Optical Components 6
1.3 Graded-Index Optics 17
1.4 Matrix Optics 24
Reading List 34
Problems 35

2 WAVE OPTICS 38

2.1 Postulates of Wave Optics 40
2.2 Monochromatic Waves 41
*2.3 Relation Between Wave Optics and Ray Optics 49
2.4 Simple Optical Components 50
2.5 Interference 58
2.6 Polychromatic and Pulsed Light 66
Reading List 72
Problems 73

3 BEAM OPTICS 74

3.1 The Gaussian Beam 75
3.2 Transmission Through Optical Components 86
3.3 Hermite–Gaussian Beams 94
*3.4 Laguerre–Gaussian and Bessel Beams 97
Reading List 100
Problems 100

4 FOURIER OPTICS 102

4.1 Propagation of Light in Free Space 105
4.2 Optical Fourier Transform 116
4.3 Diffraction of Light 121
4.4 Image Formation 127
4.5 Holography 138
Reading List 145
Problems 147

5 ELECTROMAGNETIC OPTICS 150

5.1 Electromagnetic Theory of Light 152
5.2 Electromagnetic Waves in Dielectric Media 156
5.3 Monochromatic Electromagnetic Waves 162
5.4 Elementary Electromagnetic Waves 164
5.5 Absorption and Dispersion 170
5.6 Pulse Propagation in Dispersive Media 184
*5.7 Optics of Magnetic Materials and Metamaterials 190
Reading List 193
Problems 195

6 POLARIZATION OPTICS 197

6.1 Polarization of Light 199
6.2 Reflection and Refraction 209
6.3 Optics of Anisotropic Media 215
6.4 Optical Activity and Magneto-Optics 228
6.5 Optics of Liquid Crystals 232
6.6 Polarization Devices 235
Reading List 239
Problems 240

7 PHOTONIC-CRYSTAL OPTICS 243

7.1 Optics of Dielectric Layered Media 246
7.2 One-Dimensional Photonic Crystals 265
7.3 Two- and Three-Dimensional Photonic Crystals 279
Reading List 286
Problems 288

8 GUIDED-WAVE OPTICS 289

8.1 Planar-Mirror Waveguides 291
8.2 Planar Dielectric Waveguides 299
8.3 Two-Dimensional Waveguides 308
8.4 Photonic-Crystal Waveguides 311
8.5 Optical Coupling in Waveguides 313
8.6 Sub-Wavelength Metal Waveguides (Plasmonics) 321
Reading List 322
Problems 323

9 FIBER OPTICS 325

9.1 Guided Rays 327
9.2 Guided Waves 331
9.3 Attenuation and Dispersion 348
9.4 Holey and Photonic-Crystal Fibers 359
Reading List 362
Problems 363

10 RESONATOR OPTICS 365

10.1 Planar-Mirror Resonators 367
10.2 Spherical-Mirror Resonators 378
10.3 Two- and Three-Dimensional Resonators 390
10.4 Microresonators 394
Reading List 400
Problems 400

11 STATISTICAL OPTICS 403

11.1 Statistical Properties of Random Light 405
11.2 Interference of Partially Coherent Light 419
*11.3 Transmission of Partially Coherent Light Through Optical Systems 427
11.4 Partial Polarization 436
Reading List 440
Problems 442

12 PHOTON OPTICS 444

12.1 The Photon 446
12.2 Photon Streams 458
*12.3 Quantum States of Light 471
Reading List 476
Problems 478

13 PHOTONS AND ATOMS 482

13.1 Energy Levels 483
13.2 Occupation of Energy Levels 499
13.3 Interactions of Photons with Atoms 501
13.4 Thermal Light 517
13.5 Luminescence and Light Scattering 522
Reading List 528
Problems 530

14 LASER AMPLIFIERS 532

14.1 Theory of Laser Amplification 535
14.2 Amplifier Pumping 539
14.3 Common Laser Amplifiers 547
14.4 Amplifier Nonlinearity 556
*14.5 Amplifier Noise 562
Reading List 564
Problems 565

15 LASERS 567

15.1 Theory of Laser Oscillation 569
15.2 Characteristics of the Laser Output 575
15.3 Common Lasers 590
15.4 Pulsed Lasers 605
Reading List 621
Problems 624

16 SEMICONDUCTOR OPTICS 627

16.1 Semiconductors 629
16.2 Interactions of Photons with Charge Carriers 660
Reading List 675
Problems 677

17 SEMICONDUCTOR PHOTON SOURCES 680

17.1 Light-Emitting Diodes 682
17.2 Semiconductor Optical Amplifiers 702
17.3 Laser Diodes 716
17.4 Quantum-Confined and Microcavity Lasers 728
Reading List 741
Problems 745

18 SEMICONDUCTOR PHOTON DETECTORS 748

18.1 Photodetectors 749
18.2 Photoconductors 758
18.3 Photodiodes 762
18.4 Avalanche Photodiodes 767
18.5 Array Detectors 775
18.6 Noise in Photodetectors 777
Reading List 798
Problems 800

19 ACOUSTO-OPTICS 804

19.1 Interaction of Light and Sound 806
19.2 Acousto-Optic Devices 819
*19.3 Acousto-Optics of Anisotropic Media 828
Reading List 832
Problems 832

20 ELECTRO-OPTICS 834

20.1 Principles of Electro-Optics 836
*20.2 Electro-Optics of Anisotropic Media 849
20.3 Electro-Optics of Liquid Crystals 856
*20.4 Photorefractivity 863
20.5 Electroabsorption 868
Reading List 869
Problems 871

21 NONLINEAR OPTICS 873

21.1 Nonlinear Optical Media 875
21.2 Second-Order Nonlinear Optics 879
21.3 Third-Order Nonlinear Optics 894
*21.4 Second-Order Nonlinear Optics: Coupled-Wave Theory 905
*21.5 Third-Order Nonlinear Optics: Coupled-Wave Theory 917
*21.6 Anisotropic Nonlinear Media 924
*21.7 Dispersive Nonlinear Media 927
Reading List 932
Problems 934

22 ULTRAFAST OPTICS 936

22.1 Pulse Characteristics 937
22.2 Pulse Shaping and Compression 946
22.3 Pulse Propagation in Optical Fibers 960

22.4 Ultrafast Linear Optics 973
22.5 Ultrafast Nonlinear Optics 984
22.6 Pulse Detection 999
Reading List 1011
Problems 1013

23 OPTICAL INTERCONNECTS AND SWITCHES 1016

23.1 Optical Interconnects 1018
23.2 Passive Optical Routers 1030
23.3 Photonic Switches 1038
23.4 Optical Gates 1058
Reading List 1069
Problems 1071

24 OPTICAL FIBER COMMUNICATIONS 1072

24.1 Fiber-Optic Components 1074
24.2 Optical Fiber Communication Systems 1084
24.3 Modulation and Multiplexing 1101
24.4 Fiber-Optic Networks 1106
24.5 Coherent Optical Communications 1112
Reading List 1118
Problems 1120

A FOURIER TRANSFORM 1122

A.1 One-Dimensional Fourier Transform 1122
A.2 Time Duration and Spectral Width 1124
A.3 Two-Dimensional Fourier Transform 1128
Reading List 1131

B LINEAR SYSTEMS 1132

B.1 One-Dimensional Linear Systems 1132
B.2 Two-Dimensional Linear Systems 1135
Reading List 1136

C MODES OF LINEAR SYSTEMS 1137

SYMBOLS AND UNITS 1142

AUTHORS 1159

INDEX 1161

CHAPTER

1

RAY OPTICS

1.1 POSTULATES OF RAY OPTICS 3

1.2 SIMPLE OPTICAL COMPONENTS 6

- A. Mirrors
- B. Planar Boundaries
- C. Spherical Boundaries and Lenses
- D. Light Guides

1.3 GRADED-INDEX OPTICS 17

- A. The Ray Equation
- B. Graded-Index Optical Components
- *C. The Eikonal Equation

1.4 MATRIX OPTICS 24

- A. The Ray-Transfer Matrix
- B. Matrices of Simple Optical Components
- C. Matrices of Cascaded Optical Components
- D. Periodic Optical Systems



Sir Isaac Newton (1642–1727) set forth a theory of optics in which light emissions consist of collections of corpuscles that propagate rectilinearly.



Pierre de Fermat (1601–1665) enunciated the principle that light travels along the path of least time.

Light can be described as an electromagnetic wave phenomenon governed by the same theoretical principles that govern all other forms of electromagnetic radiation, such as radio waves and X-rays. This conception of light is called **electromagnetic optics**. Electromagnetic radiation propagates in the form of two mutually coupled *vector* waves, an electric-field wave and a magnetic-field wave. Nevertheless, it is possible to describe many optical phenomena using a simplified *scalar* wave theory in which light is described by a single scalar wavefunction. This approximate way of treating light is called scalar wave optics, or simply **wave optics**.

When light waves propagate through and around objects whose dimensions are much greater than the wavelength of the light, the wave nature is not readily discerned and the behavior of light can be adequately described by rays obeying a set of geometrical rules. This model of light is called **ray optics**. From a mathematical perspective, ray optics is the limit of wave optics when the wavelength is infinitesimally small.

Thus, electromagnetic optics encompasses wave optics, which, in turn, encompasses ray optics, as illustrated in Fig. 1.0-1. Ray optics and wave optics are approximate theories that derive their validity from their successes in producing results that approximate those based on the more rigorous electromagnetic theory.

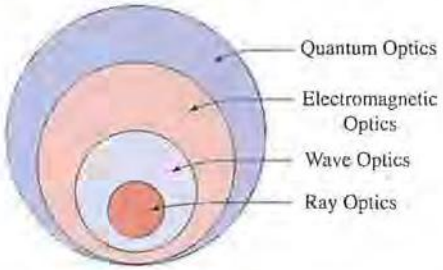


Figure 1.0-1 The theory of quantum optics provides an explanation of virtually all optical phenomena. The electromagnetic theory of light (electromagnetic optics) provides the most complete treatment of light within the confines of classical optics. Wave optics is a scalar approximation of electromagnetic optics. Ray optics is the limit of wave optics when the wavelength is very short.

Although electromagnetic optics provides the most complete treatment of light within the confines of **classical optics**, certain optical phenomena are characteristically quantum mechanical in nature and cannot be explained classically. These phenomena are described by a quantum version of electromagnetic theory known as **quantum electrodynamics**. For optical phenomena, this theory is also referred to as **quantum optics**.

Historically, the theories of optics developed roughly in the following sequence: (1) ray optics → (2) wave optics → (3) electromagnetic optics → (4) quantum optics. These models are progressively more complex and sophisticated, and were developed successively to provide explanations for the outcomes of increasingly subtle and precise optical experiments. The optimal choice of a model is the simplest one that satisfactorily describes a particular phenomenon, but it is sometimes difficult to know *a priori* which model will achieve this. Fortunately, however, experience often provides a good guide.

For pedagogical reasons, the initial chapters in this book follow the historical order indicated above. Each model of light begins with a set of postulates (provided without proof), from which a large body of results are generated. The postulates of each model are shown to arise in special cases of the next-higher-level model. In this chapter we begin with ray optics.

This Chapter

Ray optics is the simplest theory of light. Light is described by rays that travel in different optical media in accordance with a set of geometrical rules. Ray optics is therefore also called **geometrical optics**. Ray optics is an approximate theory. Although it adequately describes most of our daily experiences with light, there are many phenomena that ray optics does not adequately describe (as amply attested to by the remaining chapters of this book).

Ray optics is concerned with the *location* and *direction* of light rays. It is therefore useful in studying *image formation* — the collection of rays from each point of an object and their redirection by an optical component onto a corresponding point of an image. Ray optics permits us to determine conditions under which light is guided within a given medium, such as a glass fiber. In isotropic media, optical rays point in the direction of the flow of *optical energy*. Ray bundles can be constructed in which the density of rays is proportional to the density of light energy. When light is generated isotropically from a point source, for example, the energy associated with the rays in a given cone is proportional to the solid angle of the cone. Rays may be traced through an optical system to determine the optical energy crossing a given area.

This chapter begins with a set of postulates from which the simple rules that govern the propagation of light rays through optical media are derived. In Sec. 1.2 these rules are applied to simple optical components such as mirrors and planar or spherical boundaries between different optical media. Ray propagation in inhomogeneous (graded-index) optical media is examined in Sec. 1.3. Graded-index optics is the basis of a technology that has become an important part of modern optics.

Optical components are often centered about an optical axis, about which the rays travel at small inclinations. Such rays are called **paraxial rays**. This assumption is the basis of **paraxial optics**. The change in the position and inclination of a paraxial ray as it travels through an optical system can be efficiently described by the use of a 2×2-matrix algebra. Section 1.4 is devoted to this algebraic tool, called **matrix optics**.

1.1 POSTULATES OF RAY OPTICS

Postulates of Ray Optics

- Light travels in the form of rays. The rays are emitted by light sources and can be observed when they reach an optical detector.
- An optical medium is characterized by a quantity $n \geq 1$, called the **refractive index**. The refractive index $n = c_o/c$ where c_o is the speed of light in free space and c is the speed of light in the medium. Therefore, the time taken by light to travel a distance d is $d/c = nd/c_o$. It is proportional to the product nd, which is known as the **optical pathlength**.
- In an inhomogeneous medium the refractive index $n(\mathbf{r})$ is a function of the position $\mathbf{r} = (x, y, z)$. The optical path length along a given path between two points A and B is therefore

$$\text{Optical pathlength} = \int_A^B n(\mathbf{r})\, ds, \tag{1.1-1}$$

where ds is the differential element of length along the path. The time taken by light to travel from A to B is proportional to the optical pathlength.

■ **Fermat's Principle**. Optical rays traveling between two points, A and B, follow a path such that the time of travel (or the optical pathlength) between the two points is an extremum relative to neighboring paths. This is expressed mathematically as

$$\delta \int_A^B n(\mathbf{r})\, ds = 0, \tag{1.1-2}$$

where the symbol δ, which is read "the variation of," signifies that the optical pathlength is either minimized or maximized, or is a point of inflection. It is, however, usually a minimum, in which case:

Light rays travel along the path of least time.

Sometimes the minimum time is shared by more than one path, which are then all followed simultaneously by the rays. An example in which the pathlength is maximized is provided in Prob. 1.1-2.

In this chapter we use the postulates of ray optics to determine the rules governing the propagation of light rays, their reflection and refraction at the boundaries between different media, and their transmission through various optical components. A wealth of results applicable to numerous optical systems are obtained without the need for any other assumptions or rules regarding the nature of light.

Propagation in a Homogeneous Medium

In a homogeneous medium the refractive index is the same everywhere, and so is the speed of light. The path of minimum time, required by Fermat's principle, is therefore also the path of minimum distance. The principle of the *path of minimum distance* is known as **Hero's principle**. The path of minimum distance between two points is a straight line so that *in a homogeneous medium, light rays travel in straight lines* (Fig. 1.1-1).

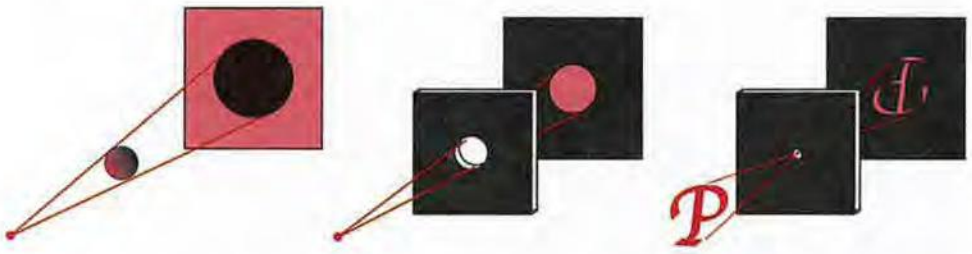


Figure 1.1-1 Light rays travel in straight lines. Shadows are perfect projections of stops.

Reflection from a Mirror

Mirrors are made of certain highly polished metallic surfaces, or metallic or dielectric films deposited on a substrate such as glass. Light reflects from mirrors in accordance with the law of reflection:

The reflected ray lies in the plane of incidence; the angle of reflection equals the angle of incidence.

The plane of incidence is the plane formed by the incident ray and the normal to the mirror at the point of incidence. The angles of incidence and reflection, θ and θ', are defined in Fig. 1.1-2(a). To prove the law of reflection we simply use Hero's principle. Examine a ray that travels from point A to point C after reflection from the planar mirror in Fig. 1.1-2(b). According to Hero's principle, for a mirror of infinitesimal thickness, the distance $\overline{AB} + \overline{BC}$ must be minimum. If C' is a mirror image of C, then $\overline{BC} = \overline{BC'}$, so that $\overline{AB} + \overline{BC'}$ must be a minimum. This occurs when $\overline{ABC'}$ is a straight line, i.e., when B coincides with B' so that $\theta = \theta'$.

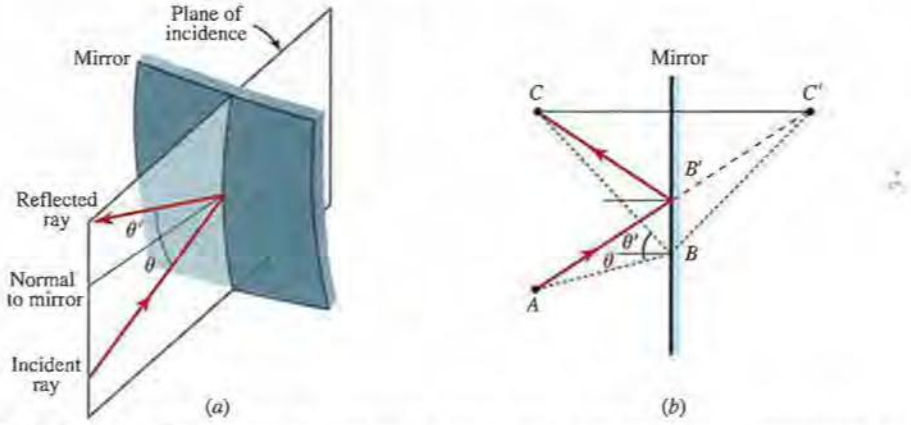


Figure 1.1-2 (a) Reflection from the surface of a curved mirror. (b) Geometrical construction to prove the law of reflection.

Reflection and Refraction at the Boundary Between Two Media

At the boundary between two media of refractive indexes n_1 and n_2 an incident ray is split into two — a reflected ray and a refracted (or transmitted) ray (Fig. 1.1-3). The reflected ray obeys the law of reflection. The refracted ray obeys the law of refraction:

The refracted ray lies in the plane of incidence; the angle of refraction θ_2 is related to the angle of incidence θ_1 by Snell's law,

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \tag{1.1-3}$$

Snell's law

The proportion in which the light is reflected and refracted is not described by ray optics.

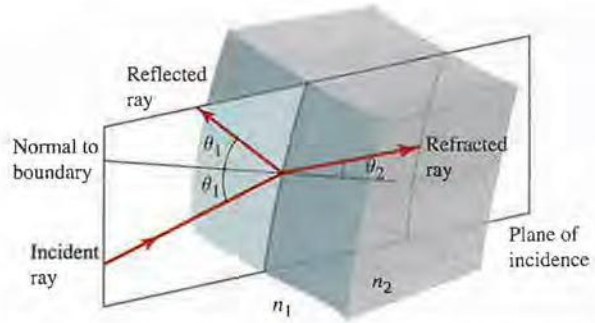


Figure 1.1-3 Reflection and refraction at the boundary between two media.

EXERCISE 1.1-1

Proof of Snell's Law. The proof of Snell's law is an exercise in the application of Fermat's principle. Referring to Fig. 1.1-4, we seek to minimize the optical pathlength $n_1\overline{AB}+n_2\overline{BC}$ between points A and C. We therefore have the following optimization problem: Minimize $n_1 d_1 \sec\theta_1 + n_2 d_2 \sec\theta_2$ with respect to the angles θ_1 and θ_2, subject to the condition $d_1 \tan\theta_1 + d_2 \tan\theta_2 = d$. Show that the solution of this constrained minimization problem yields Snell's law.

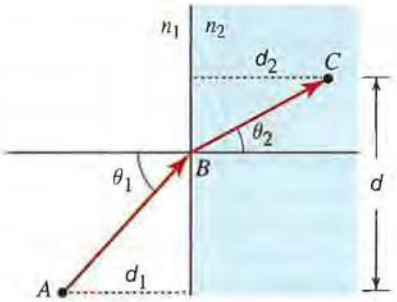


Figure 1.1-4 Construction to prove Snell's law.

The three simple rules — propagation in straight lines and the laws of reflection and refraction — are applied in Sec. 1.2 to several geometrical configurations of mirrors and transparent optical components, without further recourse to Fermat's principle.

1.2 SIMPLE OPTICAL COMPONENTS

A. Mirrors

Planar Mirrors

A planar mirror reflects the rays originating from a point P_1 such that the reflected rays appear to originate from a point P_2 behind the mirror, called the image (Fig. 1.2-1).

Paraboloidal Mirrors

The surface of a paraboloidal mirror is a paraboloid of revolution. It has the useful property of focusing all incident rays parallel to its axis to a single point called the **focus**. The distance $\overline{PF} = f$ defined in Fig. 1.2-2 is called the **focal length**. Paraboloidal

mirrors are often used as light-collecting elements in telescopes. They are also used for making parallel beams of light from point sources such as in flashlights.

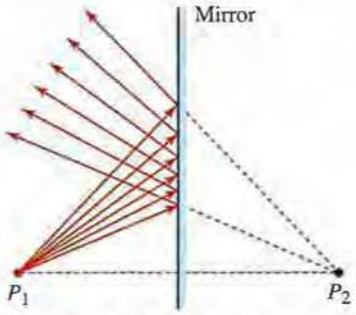


Figure 1.2-1 Reflection of light from a planar mirror.

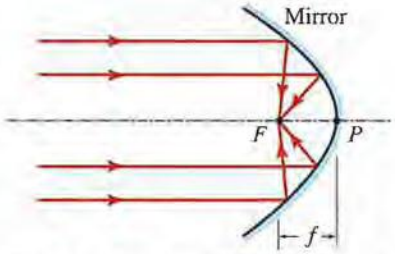


Figure 1.2-2 Focusing of light by a paraboloidal mirror.

Elliptical Mirrors

An elliptical mirror reflects all the rays emitted from one of its two foci, e.g., P_1, and images them onto the other focus, P_2 (Fig. 1.2-3). In accordance with Hero's principle, the distances traveled by the light from P_1 to P_2 along any of the paths are equal.

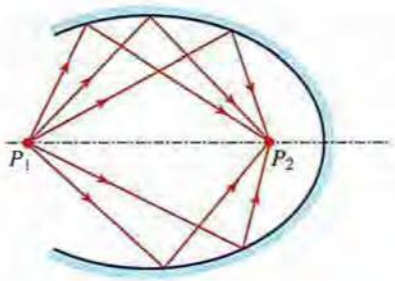


Figure 1.2-3 Reflection from an elliptical mirror.

Spherical Mirrors

A spherical mirror is easier to fabricate than a paraboloidal mirror or an elliptical mirror. However, it has neither the focusing property of the paraboloidal mirror nor the imaging property of the elliptical mirror. As illustrated in Fig. 1.2-4, parallel rays meet the axis at different points; their envelope (the dashed curve) is called the caustic curve. Nevertheless, parallel rays close to the axis are approximately focused onto a single point F at distance $(-R)/2$ from the mirror center C. By convention, R is negative for concave mirrors and positive for convex mirrors.

Paraxial Rays Reflected from Spherical Mirrors

Rays that make small angles (such that $\sin\theta \approx \theta$) with the mirror's axis are called **paraxial rays**. In the **paraxial approximation**, where only paraxial rays are considered, a spherical mirror has a focusing property like that of the paraboloidal mirror *and* an imaging property like that of the elliptical mirror. The body of rules that results from this approximation forms **paraxial optics**, also called first-order optics or Gaussian optics.

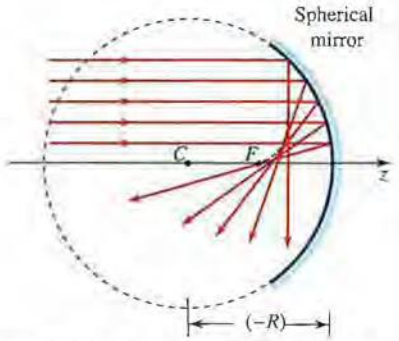


Figure 1.2-4 Reflection of parallel rays from a concave spherical mirror.

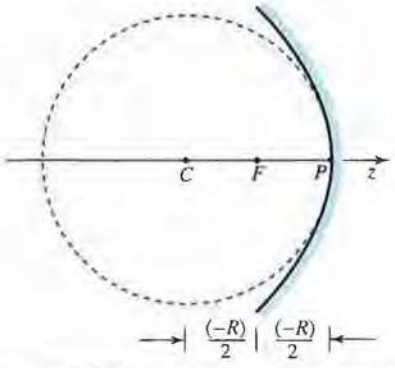


Figure 1.2-5 A spherical mirror approximates a paraboloidal mirror for paraxial rays.

A spherical mirror of radius R therefore acts like a paraboloidal mirror of focal length $f = R/2$. This is in fact plausible since at points near the axis, a parabola can be approximated by a circle with radius equal to the parabola's radius of curvature (Fig. 1.2-5).

All paraxial rays originating from each point on the axis of a spherical mirror are reflected and focused onto a single corresponding point on the axis. This can be seen (Fig. 1.2-6) by examining a ray emitted at an angle θ_1 from a point P_1 at a distance z_1 away from a concave mirror of radius R, and reflecting at angle $(-\theta_2)$ to meet the axis at a point P_2 that is a distance z_2 away from the mirror. The angle θ_2 is negative since the ray is traveling downward. Since the three angles of a triangle add to 180°, we have $\theta_1 = \theta_0 - \theta$ and $(-\theta_2) = \theta_0 + \theta$, so that $(-\theta_2) + \theta_1 = 2\theta_0$. If θ_0 is sufficiently small, the approximation $\tan\theta_0 \approx \theta_0$ may be used, so that $\theta_0 \approx y/(-R)$, from which

$$(-\theta_2) + \theta_1 \approx \frac{2y}{(-R)}, \tag{1.2-1}$$

where y is the height of the point at which the reflection occurs. Recall that R is negative since the mirror is concave. Similarly, if θ_1 and θ_2 are small, $\theta_1 \approx y/z_1$ and $(-\theta_2) = y/z_2$, so that (1.2-1) yields $y/z_1 + y/z_2 \approx 2y/(-R)$, whereupon

$$\frac{1}{z_1} + \frac{1}{z_2} \approx \frac{2}{(-R)}. \tag{1.2-2}$$

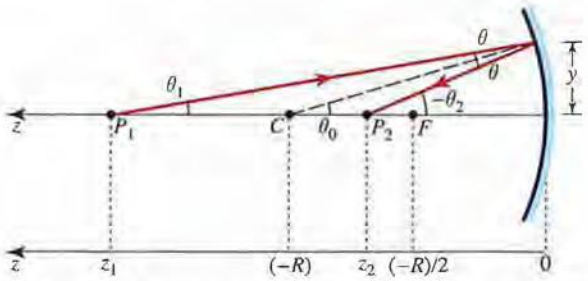


Figure 1.2-6 Reflection of paraxial rays from a concave spherical mirror of radius $R < 0$.

This relation holds regardless of y (i.e., regardless of θ_1) as long as the approximation is valid. This means that all paraxial rays originating from point P_1 arrive at P_2. The distances z_1 and z_2 are measured in a coordinate system in which the z axis points to the left. Points of negative z therefore lie to the right of the mirror.

According to (1.2-2), rays that are emitted from a point very far out on the z axis ($z_1 = \infty$) are focused to a point F at a distance $z_2 = (-R)/2$. This means that within the paraxial approximation, all rays coming from infinity (parallel to the axis of the mirror) are focused to a point at a distance f from the mirror, which is known as its focal length:

$$f = \frac{(-R)}{2}, \tag{1.2-3}$$

(1.2-3) Focal Length Spherical Mirror

Equation (1.2-2) is usually written in the form

$$\frac{1}{z_1} + \frac{1}{z_2} = \frac{1}{f}, \tag{1.2-4}$$

(1.2-4) Imaging Equation (Paraxial Rays)

which is known as the imaging equation. Both the incident and the reflected rays must be paraxial for this equation to hold.

EXERCISE 1.2-1

Image Formation by a Spherical Mirror. Show that within the paraxial approximation, rays originating from a point $P_1 = (y_1, z_1)$ are reflected to a point $P_2 = (y_2, z_2)$, where z_1 and z_2 satisfy (1.2-4) and $y_2 = -y_1 z_2/z_1$ (Fig. 1.2-7). This means that rays from each point in the plane $z = z_1$ meet at a single corresponding point in the plane $z = z_2$, so that the mirror acts as an image-formation system with magnification $-z_2/z_1$. Negative magnification means that the image is inverted.

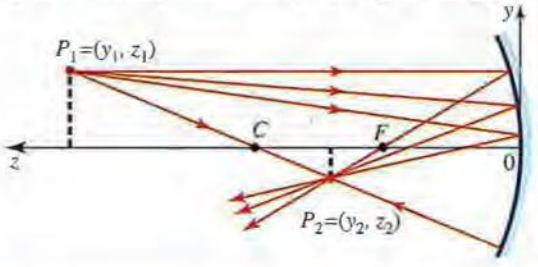


Figure 1.2-7 Image formation by a spherical mirror. Four particular rays are illustrated.

B. Planar Boundaries

The relation between the angles of refraction and incidence, θ_2 and θ_1, at a planar boundary between two media of refractive indexes n_1 and n_2 is governed by Snell's law (1.1-3). This relation is plotted in Fig. 1.2-8 for two cases:

- *External Refraction* ($n_1 < n_2$). When the ray is incident from the medium of smaller refractive index, $\theta_2 < \theta_1$ and the refracted ray bends away from the boundary.

- *Internal Refraction* ($n_1 > n_2$). If the incident ray is in a medium of higher refractive index, $\theta_2 > \theta_1$ and the refracted ray bends toward the boundary.

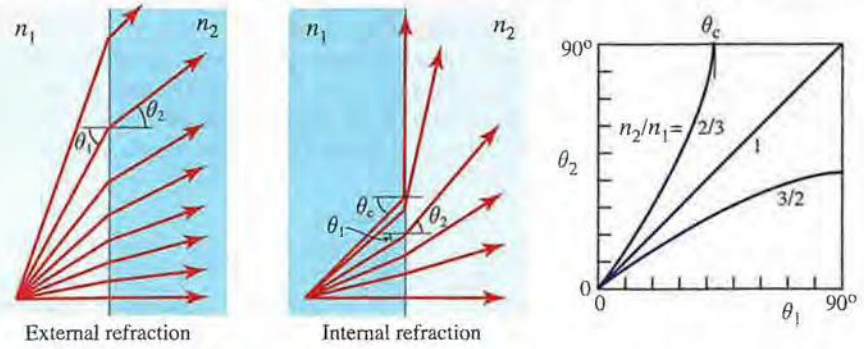


Figure 1.2-8 Relation between the angles of refraction and incidence.

The refracted rays bend in such a way as to minimize the optical pathlength, i.e., to increase the pathlength in the lower-index medium at the expense of pathlength in the higher-index medium. In both cases, when the angles are small (i.e., the rays are paraxial), the relation between θ_2 and θ_1 is approximately linear, $n_1\theta_1 \approx n_2\theta_2$, or $\theta_2 \approx (n_1/n_2)\theta_1$.

Total Internal Reflection

For internal refraction ($n_1 > n_2$), the angle of refraction is greater than the angle of incidence, $\theta_2 > \theta_1$, so that as θ_1 increases, θ_2 reaches 90° first (see Fig. 1.2-8). This occurs when $\theta_1 = \theta_c$ (the **critical angle**), with $n_1 \sin\theta_c = n_2 \sin(\pi/2) = n_2$, so that

$$\theta_c = \sin^{-1}\frac{n_2}{n_1}. \qquad (1.2\text{-}5)$$

Critical Angle

When $\theta_1 > \theta_c$, Snells' law (1.1-3) cannot be satisfied and refraction does not occur. The incident ray is totally reflected as if the surface were a perfect mirror [Fig. 1.2-9(*a*)]. The phenomenon of total internal reflection is the basis of many optical devices and systems, such as reflecting prisms [see Fig. 1.2-9(*b*)] and optical fibers (see Sec. 1.2D). It can be shown using electromagnetic optics (Fresnel's equations in Chapter 6) that all of the energy is carried by the reflected light so that the process of total internal reflection is highly efficient.

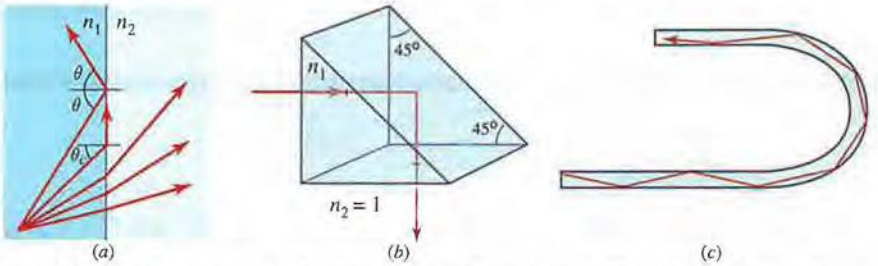


Figure 1.2-9 (*a*) Total internal reflection at a planar boundary. (*b*) The reflecting prism. If $n_1 > \sqrt{2}$ and $n_2 = 1$ (air), then $\theta_c < 45°$; since $\theta_1 = 45°$, the ray is totally reflected. (*c*) Rays are guided by total internal reflection from the internal surface of an optical fiber.

Prisms

A prism of apex angle α and refractive index n (Fig. 1.2-10) deflects a ray incident at an angle θ by an angle

$$\theta_d = \theta - \alpha + \sin^{-1}\left[\sqrt{n^2 - \sin^2\theta}\,\sin\alpha - \sin\theta\cos\alpha\right]. \tag{1.2-6}$$

This may be shown by using Snell's law twice at the two refracting surfaces of the prism. When α is very small (thin prism) and θ is also very small (paraxial approximation), (1.2-6) is approximated by

$$\theta_d \approx (n-1)\alpha. \tag{1.2-7}$$

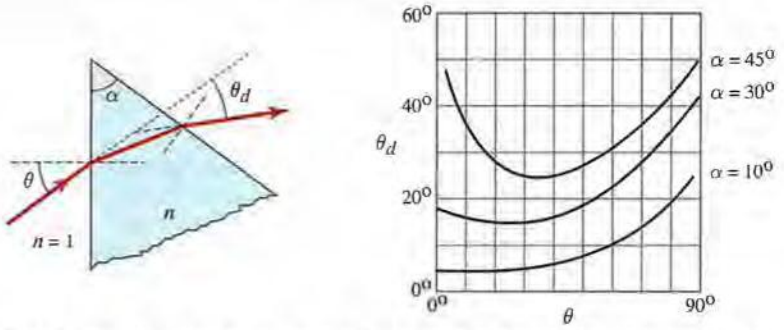
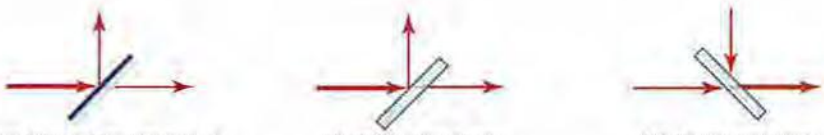


Figure 1.2-10 Ray deflection by a prism. The angle of deflection θ_d is a function of the angle of incidence θ for different apex angles α when $n = 1.5$. When both α and θ are small $\theta_d \approx (n-1)\alpha$, which is approximately independent of θ. When $\alpha = 45°$ or $\theta = 0°$, total internal reflection occurs, as illustrated in Fig. 1.2-9(*b*).

Beamsplitters

The beamsplitter is an optical component that splits the incident beam into a reflected beam and a transmitted beam, as illustrated in Fig. 1.2-11. Beamsplitters are also frequently used to combine two light beams into one [Fig. 1.2-11(*c*)]. Beamsplitters are often constructed by depositing a thin semitransparent metallic or dielectric film on a glass substrate. A thin glass plate or a prism can also serve as a beamsplitter.



(*a*) Partially reflective mirror (*b*) Thin glass plate (*c*) Beam combiner

Figure 1.2-11 Beamsplitters and combiners.

C. Spherical Boundaries and Lenses

We now examine the refraction of rays from a spherical boundary of radius R between two media of refractive indexes n_1 and n_2. By convention, R is positive for a convex boundary and negative for a concave boundary. The results are obtained by applying Snell's law, which relates the angles of incidence and refraction relative to the normal to the surface, defined by the radius vector from the center $\mathcal{C}$. These angles are to be distinguished from the angles θ_1 and θ_2, which are defined relative to the z axis. Considering only paraxial rays making small angles with the axis of the system so that $\sin\theta \approx \theta$ and $\tan\theta \approx \theta$, the following properties may be shown to hold:

- A ray making an angle θ_1 with the z axis and meeting the boundary at a point of height y where it makes an angle θ_0 with the radius vector [see Fig. 1.2-12(a)] changes direction at the boundary so that the refracted ray makes an angle θ_2 with the z axis and an angle θ_3 with the radius vector. The angle of incidence is therefore $(\theta_1 + \theta_2)$ while the angle of refraction is θ_3, so that

$$\theta_2 \approx \frac{n_1}{n_2}\theta_1 - \frac{n_2 - n_1}{n_2}\frac{y}{R}. \tag{1.2-8}$$

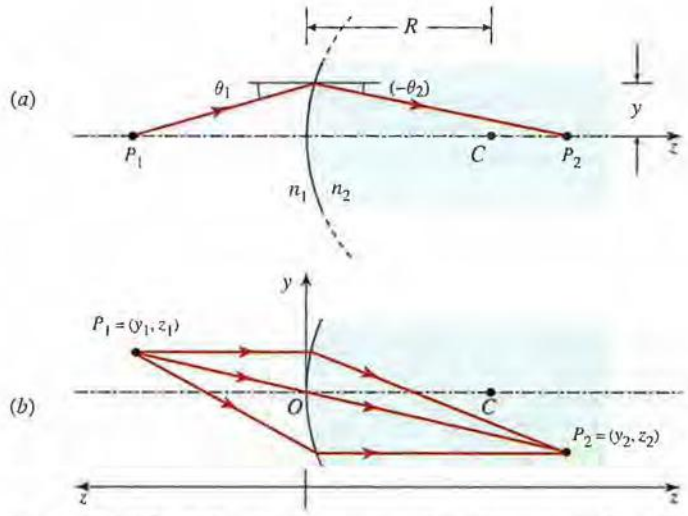


Figure 1.2-12 Refraction at a convex spherical boundary ($R > 0$).

- All paraxial rays originating from a point $P_1 = (y_1, z_1)$ in the $z = z_1$ plane meet at a point $P_2 = (y_2, z_2)$ in the $z = z_2$ plane, where

$$\frac{n_1}{z_1} + \frac{n_2}{z_2} \approx \frac{n_2 - n_1}{R} \tag{1.2-9}$$

and

$$y_2 = -\frac{n_1}{n_2}\frac{z_2}{z_1}y_1. \tag{1.2-10}$$

The $z = z_1$ and $z = z_2$ planes are said to be conjugate planes. Every point in the first plane has a corresponding point (image) in the second with magnification $-(n_1/n_2)(z_2/z_1)$. Again, negative magnification means that the image is inverted. By convention P_1 is measured in a coordinate system pointing to the left and P_2 in a coordinate system pointing to the right (e.g., if P_2 lies to the left of the boundary, then z_2 would be negative).

The similarities between these properties and those of the spherical mirror are evident. It is important to remember that the image formation properties described above are approximate. They hold only for paraxial rays. Rays of large angles do not obey these paraxial laws; the deviation results in image distortion called **aberration**.

EXERCISE 1.2-2

Image Formation. Derive (1.2-8). Prove that paraxial rays originating from P_1 pass through P_2 when (1.2-9) and (1.2-10) are satisfied.

EXERCISE 1.2-3

Aberration-Free Imaging Surface. Determine the equation of a convex aspherical (nonspherical) surface between media of refractive indexes n_1 and n_2 such that all rays (not necessarily paraxial) from an axial point P_1 at a distance z_1 to the left of the surface are imaged onto an axial point P_2 at a distance z_2 to the right of the surface [Fig. 1.2-12(*a*)]. *Hint:* In accordance with Fermat's principle the optical pathlengths between the two points must be equal for all paths.

Lenses

A spherical lens is bounded by two spherical surfaces. It is, therefore, defined completely by the radii R_1 and R_2 of its two surfaces, its thickness Δ, and the refractive index n of the material (Fig. 1.2-13). A glass lens in air can be regarded as a combination of two spherical boundaries, air-to-glass and glass-to-air.

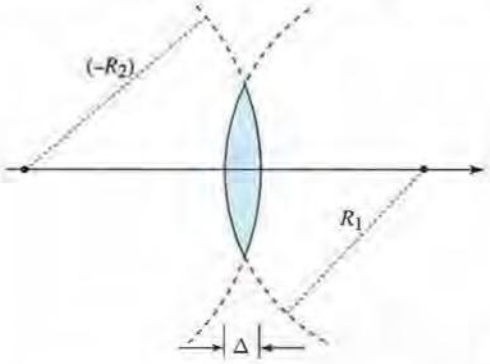


Figure 1.2-13 A biconvex spherical lens.

A ray crossing the first surface at height y and angle θ_1 with the z axis [Fig. 1.2-14(*a*)] is traced by applying (1.2-8) at the first surface to obtain the inclination angle θ of the refracted ray, which we extend until it meets the second surface. We then use (1.2-8) once more with θ replacing θ_1 to obtain the inclination angle θ_2 of the ray after refraction from the second surface. The results are in general complicated. When the lens is thin, however, it can be assumed that the incident ray emerges from the lens at about the same height y at which it enters. Under this assumption, the following relations follow:

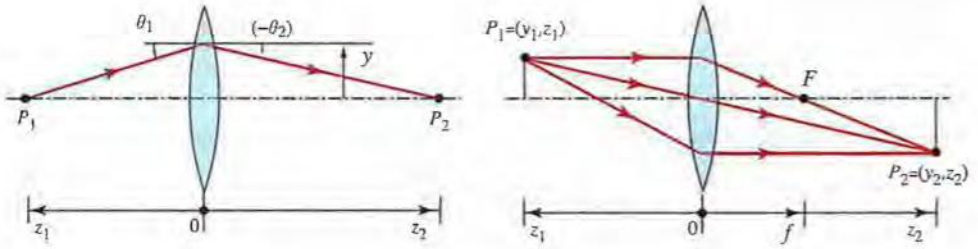
- The angles of the refracted and incident rays are related by

$$\theta_2 = \theta_1 - \frac{y}{f}, \tag{1.2-11}$$

where f, called the **focal length**, is given by

$$\boxed{\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right).} \tag{1.2-12}$$

(1.2-12) Focal Length Thin Spherical Lens



(a) (b)

Figure 1.2-14 (a) Ray bending by a thin lens. (b) Image formation by a thin lens.

- All rays originating from a point $P_1 = (y_1, z_1)$ meet at a point $P_2 = (y_2, z_2)$ [Fig. 1.2-14(b)], where

$$\boxed{\frac{1}{z_1} + \frac{1}{z_2} = \frac{1}{f}} \tag{1.2-13}$$

(1.2-13) Imaging Equation

and

$$\boxed{y_2 = -\frac{z_2}{z_1} y_1.} \tag{1.2-14}$$

(1.2-14) Magnification

These results are identical to those for the spherical mirror [see (1.2-4) and Exercise 1.2-1].

These equations indicate that each point in the $z = z_1$ plane is imaged onto a corresponding point in the $z = z_2$ plane with the magnification factor $-z_2/z_1$. The magnification is unity when $z_1 = z_2 = 2f$. The focal length f of a lens therefore completely determines its effect on paraxial rays. As indicated earlier, P_1 and P_2 are measured in coordinate systems pointing to the left and right, respectively, and the radii of curvatures R_1 and R_2 are positive for convex surfaces and negative for concave surfaces. For the biconvex lens shown in Fig. 1.2-13, R_1 is positive and R_2 is negative, so that the two terms of (1.2-12) add and provide a positive f.

EXERCISE 1.2-4

Proof of the Thin Lens Formulas. Using (1.2-8), along with the definition of the focal length given in (1.2-12), prove (1.2-11) and (1.2-13).

It is emphasized once again that the foregoing relations hold only for paraxial rays. The presence of nonparaxial rays results in aberrations, as illustrated in Fig. 1.2-15.

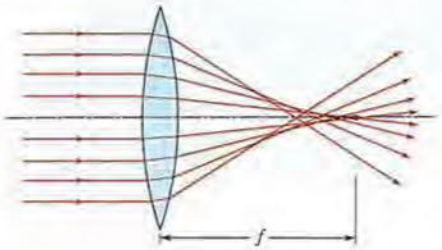


Figure 1.2-15 Nonparaxial rays do not meet at the paraxial focus. The dashed envelope of the refracted rays is called the caustic curve.

D. Light Guides

Light may be guided from one location to another by use of a set of lenses or mirrors, as illustrated schematically in Fig. 1.2-16. Since refractive elements (such as lenses) are usually partially reflective and since mirrors are partially absorptive, the cumulative loss of optical power will be significant when the number of guiding elements is large. Components in which these effects are minimized can be fabricated (e.g., antireflection-coated lenses), but the system is generally cumbersome and costly.

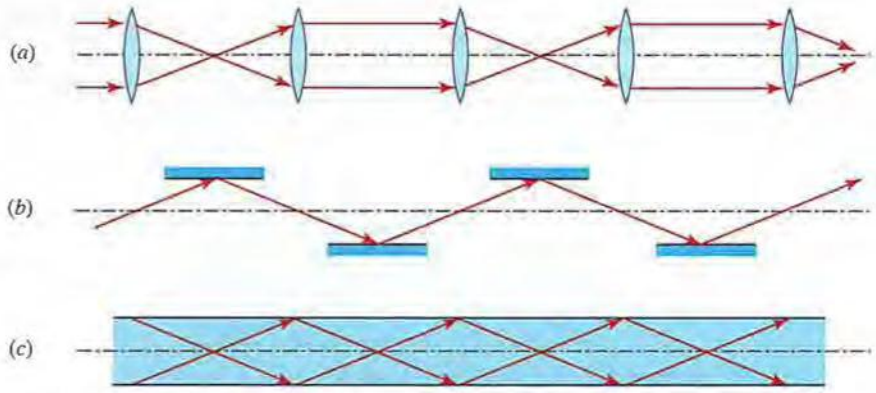


Figure 1.2-16 Guiding light: (*a*) lenses; (*b*) mirrors; (*c*) total internal reflection.

An ideal mechanism for guiding light is that of total internal reflection at the boundary between two media of different refractive indexes. Rays are reflected repeatedly without undergoing refraction. Glass fibers of high chemical purity are used to guide light for tens of kilometers with relatively low loss of optical power.

An optical fiber is a light conduit made of two concentric glass (or plastic) cylinders (Fig. 1.2-17). The inner, called the core, has a refractive index n_1, and the outer, called

the cladding, has a slightly smaller refractive index, $n_2 < n_1$. Light rays traveling in the core are totally reflected from the cladding if their angle of incidence is greater than the critical angle, $\overline{\theta} > \theta_c = \sin^{-1}(n_2/n_1)$. The rays making an angle $\theta = 90° - \overline{\theta}$ with the optical axis are therefore confined in the fiber core if $\theta < \overline{\theta_c}$, where $\overline{\theta_c} = 90° - \theta_c = \cos^{-1}(n_2/n_1)$. Optical fibers are used in optical communication systems (see Chapters 9 and 24). Some important properties of optical fibers are derived in Exercise 1.2-5.

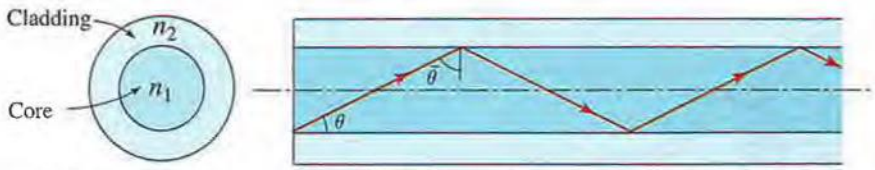


Figure 1.2-17 The optical fiber. Light rays are guided by multiple total internal reflections. Here θ represents the angle measured from the axis of the optical fiber so that its complement $\overline{\theta} = 90° - \theta$ is the angle of incidence at the dielectric interface.

EXERCISE 1.2-5

Numerical Aperture and Angle of Acceptance of an Optical Fiber. An optical fiber is illuminated by light from a source (e.g., a light-emitting diode, LED). The refractive indexes of the core and cladding of the fiber are n_1 and n_2, respectively, and the refractive index of air is 1 (Fig. 1.2-18). Show that the half-angle θ_a of the cone of rays accepted by the fiber (transmitted through the fiber without undergoing refraction at the cladding) is given by

$$\text{NA} = \sin\theta_a = \sqrt{n_1^2 - n_2^2}. \tag{1.2-15}$$

Numerical Aperture
Optical Fiber

The angle θ_a is called the acceptance angle and the parameter $\text{NA} \equiv \sin\theta_a$ is known as the **numerical aperture** of the fiber. Calculate the numerical aperture and acceptance angle for a silica-glass fiber with $n_1 = 1.475$ and $n_2 = 1.460$.

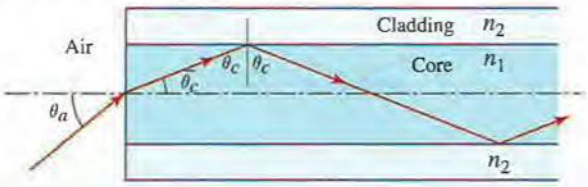


Figure 1.2-18 Acceptance angle of an optical fiber.

Trapping of Light in Media of High Refractive Index

It is often difficult for light originating inside a medium of large refractive index to be extracted into air, especially if the surfaces of the medium are parallel. This occurs since certain rays undergo multiple total internal reflections without ever refracting into air. The principle is illustrated in Exercise 1.2-6.

EXERCISE 1.2-6

Light Trapped in a Light-Emitting Diode.

(a) Assume that light is generated in all directions inside a material of refractive index n cut in the shape of a parallelepiped (Fig. 1.2-19). The material is surrounded by air with unity refractive index. This process occurs in light-emitting diodes (see Chapter 17). What is the angle of the cone of light rays (inside the material) that will emerge from each face? What happens to the other rays? What is the numerical value of this angle for GaAs ($n = 3.6$)?

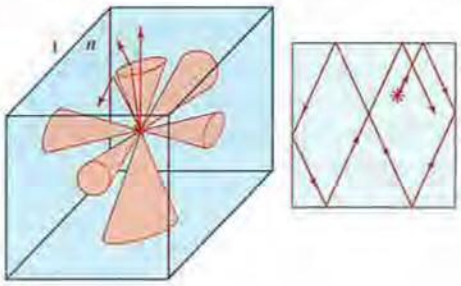


Figure 1.2-19 Trapping of light in a parallelepiped of high refractive index.

(b) Assume that when light is generated isotropically the amount of optical power associated with the rays in a given cone is proportional to the solid angle of the cone. Show that the ratio of the optical power that is extracted from the material to the total generated optical power is $3\left(1-\sqrt{1-1/n^2}\right)$, provided that $n > \sqrt{2}$. What is the numerical value of this ratio for GaAs?

1.3 GRADED-INDEX OPTICS

A graded-index (GRIN) material has a refractive index that varies with position in accordance with a continuous function $n(\mathbf{r})$. These materials are often fabricated by adding impurities (dopants) of controlled concentrations. In a GRIN medium the optical rays follow curved trajectories, instead of straight lines. By appropriate choice of $n(\mathbf{r})$, a GRIN plate can have the same effect on light rays as a conventional optical component, such as a prism or lens.

A. The Ray Equation

To determine the trajectories of light rays in an inhomogeneous medium with refractive index $n(\mathbf{r})$, we use Fermat's principle,

$$\delta \int_A^B n(\mathbf{r})\, ds = 0, \tag{1.3-1}$$

where ds is a differential length along the ray trajectory between A and B. If the trajectory is described by the function $x(s)$, $y(s)$, and $z(s)$, where s is the length of the trajectory (Fig. 1.3-1), then using the calculus of variations it can be shown† that $x(s)$,

† This derivation is beyond the scope of this book; see, e.g., R. Weinstock, *Calculus of Variations*, Dover, 1974.

$y(s)$, and $z(s)$ must satisfy three partial differential equations,

$$\frac{d}{ds}\left(n\frac{dx}{ds}\right) = \frac{\partial n}{\partial x}, \qquad \frac{d}{ds}\left(n\frac{dy}{ds}\right) = \frac{\partial n}{\partial y}, \qquad \frac{d}{ds}\left(n\frac{dz}{ds}\right) = \frac{\partial n}{\partial z}. \tag{1.3-2}$$

By defining the vector $\mathbf{r}(s)$, whose components are $x(s)$, $y(s)$, and $z(s)$, (1.3-2) may be written in the compact vector form

$$\boxed{\frac{d}{ds}\left(n\frac{d\mathbf{r}}{ds}\right) = \nabla n,} \tag{1.3-3}$$

Ray Equation

where ∇n, the gradient of n, is a vector with Cartesian components $\partial n/\partial x$, $\partial n/\partial y$, and $\partial n/\partial z$. Equation (1.3-3) is known as the **ray equation**.

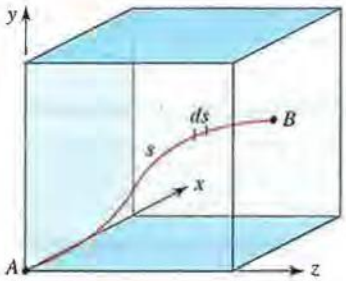


Figure 1.3-1 The ray trajectory is described parametrically by three functions $x(s)$, $y(s)$, and $z(s)$, or by two functions $x(z)$ and $y(z)$.

One approach to solving the ray equation is to describe the trajectory by two functions $x(z)$ and $y(z)$, write $ds = dz\sqrt{1 + (dx/dz)^2 + (dy/dz)^2}$, and substitute in (1.3-3) to obtain two partial differential equations for $x(z)$ and $y(z)$. The algebra is generally not trivial, but it simplifies considerably when the paraxial approximation is used.

The Paraxial Ray Equation

In the paraxial approximation, the trajectory is almost parallel to the z axis, so that $ds \approx dz$ (Fig. 1.3-2). The ray equations (1.3-2) then simplify to

$$\boxed{\frac{d}{dz}\left(n\frac{dx}{dz}\right) \approx \frac{\partial n}{\partial x}, \qquad \frac{d}{dz}\left(n\frac{dy}{dz}\right) \approx \frac{\partial n}{\partial y}.} \tag{1.3-4}$$

Paraxial Ray Equations

Given $n = n(x, y, z)$, these two partial differential equations may be solved for the trajectory $x(z)$ and $y(z)$.

In the limiting case of a homogeneous medium for which n is independent of x, y, z, (1.3-4) gives $d^2x/dz^2 = 0$ and $d^2y/dz^2 = 0$, from which it follows that x and y are linear functions of z, so that the trajectories are straight lines. More interesting cases will be examined subsequently.

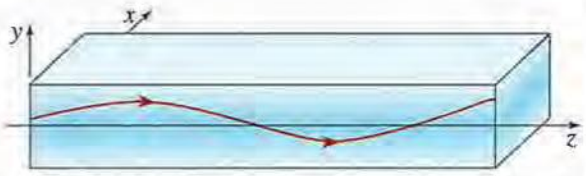


Figure 1.3-2 Trajectory of a paraxial ray in a graded-index medium.

B. Graded-Index Optical Components

Graded-Index Slab

Consider a slab of material whose refractive index $n = n(y)$ is uniform in the x and z directions but varies continuously in the y direction (Fig. 1.3-3). The trajectories of paraxial rays in the y–z plane are described by the paraxial ray equation

$$\frac{d}{dz}\left(n\frac{dy}{dz}\right) = \frac{dn}{dy}, \tag{1.3-5}$$

from which

$$\frac{d^2y}{dz^2} = \frac{1}{n(y)}\frac{dn(y)}{dy}. \tag{1.3-6}$$

Given $n(y)$ and initial conditions (y and dy/dz at $z = 0$), (1.3-6) can be solved for the function $y(z)$, which describes the ray trajectories.

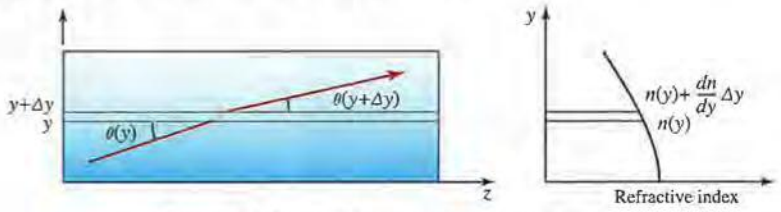


Figure 1.3-3 Refraction in a graded-index slab.

□ **Derivation of the Paraxial Ray Equation in a Graded-Index Slab Using Snell's Law.** Equation (1.3-6) may also be derived by the direct use of Snell's law (Fig. 1.3-3). Let $\theta(y) \approx dy/dz$ be the angle that the ray makes with the z axis at the position (y, z). After traveling through a layer of thickness Δy the ray changes its angle to $\theta(y + \Delta y)$. The two angles are related by Snell's law where θ, as defined in Fig. 1.3-3, is the complement of the angle of incidence (refraction):

$$\begin{aligned} n(y)\cos\theta(y) &= n(y+\Delta y)\cos\theta(y+\Delta y) \\ &= \left[n(y) + \frac{dn}{dy}\Delta y\right]\left[\cos\theta(y) - \frac{d\theta}{dy}\Delta y \sin\theta(y)\right], \end{aligned} \tag{1.3-7}$$

where we have applied the expansion $f(y+\Delta y) = f(y) + (df/dy)\,\Delta y$ to the functions $f(y) = n(y)$ and $f(y) = \cos\theta(y)$. In the limit $\Delta y \to 0$, after eliminating the term in $(\Delta y)^2$, we obtain the differential equation

$$\frac{dn}{dy} = n\frac{d\theta}{dy}\tan\theta. \tag{1.3-8}$$

For paraxial rays θ is very small so that $\tan\theta \approx \theta$. Substituting $\theta = dy/dz$ in (1.3-8), we obtain (1.3-6). ■

EXAMPLE 1.3-1. ***Slab with Parabolic Index Profile.*** An important particular distribution for the graded refractive index is

$$n^2(y) = n_0^2\left(1 - \alpha^2 y^2\right). \tag{1.3-9}$$

This is a symmetric function of y that has its maximum value at $y = 0$ (Fig. 1.3-4). A glass slab with this profile is known by the trade name SELFOC. Usually, α is chosen to be sufficiently small so that $\alpha^2 y^2 \ll 1$ for all y of interest. Under this condition, $n(y) = n_0\sqrt{1 - \alpha^2 y^2} \approx n_0(1 - \frac{1}{2}\alpha^2 y^2)$; i.e., $n(y)$ is a parabolic distribution. Also, because $n(y) - n_0 \ll n_0$, the fractional change of the refractive index is very small. Taking the derivative of (1.3-9), the right-hand side of (1.3-6) is $(1/n)dn/dy = -(n_0/n)^2\alpha^2 y \approx -\alpha^2 y$, so that (1.3-6) becomes

$$\frac{d^2y}{dz^2} \approx -\alpha^2 y. \tag{1.3-10}$$

The solutions of this equation are harmonic functions with period $2\pi/\alpha$. Assuming an initial position $y(0) = y_0$ and an initial slope $dy/dz = \theta_0$ at $z = 0$ inside the GRIN medium,

$$y(z) = y_0 \cos \alpha z + \frac{\theta_0}{\alpha} \sin \alpha z, \tag{1.3-11}$$

from which the slope of the trajectory is

$$\theta(z) = \frac{dy}{dz} = -y_0 \alpha \sin \alpha z + \theta_0 \cos \alpha z. \tag{1.3-12}$$

The ray oscillates about the center of the slab with a period (distance) $2\pi/\alpha$ known as the **pitch**, as illustrated in Fig. 1.3-4.

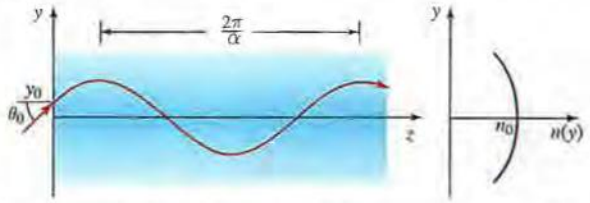


Figure 1.3-4 Trajectory of a ray in a GRIN slab of parabolic index profile (SELFOC).

The maximum excursion of the ray is $y_{\max} = \sqrt{y_0^2 + (\theta_0/\alpha)^2}$ and the maximum angle is $\theta_{\max} = \alpha y_{\max}$. The validity of this approximate analysis is ensured if $\theta_{\max} \ll 1$. If $2y_{\max}$ is smaller than the thickness of the slab, the ray remains confined and the slab serves as a light guide. Figure 1.3-5 shows the trajectories of a number of rays transmitted through a SELFOC slab. Note that all rays have the same pitch. This GRIN slab may be used as a lens, as demonstrated in Exercise 1.3-1.

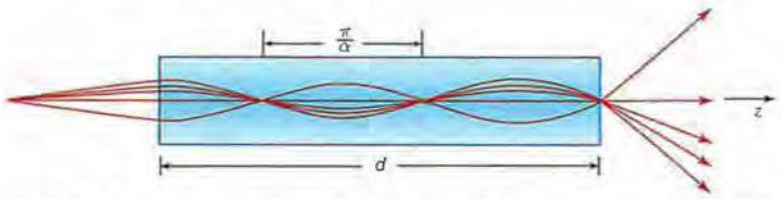


Figure 1.3-5 Trajectories of rays from an external point source in a SELFOC slab.

EXERCISE 1.3-1

The GRIN Slab as a Lens. Show that a SELFOC slab of length $d < \pi/2\alpha$ and refractive index given by (1.3-9) acts as a cylindrical lens (a lens with focusing power in the y–z plane) of focal length

$$f \approx \frac{1}{n_0 d\,\alpha \sin \alpha} \tag{1.3-13}$$

Show that the principal point (defined in Fig. 1.3-6) lies at a distance from the slab edge $\overline{AH} \approx (1/n_0\alpha)\tan(\alpha d/2)$. Sketch the ray trajectories in the special cases $d = \pi/\alpha$ and $\pi/2\alpha$.

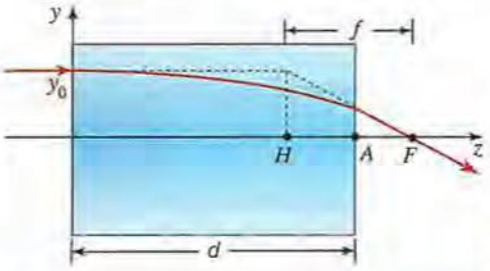


Figure 1.3-6 The SELFOC slab used as a lens; F is the focal point and H is the principal point.

Graded-Index Fibers

A graded-index fiber is a glass cylinder with a refractive index n that varies as a function of the radial distance from its axis. In the paraxial approximation, the ray trajectories are governed by the paraxial ray equations (1.3-4). Consider, for example, the distribution

$$n^2 = n_0^2\left[1 - \alpha^2\left(x^2 + y^2\right)\right]. \tag{1.3-14}$$

Substituting (1.3-14) into (1.3-4) and assuming that $\alpha^2(x^2 + y^2) \ll 1$ for all x and y of interest, we obtain

$$\frac{d^2x}{dz^2} \approx -\alpha^2 x, \qquad \frac{d^2y}{dz^2} \approx -\alpha^2 y. \tag{1.3-15}$$

Both x and y are therefore harmonic functions of z with period $2\pi/\alpha$. The initial positions (x_0, y_0) and angles ($\theta_{x0} = dx/dz$ and $\theta_{y0} = dy/dz$) at $z = 0$ determine the amplitudes and phases of these harmonic functions. Because of the circular symmetry, there is no loss of generality in choosing $x_y = 0$. The solution of (1.3-15) is then

$$\begin{aligned} x(z) &= \frac{\theta_{x0}}{\alpha}\sin \alpha z \\ y(z) &= \frac{\theta_{y0}}{\alpha}\sin \alpha z + y_0 \cos \alpha z. \end{aligned} \tag{1.3-16}$$

If $\theta_{x0} = 0$, i.e., the incident ray lies in a meridional plane (a plane passing through the axis of the cylinder, in this case the y–z plane), the ray continues to lie in that plane following a sinusoidal trajectory similar to that in the GRIN slab [Fig. 1.3-7(*a*)].

On the other hand, if $\theta_{y0} = 0$, and $\theta_{x0} = \alpha y_0$, then

$$\begin{aligned} x(z) &= y_0 \sin \alpha z \\ y(z) &= y_0 \cos \alpha z, \end{aligned} \tag{1.3-17}$$

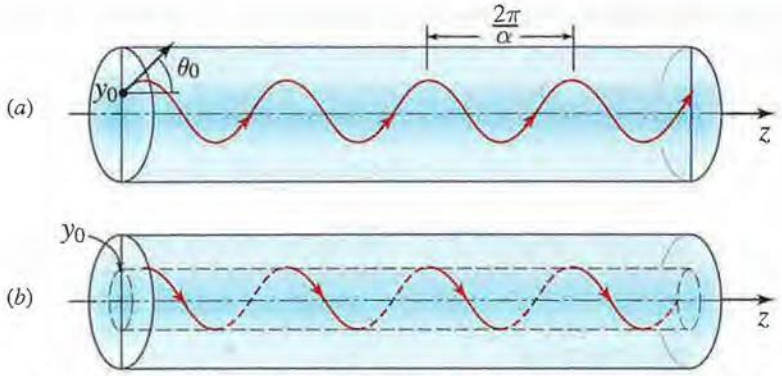


Figure 1.3-7 (*a*) Meridional and (*b*) helical rays in a graded-index fiber with parabolic index profile.

so that the ray follows a helical trajectory lying on the surface of a cylinder of radius y_0 [Fig. 1.3-7(*b*)]. In both cases the ray remains confined within the fiber, so that the fiber serves as a light guide. Other helical patterns are generated with different incident rays.

Graded-index fibers and their use in optical communications are discussed in Chapters 9 and 24.

EXERCISE 1.3-2

Numerical Aperture of the Graded-Index Fiber. Consider a graded-index fiber with the index profile provided in (1.3-14) and radius a. A ray is incident from air into the fiber at its center, which then makes an angle θ_0 with the fiber axis in the medium (see Fig. 1.3-8). Show, in the paraxial approximation, that the numerical aperture is

$$\text{NA} \equiv \sin\theta_a \approx n_0 a\alpha, \tag{1.3-18}$$

Numerical Aperture
Graded-Index Fiber

where θ_a is the maximum acceptance angle for which the ray trajectory is confined within the fiber. Compare this to the numerical aperture of a step-index fiber such as the one discussed in Exercise 1.2-5. To make the comparison fair, take the refractive indexes of the core and cladding of the step-index fiber to be $n_1 = n_0$ and $n_2 = n_0\sqrt{1-\alpha^2a^2} \approx n_0(1-\frac{1}{2}\alpha^2a^2)$, respectively.

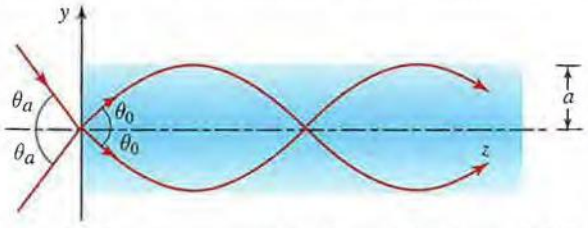


Figure 1.3-8 Acceptance angle of a graded-index optical fiber.

*C. The Eikonal Equation

The ray trajectories are often characterized by the surfaces to which they are normal. Let $S(\mathbf{r})$ be a scalar function such that its equilevel surfaces, $S(\mathbf{r}) = \text{constant}$, are everywhere normal to the rays (Fig. 1.3-9). If $S(\mathbf{r})$ is known, the ray trajectories can readily be constructed since the normal to the equilevel surfaces at a position $\mathbf{r}$ is in the direction of the gradient vector $\nabla S(\mathbf{r})$. The function $S(\mathbf{r})$, called the **eikonal**, is akin to the potential function $V(\mathbf{r})$ in electrostatics; the role of the optical rays is played by the lines of electric field $\mathbf{E} = -\nabla V$.

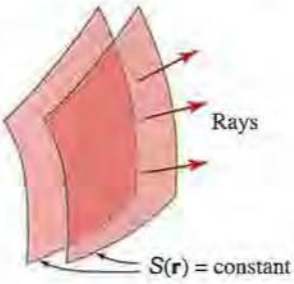


Figure 1.3-9 Ray trajectories are normal to the surfaces of constant $S(\mathbf{r})$.

To satisfy Fermat's principle (which is the main postulate of ray optics) the eikonal $S(\mathbf{r})$ must satisfy a partial differential equation known as the **eikonal equation**,

$$\left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{\partial S}{\partial y}\right)^2 + \left(\frac{\partial S}{\partial z}\right)^2 = n^2, \tag{1.3-19}$$

which is usually written in the vector form

$$\boxed{|\nabla S|^2 = n^2,} \tag{1.3-20}$$

Eikonal Equation

where $|\nabla S^2| = \nabla S \cdot \nabla S$. The proof of the eikonal equation from Fermat's principle is a mathematical exercise that lies beyond the scope of this book.† Fermat's principle (and the ray equation) can also be shown to follow from the eikonal equation. Therefore, either the eikonal equation or Fermat's principle may be regarded as the principal postulate of ray optics.

Integrating the eikonal equation (1.3-20) along a ray trajectory between points A and B gives

$$S(\mathbf{r}_B) - S(\mathbf{r}_A) = \int_A^B |\nabla S|\, ds = \int_A^B n\, ds = \text{optical pathlength between } A \text{ and } B. \tag{1.3-21}$$

This means that the difference $S(\mathbf{r}_B) - S(\mathbf{r}_A)$ represents the optical pathlength between A and B. In the electrostatics analogy, the optical pathlength plays the role of the potential difference.

To determine the ray trajectories in an inhomogeneous medium of refractive index $n(\mathbf{r})$, we can either solve the ray equation (1.3-3), as we have done earlier, or solve the eikonal equation for $S(\mathbf{r})$, from which we calculate the gradient ∇S.

† See, e.g., M. Born and E. Wolf, *Principles of Optics*, Cambridge University Press, 7th expanded and corrected ed. 2002.

If the medium is homogeneous, i.e., $n(\mathbf{r})$ is constant, the magnitude of ∇S is constant, so that the wavefront normals (rays) must be straight lines. The surfaces $S(\mathbf{r}) =$ constant may be parallel planes or concentric spheres, as illustrated in Fig. 1.3-10.

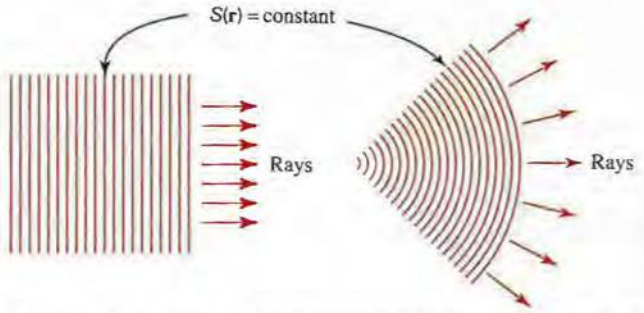


Figure 1.3-10 Rays and surfaces of constant $S(\mathbf{r})$ in a homogeneous medium.

The eikonal equation is revisited from the point-of-view of the relation between ray optics and wave optics in Sec. 2.3.

1.4 MATRIX OPTICS

Matrix optics is a technique for tracing paraxial rays. The rays are assumed to travel only within a single plane, so that the formalism is applicable to systems with planar geometry and to meridional rays in circularly symmetric systems.

A ray is described by its position and its angle with respect to the optical axis. These variables are altered as the ray travels through the system. In the paraxial approximation, the position and angle at the input and output planes of an optical system are related by two *linear* algebraic equations. As a result, the optical system is described by a 2×2 matrix called the ray transfer matrix.

The convenience of using matrix methods lies in the fact that the ray-transfer matrix of a cascade of optical components (or systems) is a product of the ray-transfer matrices of the individual components (or systems). Matrix optics therefore provides a formal mechanism for describing complex optical systems in the paraxial approximation.

A. The Ray-Transfer Matrix

Consider a circularly symmetric optical system formed by a succession of refracting and reflecting surfaces all centered about the same axis (optical axis). The z axis lies along the optical axis and points in the general direction in which the rays travel. Consider rays in a plane containing the optical axes, say the y–z plane. We proceed to trace a ray as it travels through the system, i.e., as it crosses the transverse planes at different axial distances. A ray crossing the transverse plane at z is completely characterized by the coordinate of y of its crossing point and the angle θ (Fig. 1.4-1).

An optical system is a set of optical components placed between two transverse planes at z_1 and z_2, referred to as the input and output planes, respectively. The system is characterized completely by its effect on an incoming ray of arbitrary position and

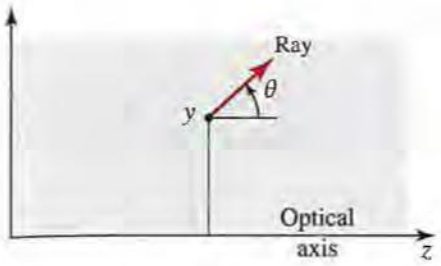


Figure 1.4-1 A ray is characterized by its coordinate y and its angle θ.

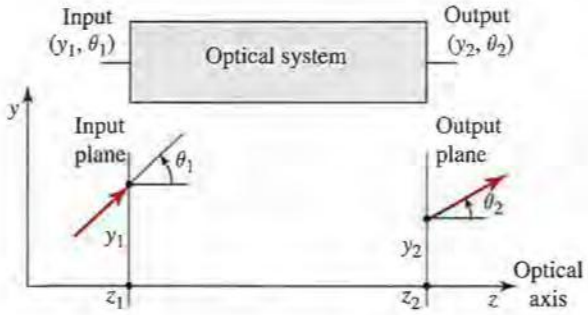


Figure 1.4-2 A ray enters an optical system at location z_1 with position y_1 and angle θ_1 and leaves at position y_2 and angle θ_2.

direction (y_1, θ_1). It steers the ray so that it has new position and direction (y_2, θ_2) at the output plane (Fig. 1.4-2).

In the paraxial approximation, when all angles are sufficiently small so that $\sin\theta \approx \theta$, the relation between (y_2, θ_2) and (y_1, θ_1) is linear and can generally be written in the form

$$y_2 = Ay_1 + B\theta_1 \tag{1.4-1}$$

$$\theta_2 = Cy_1 + D\theta_1, \tag{1.4-2}$$

where A, B, C, and D are real numbers. Equations (1.4-1) and (1.4-2) may be conveniently written in matrix form as

$$\begin{bmatrix} y_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix}. \tag{1.4-3}$$

The matrix $\mathbf{M}$, whose elements are A, B, C, and D, characterizes the optical system completely since it permits (y_2, θ_2) to be determined for any (y_1, θ_1). It is known as the **ray-transfer matrix**. As will be seen in the examples provided in Sec. 1.4B, angles that turn out to be negative point downward from the z axis in their direction of travel. Radii that turn out to be negative indicate concave surfaces whereas those that are positive indicate convex surfaces.

EXERCISE 1.4-1

Special Forms of the Ray-Transfer Matrix. Consider the following situations in which one of the four elements of the ray-transfer matrix vanishes:

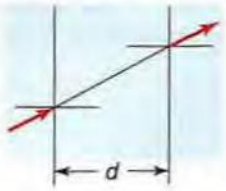
(a) Show that if $A = 0$, all rays that enter the system at the same angle leave at the same position, so that parallel rays in the input are focused to a single point at the output.

(b) What are the special features of each of the systems for which $B = 0$, $C = 0$, or $D = 0$?

B. Matrices of Simple Optical Components

Free-Space Propagation

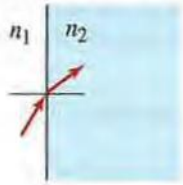
Since rays travel along straight lines in a medium of uniform refractive index such as free space, a ray traversing a distance d is altered in accordance with $y_2 = y_1 + \theta_1 d$ and $\theta_2 = \theta_1$. The ray-transfer matrix is therefore



$$\mathbf{M} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}. \qquad (1.4\text{-}4)$$

Refraction at a Planar Boundary

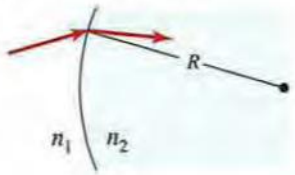
At a planar boundary between two media of refractive indexes n_1 and n_2, the ray angle changes in accordance with Snell's law $n_1 \sin\theta_1 = n_2 \sin\theta_2$. In the paraxial approximation, $n_1\theta_1 \approx n_2\theta_2$. The position of the ray is not altered, $y_2 = y_1$. The ray-transfer matrix is



$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{n_1}{n_2} \end{bmatrix}. \qquad (1.4\text{-}5)$$

Refraction at a Spherical Boundary

The relation between θ_1 and θ_2 for paraxial rays refracted at a spherical boundary between two media is provided in (1.2-8). The ray height is not altered, $y_2 \approx y_1$. The ray-transfer matrix is

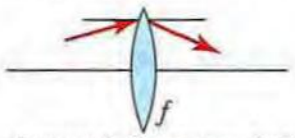


Convex: $R > 0$; concave: $R < 0$

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ -\frac{(n_2 - n_1)}{n_2 R} & \frac{n_1}{n_2} \end{bmatrix}. \qquad (1.4\text{-}6)$$

Transmission Through a Thin Lens

The relation between θ_1 and θ_2 for paraxial rays transmitted through a thin lens of focal length f is given in (1.2-11). Since the height remains unchanged ($y_2 = y_1$), we have



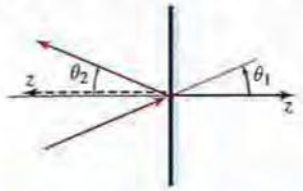
Convex: $f > 0$; concave: $f < 0$

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix}. \qquad (1.4\text{-}7)$$

Reflection from a Planar Mirror

Upon reflection from a planar mirror, the ray position is not altered, $y_2 = y_1$. Adopting the convention that the z axis points in the general direction of travel of the rays, i.e., toward the mirror for the incident rays and away from it for the reflected rays, we

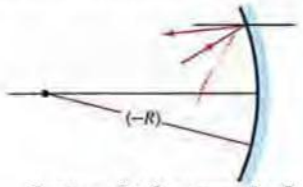
conclude that $\theta_2 = \theta_1$. The ray-transfer matrix is therefore the identity matrix



$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \qquad (1.4\text{-}8)$$

Reflection from a Spherical Mirror

Using (1.2-1), and the convention that the z axis follows the general direction of the rays as they reflect from mirrors, we similarly obtain



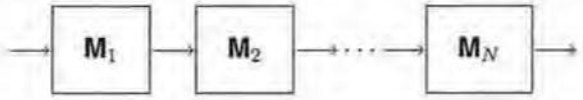
Concave: $R < 0$; convex: $R > 0$

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ \frac{2}{R} & 1 \end{bmatrix}. \qquad (1.4\text{-}9)$$

Note the similarity between the ray-transfer matrices of a spherical mirror (1.4-9) and a thin lens (1.4-7). A mirror with radius of curvature R bends rays in a manner that is identical to that of a thin lens with focal length $f = -R/2$.

C. Matrices of Cascaded Optical Components

A cascade of N optical components or systems whose ray-transfer matrices are $\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_N$ is equivalent to a single optical system of ray-transfer matrix

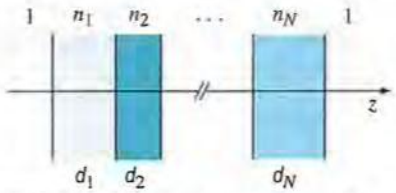


$$\mathbf{M} = \mathbf{M}_N \cdots \mathbf{M}_2 \mathbf{M}_1. \qquad (1.4\text{-}10)$$

Note the order of matrix multiplication: The matrix of the system that is crossed by the rays is first placed to the right, so that it operates on the column matrix of the incident ray first. A sequence of matrix multiplications is not, in general, commutative, although it is associative.

EXERCISE 1.4-2

A Set of Parallel Transparent Plates. Consider a set of N parallel planar transparent plates of refractive indexes $n_1, n_2, \ldots n_N$ and thicknesses $d_1, d_2, \ldots d_N$, placed in air $(n = 1)$ normal to the z axis. Using induction, show that the ray-transfer matrix is

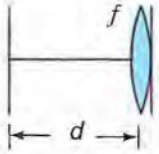


$$\mathbf{M} = \begin{bmatrix} 1 & \sum_{i=1}^{N} \frac{d_i}{n_i} \\ 0 & 1 \end{bmatrix}. \qquad (1.4\text{-}11)$$

Note that the order in which the plates are placed does not affect the overall ray-transfer matrix. What is the ray-transfer matrix of an inhomogeneous transparent plate of thickness d_0 and refractive index $n(z)$?

EXERCISE 1.4-3

A Gap Followed by a Thin Lens. Show that the ray-transfer matrix of a distance d of free space followed by a lens of focal length f is



$$\mathbf{M} = \begin{bmatrix} 1 & d \\ -\frac{1}{f} & 1 - \frac{d}{f} \end{bmatrix}. \qquad (1.4\text{-}12)$$

EXERCISE 1.4-4

Imaging with a Thin Lens. Derive an expression for the ray-transfer matrix of a system comprised of free space/thin lens/free space, as shown in Fig. 1.4-3. Show that if the imaging condition $(1/d_1 + 1/d_2 = 1/f)$ is satisfied, all rays originating from a single point in the input plane reach the output plane at the single point y_2, regardless of their angles. Also show that if $d_2 = f$, all parallel incident rays are focused by the lens onto a single point in the output plane.

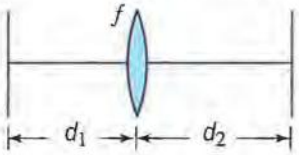


Figure 1.4-3 Single-lens imaging system.

EXERCISE 1.4-5

Imaging with a Thick Lens. Consider a glass lens of refractive index n, thickness d, and two spherical surfaces of equal radii R (Fig. 1.4-4). Determine the ray-transfer matrix of the system between the two planes at distances d_1 and d_2 from the vertices of the lens. The lens is placed in air (refractive index = 1). Show that the system is an imaging system (i.e., the input and output planes are conjugate) if

$$\frac{1}{z_1} + \frac{1}{z_2} = \frac{1}{f} \quad \text{or} \quad s_1 s_2 = f^2, \qquad (1.4\text{-}13)$$

where

$$z_1 = d_1 + h, \qquad s_1 = z_1 - f \qquad (1.4\text{-}14)$$

$$z_2 = d_2 + h, \qquad s_2 = z_2 - f \qquad (1.4\text{-}15)$$

and

$$h = \frac{(n-1)fd}{nR} \qquad (1.4\text{-}16)$$

$$\frac{1}{f} = \frac{(n-1)}{R}\left[2 - \frac{n-1}{n}\frac{d}{R}\right]. \qquad (1.4\text{-}17)$$

The points F_1 and F_2 are known as the front and back **focal points**, respectively. The points P_1 and P_2 are known as the first and second **principal points**, respectively. Show the importance of these points by tracing the trajectories of rays that are incident parallel to the optical axis.

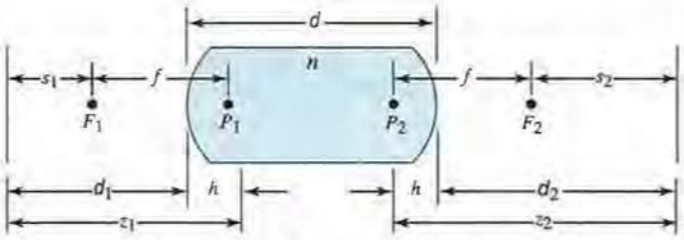


Figure 1.4-4 Imaging with a thick lens. P_1 and P_2 are the principal points and F_1 and F_2 are the focal points.

D. Periodic Optical Systems

A periodic optical system is a cascade of identical unit systems. An example is a sequence of equally spaced identical relay lenses used to guide light, as shown in Fig. 1.2-16(a). Another example is the reflection of light between two mirrors that form an optical resonator (see Sec. 10.2A); in that case, the ray repeatedly traverses the same unit system (a round trip of reflections). Even a homogeneous medium, such as a glass fiber, may be considered as a periodic system if it is divided into contiguous identical segments of equal length. We proceed to formulate a general theory of ray propagation in periodic optical systems using matrix methods.

Difference Equation for the Ray Position

A periodic system is composed of a cascade of identical unit systems (stages), each with a ray-transfer matrix (A, B, C, D), as shown in Fig. 1.4-5. A ray enters the system with initial position y_0 and slope θ_0. To determine the position and slope (y_m, θ_m) of the ray at the exit of the mth stage, we apply the $ABCD$ matrix m times,

$$\begin{bmatrix} y_m \\ \theta_m \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^m \begin{bmatrix} y_0 \\ \theta_0 \end{bmatrix}. \tag{1.4-18}$$

We can also iteratively apply the relations

$$y_{m+1} = Ay_m + B\theta_m \tag{1.4-19}$$

$$\theta_{m+1} = Cy_m + D\theta_m \tag{1.4-20}$$

to determine (y_1, θ_1) from (y_0, θ_0), then (y_2, θ_2) from (y_1, θ_1), and so on, using a software routine.

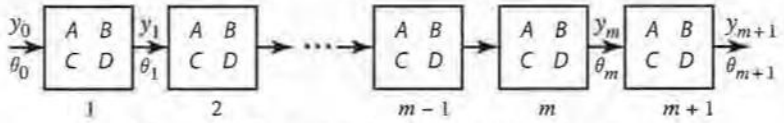


Figure 1.4-5 A cascade of identical optical systems.

It is of interest to derive equations that govern the dynamics of the position y_m, $m = 0, 1, \ldots$, irrespective of the angle θ_m. This is achieved by eliminating θ_m from

(1.4-19) and (1.4-20). From (1.4-19)

$$\theta_m = \frac{y_{m+1} - Ay_m}{B}. \tag{1.4-21}$$

Replacing m with $m + 1$ in (1.4-21) yields

$$\theta_{m+1} = \frac{y_{m+2} - Ay_{m+1}}{B}. \tag{1.4-22}$$

Substituting (1.4-21) and (1.4-22) into (1.4-20) gives

$$\boxed{y_{m+2} = 2by_{m+1} - F^2 y_m,} \tag{1.4-23}$$

Recurrence Relation for Ray Position

where

$$b = \frac{A + D}{2} \tag{1.4-24}$$

$$F^2 = AD - BC = \det[\mathbf{M}], \tag{1.4-25}$$

and $\det[\mathbf{M}]$ is the determinant of $\mathbf{M}$.

Equation (1.4-23) is a linear difference equation governing the ray position y_m. It can be solved iteratively by computing y_2 from y_0 and y_1, then y_3 from y_1 and y_2, and so on. The quantity y_1 may be computed from y_0 and θ_0 by use of (1.4-19) with $m = 0$.

It is useful, however, to derive an explicit expression for y_m by solving the difference equation (1.4-23). As with linear differential equations, a solution satisfying a linear difference equation and the initial conditions is a unique solution. It is therefore appropriate to make a judicious guess for the solution of (1.4-23). We use a trial solution of the geometric form

$$y_m = y_0 h^m, \tag{1.4-26}$$

where h is a constant. Substituting (1.4-26) into (1.4-23) immediately shows that the trial solution is suitable provided that h satisfies the quadratic algebraic equation

$$h^2 - 2bh + F^2 = 0, \tag{1.4-27}$$

from which

$$h = b \pm j\sqrt{F^2 - b^2}. \tag{1.4-28}$$

The results can be presented in a more compact form by defining the variable

$$\varphi = \cos^{-1}(b/F), \tag{1.4-29}$$

so that $b = F\cos\varphi$, $\sqrt{F^2 - b^2} = F\sin\varphi$, and therefore $h = F(\cos\varphi \pm j\sin\varphi) = F\exp(\pm j\varphi)$, whereupon (1.4-26) becomes $y_m = y_0 F^m \exp(\pm jm\varphi)$.

A general solution may be constructed from the two solutions with positive and negative signs by forming their linear combination. The sum of the two exponential functions can always be written as a harmonic (circular) function, so that

$$y_m = y_{\max} F^m \sin(m\varphi + \varphi_0), \tag{1.4-30}$$

where $y_{\max}$ and φ_0 are constants to be determined from the initial conditions y_0 and y_1. In particular, setting $m = 0$ we obtain $y_{\max} = y_0/\sin\varphi_0$.

The parameter F is related to the determinant of the ray-transfer matrix of the unit system by $F = \sqrt{\det[\mathbf{M}]}$. It can be shown that regardless of the unit system, $\det[\mathbf{M}] = n_1/n_2$, where n_1 and n_2 are the refractive indexes of the initial and final sections of the unit system. This general result is easily verified for the ray-transfer matrices of all the optical components considered in this section. Since the determinant of a product of two matrices is the product of their determinants, it follows that the relation $\det[\mathbf{M}] = n_1/n_2$ is applicable to any cascade of these optical components. For example, if $\det[\mathbf{M}_1] = n_1/n_2$ and $\det[\mathbf{M}_2] = n_2/n_3$, then $\det[\mathbf{M}_2\mathbf{M}_1] = (n_2/n_3)(n_1/n_2) = n_1/n_3$. In most applications the first and last stages are air ($n = 1$) $n_1 = n_2$, so that $\det[\mathbf{M}] = 1$ and $F = 1$, in which case the solution for the ray position is

$$y_m = y_{\max}\sin(m\varphi + \varphi_0). \tag{1.4-31}$$

Ray Position
Periodic System

We shall assume henceforth that $F = 1$. The corresponding solution for the ray angle is obtained by use of the relation $\theta_m = (y_{m+1} - Ay_m)/B$, which is derived from (1.4-19).

Condition for a Harmonic Trajectory

For y_m to be a harmonic (instead of hyperbolic) function, $\varphi = \cos^{-1} b$ must be real. This requires that

$$|b| \le 1 \quad \text{or} \quad \tfrac{1}{2}|A + D| \le 1. \tag{1.4-32}$$

Stability Condition

If, instead, $|b| > 1$, φ is then imaginary and the solution is a hyperbolic function (cosh or sinh), which increases without bound, as illustrated in Fig. 1.4-6(a). A harmonic solution ensures that y_m is bounded for all m, with a maximum value of $y_{\max}$. The bound $|b| \le 1$ therefore provides a condition of **stability** (boundedness) of the ray trajectory.

Since y_m and y_{m+1} are both harmonic functions, so too is the ray angle corresponding to (1.4-31), by virtue of (1.4-21) and trigonometric identities. Thus, $\theta_m = \theta_{\max}\sin(m\varphi + \varphi_1)$, where the constants $\theta_{\max}$ and φ_1 are determined by the initial conditions. The maximum angle $\theta_{\max}$ must be sufficiently small so that the paraxial approximation, which underlies this analysis, is applicable.

Condition for a Periodic Trajectory

The harmonic function (1.4-31) is periodic in m if it is possible to find an integer s such that $y_{m+s} = y_m$ for all m. The smallest integer is the period. The ray then retraces its path after s stages. This condition is satisfied if $s\varphi = 2\pi q$, where q is an integer. Thus, the necessary and sufficient condition for a periodic trajectory is that $\varphi/2\pi$ is a rational number q/s. If $\varphi = 6\pi/11$, for example, then $\varphi/2\pi = \frac{3}{11}$ and the trajectory is periodic with period $s = 11$ stages. This case is illustrated in Fig. 1.4-6(b). Periodic optical systems will be revisited in Chapter 7.

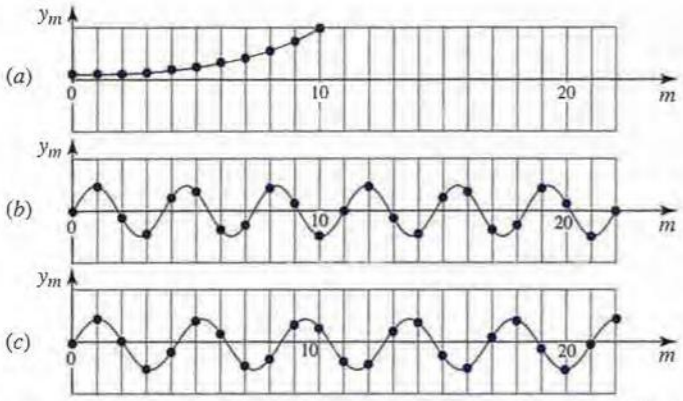


Figure 1.4-6 Examples of trajectories in periodic optical systems: (*a*) unstable trajectory ($b > 1$); (*b*) stable and periodic trajectory ($\varphi = 6\pi/11$; period = 11 stages); (*c*) stable but nonperiodic trajectory ($\varphi = 1.5$).

Summary

A paraxial ray ($\theta_{\max} \ll 1$) traveling through a cascade of identical unit optical systems, each with a ray-transfer matrix with elements (A, B, C, D) such that $AD - BC = 1$, follows a harmonic (and therefore bounded) trajectory if the condition $|\frac{1}{2}(A + D)| \leq 1$, called the stability condition, is satisfied. The position at the mth stage is then $y_m = y_{\max} \sin(m\varphi + \varphi_0)$, $m = 0, 1, 2, \ldots$, where $\varphi = \cos^{-1}[\frac{1}{2}(A + D)]$. The constants $y_{\max}$ and φ_0 are determined from the initial positions y_0 and $y_1 = Ay_0 + B\theta_0$, where θ_0 is the initial ray inclination. The ray angles are related to the positions by $\theta_m = (y_{m+1} - Ay_m)/B$ and follow a harmonic function $\theta_m = \theta_{\max} \sin(m\varphi + \varphi_1)$. The ray trajectory is periodic with period s if $\varphi/2\pi$ is a rational number q/s.

EXAMPLE 1.4-1. ***A Sequence of Equally Spaced Identical Lenses.*** A set of identical lenses of focal length f separated by distance d, as shown in Fig. 1.4-7, may be used to relay light between two locations. The unit system, a distance of d of free space followed by a lens, has a ray-transfer matrix given by (1.4-12); $A = 1$, $B = d$, $C = -1/f$, $D = 1 - d/f$. The parameter $b = \frac{1}{2}(A + D) = 1 - d/2f$ and the determinant is unity. The condition for a stable ray trajectory, $|b| \leq 1$ or $-1 \leq b \leq 1$, is therefore

$$0 \leq d \leq 4f, \tag{1.4-33}$$

so that the spacing between the lenses must be smaller than four times the focal length. Under this condition the positions of paraxial rays obey the harmonic function

$$y_m = y_{\max} \sin(m\varphi + \varphi_0), \qquad \varphi = \cos^{-1}\left(1 - \frac{d}{2f}\right). \tag{1.4-34}$$

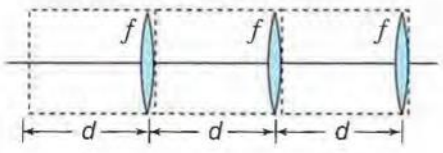


Figure 1.4-7 A periodic sequence of lenses.

When $d = 2f$, $\varphi = \pi/2$, and $\varphi/2\pi = \frac{1}{4}$, so that the trajectory of an arbitrary ray is periodic with period equal to four stages. When $d = f$, $\varphi = \pi/3$, and $\varphi/2\pi = \frac{1}{6}$, so that the ray trajectory is periodic and retraces itself each six stages. These cases are illustrated in Fig. 1.4-8.

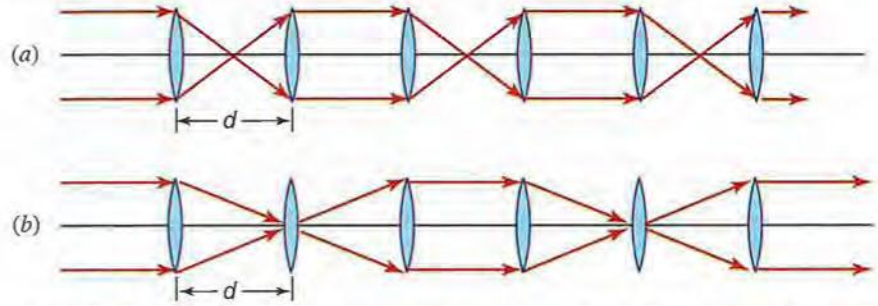


Figure 1.4-8 Examples of stable ray trajectories in a periodic lens system: (*a*) $d = 2f$; (*b*) $d = f$.

EXERCISE 1.4-6

A Periodic Set of Pairs of Different Lenses. Examine the trajectories of paraxial rays through a periodic system comprising a sequence of lens pairs with alternating focal lengths f_1 and f_2, as shown in Fig. 1.4-9. Show that the ray trajectory is bounded (stable) if

$$0 \le \left(1 - \frac{d}{2f_1}\right)\left(1 - \frac{d}{2f_2}\right) \le 1. \tag{1.4-35}$$

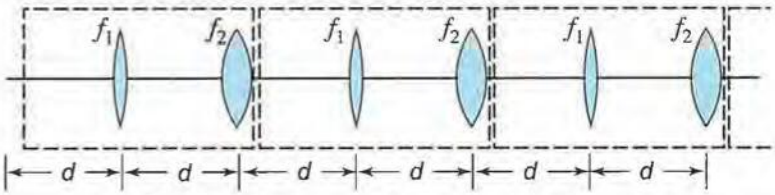


Figure 1.4-9 A periodic sequence of lens pairs.

EXERCISE 1.4-7

An Optical Resonator. Paraxial rays are reflected repeatedly between two spherical mirrors of radii R_1 and R_2 separated by a distance d (Fig. 1.4-10). Regarding this as a periodic system whose unit system is a single round trip between the mirrors, determine the condition of stability for the ray trajectory. Optical resonators will be studied in detail in Chapter 10.

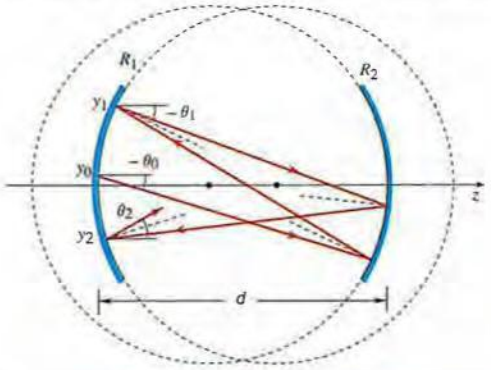


Figure 1.4-10 The optical resonator as a periodic optical system.

READING LIST

General

F. L. Pedrotti, L. M. Pedrotti, and L. S. Pedrotti, *Introduction to Optics*, Prentice Hall, 3rd ed. 2006.

K. K. Sharma, *Optics: Principles and Applications*, Academic Press, 2006.

A. Walther, *The Ray and Wave Theory of Lenses*, Cambridge University Press, 1995, paperback ed. 2006.

K. D. Moeller, *Optics: Learning by Computing with Examples Using Maple, MathCad, Mathematica, and MATLAB*, Springer-Verlag, 2nd ed. 2006.

T.-C. Poon and T. Kim, *Engineering Optics with MATLAB*, World Scientific, 2006.

A. Siciliano, *Optics: Problems and Solutions*, World Scientific, 2006.

G. Chartier, *Introduction to Optics*, Springer-Verlag, 2005.

J. Strong, *Concepts of Classical Optics*, Freeman, 1958; Dover, paperback ed. 2004.

G. Brooker, *Modern Classical Optics*, Oxford University Press, 2003.

M. Born and E. Wolf, *Principles of Optics*, Cambridge University Press, 7th expanded and corrected ed. 2002.

E. Hecht, *Optics*, Addison–Wesley, 4th ed. 2002.

M. Mansuripur, *Classical Optics and Its Applications*, Cambridge University Press, 2002.

M. P. Keating *Geometric, Physical, and Visual Optics*, Butterworth–Heinemann, 2nd ed. 2002.

M. Young, *Optics and Lasers Including Fibers and Optical Waveguides*, Springer-Verlag, 1977, 5th ed. 2000.

J. R. Meyer-Arendt, *Introduction to Classical and Modern Optics*, Prentice Hall, 1972, 4th ed. 1995.

J. W. Blaker and W. M. Rosenblum, *Optics—An Introduction for Students of Engineering*, Macmillan, 1993.

D. T. Moore, ed., *Selected Papers on Gradient-Index Optics*, SPIE Optical Engineering Press (Milestone Series Volume 67), 1993.

F. A. Jenkins and H. E. White, *Fundamentals of Optics*, McGraw–Hill, 1937, 4th revised ed. 1991.

P. P. Banerjee and T.-C. Poon, *Principles of Applied Optics*, Aksen Associates, 1991.

R. D. Guenther, *Modern Optics*, Wiley, 1990.

E. Hecht and A. Zajac, *Optics*, Addison–Wesley, 1974, 2nd ed. 1990.

W. T. Welford, *Optics*, Oxford University Press, 1976, 3rd ed. 1988.

R. W. Wood, *Physical Optics*, Macmillan, 3rd ed. 1934; Optical Society of America, 1988.

M. V. Klein and T. E. Furtak, *Optics*, Wiley, 1982, 2nd ed. 1986.

E. W. Marchand, *Gradient-Index Optics*, Academic Press, 1978.

F. P. Carlson, *Introduction to Applied Optics for Engineers*, Academic Press, 1977.

R. W. Ditchburn, *Light*, Academic Press, 3rd ed. 1976.

E. Hecht, *Schaum's Outline of Optics*, McGraw–Hill, paperback ed. 1974.

B. B. Rossi, *Optics*, Addison–Wesley, 1957, reprinted 1965.

J. M. Stone, *Radiation and Optics*, McGraw–Hill, 1963.

A. Sommerfeld, *Lectures on Theoretical Physics: Optics*, Academic Press, paperback ed. 1954.

Geometrical Optics

Yu. A. Kravtsov, *Geometrical Optics in Engineering Physics*, Alpha Science, 2005.

J. E. Greivenkamp, *Field Guide to Geometrical Optics*, SPIE Optical Engineering Press, 2004.

K. B. Wolf, *Geometric Optics on Phase Space*, Springer-Verlag, 2004.

M. Katz, *Introduction to Geometrical Optics*, World Scientific, 2002.

R. Ditteon, *Modern Geometrical Optics*, Wiley, 1998.

F. Colombini and N. Lerner, eds., *Geometrical Optics and Related Topics*, Birkhäuser, 1997.

P. Mouroulis and J. Macdonald, *Geometrical Optics and Optical Design*, Oxford University Press, 1997.

D. S. Loshin, *The Geometrical Optics Workbook*, Butterworth–Heinemann, 1991.

G. A. Fry, *Geometrical Optics*, Chilton, 1969, reprinted 1981.

W. T. Welford and R. Winston, *The Optics of Nonimaging Concentrators*, Academic Press, 1978.

O. N. Stavroudis, *The Optics of Rays, Wavefronts, and Caustics*, Academic Press, 1972.

H.-G. Zimmer, *Geometrical Optics*, Springer-Verlag, 1970.

A. Nussbaum, *Geometric Optics: An Introduction*, Addison–Wesley, 1968.

R. K. Luneburg and M. Herzberger, *Mathematical Theory of Optics*, University of California Press, 1964, reprinted 1966.

Optical System Design

H. Gross, ed., *Handbook of Optical Systems*, Wiley, 2005.

D. Malacara and Z. Malacara, *Handbook of Optical Design*, Marcel Dekker, 1994, 2nd ed. 2004.

R. E. Fischer and B. Tadic-Galeb, *Optical System Design*, McGraw–Hill, 2000.

W. J. Smith, *Modern Optical Engineering: The Design of Optical Systems*, McGraw–Hill, 1966, 3rd ed. 2000.

A. Nussbaum, *Optical System Design*, Prentice Hall, 1998.

D. C. O'Shea, *Elements of Modern Optical Design*, Wiley, 1985.

R. Kingslake, *Optical System Design*, Academic Press, 1983.

L. Levi, *Applied Optics: A Guide to Optical System Design*, Wiley, Volume 1, 1968; Volume 2, 1980.

Matrix Optics

A. Gerrard and J. M. Burch, *Introduction to Matrix Methods in Optics*, Wiley, 1975; Dover, paperback ed. 1994.

J. W. Blaker, *Geometric Optics: The Matrix Theory*, Marcel Dekker, 1971.

W. Brouwer, *Matrix Methods in Optical Instrument Design*, Benjamin, 1964.

Popular and Historical

R. J. Weiss, *A Brief History of Light and Those that Lit the Way*, World Scientific, 1996.

A. R. Hall, *All was Light: An Introduction to Newton's Opticks*, Clarendon Press/Oxford University Press, 1993.

R. Kingslake, *A History of the Photographic Lens*, Academic Press, 1989.

M. I. Sobel, *Light*, University of Chicago Press, 1987.

A. I. Sabra, *Theories of Light from Descartes to Newton*, Cambridge University Press, 1981.

I. Newton, *Opticks or a Treatise of the Reflections, Refractions, Inflections & Colours of Light*, 4th ed. 1704; Dover, reissued 1979.

A. C. S. van Heel and C. H. F. Velze, *What is Light?*, McGraw–Hill, 1968, reprinted 1978.

V. Ronchi, *The Nature of Light: An Historical Survey*, Harvard University Press, 1970.

S. Tolansky, *Revolution in Optics*, Penguin, 1968.

S. Tolansky, *Curiosities of Light Rays and Light Waves*, Elsevier, 1965.

W. H. Bragg, *Universe of Light*, Dover, paperback ed. 1959.

E. Rüchardt, *Light, Visible and Invisible*, University of Michigan Press, 1958.

PROBLEMS

1.1-2 **Fermat's Principle with Maximum Time.** Consider the elliptical mirror shown in Fig. P1.1-2(*a*), whose foci are denoted A and B. Geometrical properties of the ellipse dictate that the pathlength $\overline{APB}$ is identical to the pathlengths $\overline{AP'B}$ and $\overline{AP''B}$ for adjacent points on the ellipse.

(a) Now consider another mirror with a radius of curvature smaller than that of the elliptical mirror, but tangent to it at P, as displayed in Fig. P1.1-2(*b*). Show that the path $\overline{APB}$ followed by the light ray in traveling between points A and B is a path of *maximum* time, i.e., is greater than the adjacent paths $\overline{AQ'B}$ and $\overline{AQ''B}$.

(b) Finally, consider a mirror that crosses the ellipse, but is tangent to it at P, as illustrated in Fig. P1.1-2(*c*). Show that the possible ray paths $\overline{AQ'B}$, $\overline{APB}$, and $\overline{AQ''B}$ exhibit a point of inflection.

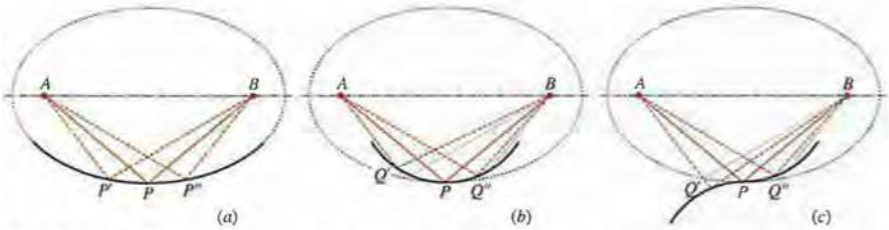


Figure P1.1-2 (*a*) Reflection from an elliptical mirror. (*b*) Reflection from an inscribed tangential mirror with greater curvature. (*c*) Reflection from a tangential mirror with curvature changing from concave to convex.

1.2-7 **Transmission through Planar Plates.**

(a) Use Snell's law to show that a ray entering a planar plate of thickness d and refractive index n_1 (placed in air; $n \approx 1$) emerges parallel to its initial direction. The ray need not be paraxial. Derive an expression for the lateral displacement of the ray as a function of the angle of incidence θ. Explain your results in terms of Fermat's principle.

(b) If the plate instead comprises a stack of N parallel layers stacked against each other with thicknesses $d_1, d_2, \ldots, d_N$ and refractive indexes $n_1, n_2, \ldots, n_N$, show that the transmitted ray is parallel to the incident ray. If θ_m is the angle of the ray in the mth layer, show that $n_m \sin\theta_m = \sin\theta$, $m = 1, 2, \ldots$.

1.2-8 **Lens in Water.** Determine the focal length f of a biconvex lens with radii 20 cm and 30 cm and refractive index $n = 1.5$. What is the focal length when the lens is immersed in water $(n = \frac{4}{3})$?

1.2-9 **Numerical Aperture of a Cladless Fiber.** Determine the numerical aperture and the acceptance angle of an optical fiber if the refractive index of the core is $n_1 = 1.46$ and the cladding is stripped out (replaced with air $n_2 \approx 1$).

1.2-10 **Fiber Coupling Spheres.** Tiny glass balls are often used as lenses to couple light into and out of optical fibers. The fiber end is located at a distance f from the sphere. For a sphere of radius $a = 1$ mm and refractive index $n = 1.8$, determine f such that a ray parallel to the optical axis at a distance $y = 0.7$ mm is focused onto the fiber, as illustrated in Fig. P1.2-10.

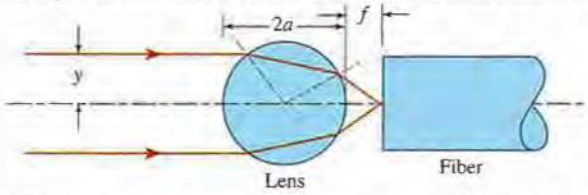


Figure P1.2-10 Focusing light into an optical fiber with a spherical glass ball.

1.2-11 **Extraction of Light from a High-Refractive-Index Medium.** Assume that light is generated isotropically in all directions inside a material of refractive index $n = 3.7$ cut in the shape of a parallelepiped and placed in air $(n = 1)$ (see Exercise 1.2-6).

(a) If a reflective material acting as a perfect mirror is coated on all sides except the front side, determine the percentage of light that may be extracted from the front side.

(b) If another transparent material of refractive index $n = 1.4$ is placed on the front side, would that help extract some of the trapped light?

1.3-3 **Axially Graded Plate.** A plate of thickness d is oriented normal to the z axis. The refractive index $n(z)$ is graded in the z direction. Show that a ray entering the plate from air at an incidence angle θ_0 in the y–z plane makes an angle $\theta(z)$ at position z in the medium given by $n(z)\sin\theta(z) = \sin\theta_0$. Show that the ray emerges into air parallel to the original incident ray. *Hint:* You may use the results of Prob. 1.2-7. Show that the ray position $y(z)$ inside the plate obeys the differential equation $(dy/dz)^2 = (n^2/\sin^2\theta - 1)^{-1}$.

1.3-4 **Ray Trajectories in GRIN Fibers.** Consider a graded-index optical fiber with cylindrical symmetry about the z axis and refractive index $n(\rho)$, $\rho = \sqrt{x^2 + y^2}$. Let (ρ, ϕ, z) be the position vector in a cylindrical coordinate system. Rewrite the paraxial ray equations, (1.3-4), in a cylindrical system and derive differential equations for ρ and ϕ as functions of z.

1.4-8 **Ray-Transfer Matrix of a Lens System.** Determine the ray-transfer matrix for an optical system made of a thin convex lens of focal length f and a thin concave lens of focal length $-f$ separated by a distance f. Discuss the imaging properties of this composite lens.

1.4-9 **Ray-Transfer Matrix of a GRIN Plate.** Determine the ray-transfer matrix of a SELFOC plate [i.e., a graded-index material with parabolic refractive index $n(y) \approx n_0(1 - \frac{1}{2}\alpha^2 y^2)$] of thickness d.

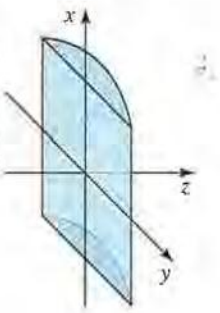
1.4-10 **The GRIN Plate as a Periodic System.** Consider the trajectories of paraxial rays inside a SELFOC plate normal to the z axis. This system may be regarded as a periodic system made of a sequence of identical contiguous plates, each of thickness d. Using the result of Prob. 1.4-9, determine the stability condition of the ray trajectory. Is this condition dependent on the choice of d?

1.4-11 **Recurrence Relation for a Planar-Mirror Resonator.** Consider a planar-mirror optical resonator, with mirror separation d, as a periodic optical system. Determine the unit ray-transfer matrix for this system, demonstrating that $b = 1$ and $F = 1$. Show that there is then only a single root to the quadratic equation (1.4-27) so that the ray position must then take the form $\alpha + m\beta$, where α and β are constants.

1.4-12 **4×4 Ray-Transfer Matrix for Skewed Rays.** Matrix methods may be generalized to describe skewed paraxial rays in circularly symmetric systems, and to astigmatic (non-circularly symmetric) systems. A ray crossing the plane $z = 0$ is generally characterized by four variables — the coordinates (x, y) of its position in the plane, and the angles (θ_x, θ_y) that its projections in the x–z and y–z planes make with the z axis. The emerging ray is also characterized by four variables linearly related to the initial four variables. The optical system may then be characterized completely, within the paraxial approximation, by a 4×4 matrix.

(a) Determine the 4×4 ray-transfer matrix of a distance d in free space.

(b) Determine the 4×4 ray-transfer matrix of a thin cylindrical lens with focal length f oriented in the y direction. The cylindrical lens has focal length f for rays in the y–z plane, and no focusing power for rays in the x–z plane.



CHAPTER

24

OPTICAL FIBER COMMUNICATIONS

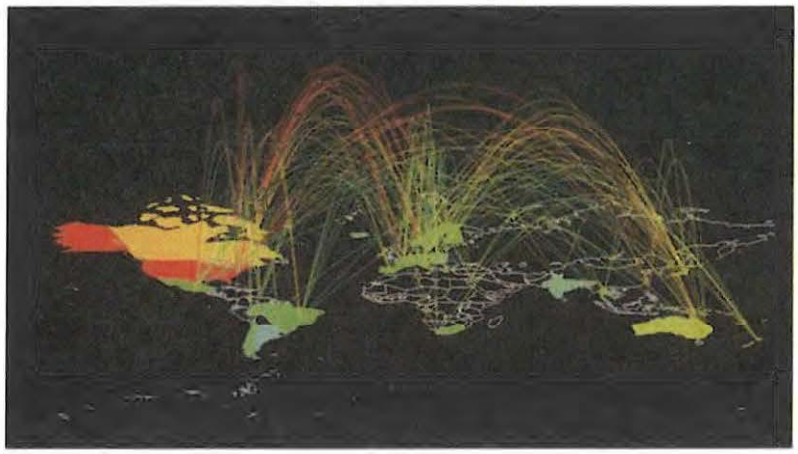
24.1 FIBER-OPTIC COMPONENTS 1074
- A. Optical Fibers
- B. Sources for Optical Transmitters
- C. Optical Amplifiers
- D. Detectors for Optical Receivers

24.2 OPTICAL FIBER COMMUNICATION SYSTEMS 1084
- A. Evolution of Optical Fiber Communication Systems
- B. Performance of Optical Fiber Systems
- C. Attenuation- and Dispersion-Limited Systems
- D. Attenuation and Dispersion Compensation and Management
- E. Soliton Optical Communications

24.3 MODULATION AND MULTIPLEXING 1101
- A. Modulation
- B. Multiplexing
- C. Wavelength-Division Multiplexing (WDM)

24.4 FIBER-OPTIC NETWORKS 1106
- A. Network Topologies and Multiple Access
- B. Wavelength-Division Multiplexing (WDM) Networks

24.5 COHERENT OPTICAL COMMUNICATIONS 1112



Intercontinental optical fiber communications network.

Until the mid-1970s virtually all communication systems relied on the transmission of information over electrical cables or have made use of radio-frequency and microwave electromagnetic radiation propagating in free space. It would appear that the use of light would have been a more natural choice for communications since, unlike electricity and radio waves, it did not have to be discovered. The reasons for the delay in the development of this technology are twofold: the difficulty of producing a light source that could be rapidly switched on and off and therefore could encode information at a high rate, and the fact that light is easily obstructed by opaque objects such as clouds, fog, smoke, and haze. Unlike radio-frequency and microwave radiation, light is rarely suitable for free-space communication.

Lightwave communication has come into its own, however, and indeed it is now the preferred technology in many applications including the transmission of data, voice, video, and telemetry in short-distance communication and local-area networks, as well as long-haul communication and internet traffic. Lightwave technology affords the user enormous transmission capacity, distant spacings of repeaters, immunity from electromagnetic interference, and relative ease of installation. Lightwave transmission is the only technology capable of meeting the vast and exponentially increasing demands of global communication, and its is now reaching individual dwellings via fiber to the home (FTTH) broadband systems.

The spectacular successes of lightwave communication have their roots in two critical photonic inventions: the development of the light-emitting diode (LED) and the development of the low-loss optical fiber as a light conduit. Suitable detectors of light have been available for some time, although their performance has been improved dramatically in recent years. Interest in optical communications was initially stirred by the invention of the laser in the early 1960s. However, the first generation of optical fiber communication systems made use of LED sources and indeed many present local-area communication systems continue to do so. Nevertheless, most lightwave communication systems (such as long-haul single-mode optical fiber systems and short-haul free-space systems) do benefit from the large optical power, narrow linewidth, and high directivity provided by the laser.

This Chapter

This chapter is an introduction to optical fiber communication systems and networks. A point-to-point communication link comprises three basic elements, as illustrated in Fig. 24.0-1: a compact light source modulated by the electrical signal, a low-loss/low-dispersion optical fiber, and a photodetector converting the optical signal back into an electrical signal. These optical components have been discussed in detail in Chapters 17, 9, and 18, respectively. Optical amplifiers have also proved themselves to be very valuable in fiber systems; these devices are discussed in Chapter 14. To make this chapter self-contained, Sec. 24.1 provides an abbreviated summary of the pertinent properties of fibers, sources, detectors, and amplifiers, examining their role in the context of the overall design, operation, and performance of an optical communication link. Other optical accessories such as splices, connectors, couplers, switches, and multiplexing devices are also essential to the successful operation of fiber links and networks; the principles of some of these devices are described in Chapter 23 and in other parts of this book.

Section 24.2 introduces the basic design principles applicable to long-distance digital and analog optical fiber communication systems using intensity modulation. The maximum fiber length that can be used to transmit data, at a given rate and with a prescribed level of performance, is determined. Performance deteriorates if the data rate exceeds the fiber bandwidth, or if the received power is smaller than the receiver



Figure 24.0-1 An optical fiber communication system. An electrical signal is converted into an optical signal (E/O) by modulating an optical source. The optical signal is transmitted through the fiber to the receiver. At the receiver, the optical signal is converted back into an electrical signal by use of a detector and demodulator (O/E). For long fibers, optical amplifiers (OA) may be used to boost the weakened optical signal. Alternatively, several optical links may be cascaded to form a longer link by use of an intermediate process of electrical amplification and signal regeneration between adjacent links. Such units are called **regenerators** or **repeaters**.

sensitivity (so that the signal cannot be distinguished from noise). This is followed, in Sec. 24.3, by an introduction to modulation and multiplexing systems used in optical fiber communications.

Fiber-optic networks are communication links connecting multiple users that are distributed in some geographic area, and controlled by a set of routers and switches. Section 24.4 provides an introduction to such networks, including **wavelength-division multiplexing (WDM)** networks.

Coherent optical communication systems, which are introduced in Sec. 24.5, use light not as a source of controllable power but rather as an electromagnetic wave of controllable amplitude, phase, or frequency. These systems are the natural extension to higher frequencies of conventional radio and microwave communications. They provide substantial gains in receiver sensitivity, permitting greater spacings between repeaters and increased data rates.

24.1 FIBER-OPTIC COMPONENTS

A. Optical Fibers

An optical fiber is a cylindrical dielectric waveguide made of low-loss materials, usually fused silica glass of high chemical purity. In the simplest optical fiber, the step-index fiber, the core of the waveguide has a constant refractive index slightly higher than that of the cladding (the outer medium) so that light is guided along the fiber axis by total internal reflection.

As described in Chapter 9, the transmission of light through the fiber may be studied by examining the trajectories of rays within the core. In accordance with a more complete analysis based on electromagnetic theory, light travels in the fiber in the form of modes; each is a wave with a distinct spatial distribution, polarization, propagation constant, group velocity, and attenuation coefficient. There is, however, a correspondence between each mode and a ray that bounces within the core in a distinct trajectory.

The step-index fiber is characterized by its core radius a, the refractive indexes of the core and cladding, n_1 and n_2, and the fractional refractive index change $\Delta = (n_1 - n_2)/n_1$, which is usually very small ($\Delta = 0.001$–0.02). Light rays making angles with the fiber axis smaller than the complement of the critical angle, $\overline{\theta}_c = \cos^{-1}(n_2/n_1)$, are guided within the core by multiple total internal reflections at the core–cladding boundary. The angle $\overline{\theta}_c$ in the fiber corresponds to an acceptance angle

$\theta_a = \sin^{-1}(\text{NA})$ for rays incident from air into the fiber, where

$$\text{NA} = \sin\theta_a = \sqrt{n_1^2 - n_2^2} \approx n_1\sqrt{2\Delta} \tag{24.1-1}$$

Numerical Aperture

is the numerical aperture.

Multimode Fibers (MMF)

The number of guided modes M is governed by the V parameter, $V = 2\pi(a/\lambda_o)\,\text{NA}$, where a/λ_o is the ratio of the core radius a to the wavelength λ_o. In a fiber with $V \gg 1$, there are a large number of modes, $M \approx V^2/2$. Since the modes travel with different group velocities, this results in pulse spreading, which increases linearly with the fiber length, an effect called **modal dispersion**. When an impulse of light travels a distance L in the fiber, it arrives as a sequence of pulses centered at the mode delay times, as illustrated in Fig. 24.1-1. The composite pulse has an approximate RMS width

$$\sigma_{\text{fiber}} \approx \frac{\Delta}{2c_1} L, \tag{24.1-2}$$

Response Time (Step-Index MMF)

where $c_1 = c_o/n_1$. It is therefore more desirable to use fibers with small Δ. For example, if $n_1 = 1.46$ and $\Delta = 0.01$, the response time per km $\approx \Delta/2c_1 \approx 24$ ns/km. For a 100-km fiber, an impulse spreads to a width of 2.4 μs.

Modal dispersion can also be reduced by use of **graded-index (GRIN)** fibers. In such fibers the refractive index of the core varies gradually from a maximum value n_1 on the fiber axis to a minimum value n_2 at the core–cladding boundary. Rays follow curved trajectories, with paths shorter than those in the step-index fiber. The axial ray travels the shortest distance at the smallest phase velocity (largest refractive index), whereas oblique rays travel longer distances at higher phase velocities (smaller refractive indexes), so that the delay times are approximately equalized. If the fiber is graded optimally (using an approximately parabolic profile), then the pulse spreading rate (ps/km) is equal to that of the equivalent step-index fiber multiplied by a factor of $\Delta/2$. For example, for $\Delta = 0.01$, the pulse spread is reduced by a factor of 500. This factor, however, is usually not fully met in practical graded-index fibers because of the difficulty of achieving ideal index profiles.

Single-Mode Fibers (SMF)

When the core radius a and the numerical aperture NA of a step-index fiber are sufficiently small so that $V < 2.405$, only a single mode is allowed and the fiber is called a **single-mode fiber (SMF)**. One advantage of using an SMF is the elimination of pulse spreading caused by modal dispersion. Pulse spreading occurs, nevertheless, since the initial pulse has a finite spectral linewidth and the group velocities (and therefore the delay times) are wavelength dependent. This effect is called **chromatic dispersion**.

There are two origins of chromatic dispersion: **material dispersion**, which results from the dependence of the refractive index on the wavelength, and **waveguide dispersion**, which is a consequence of the dependence of the group velocity of the mode on the ratio between the core radius and the wavelength. Material dispersion is usually larger than waveguide dispersion.

A short optical pulse of spectral width σ_λ spreads to a temporal width

$$\sigma_{\text{fiber}} = |D|\sigma_\lambda L, \qquad (24.1\text{-}3)$$

Response Time (SMF)

proportional to the propagation distance L (km) and to the source linewidth σ_λ (nm), where D is the dispersion coefficient (ps/km-nm). The parameter D involves a combination of material and waveguide dispersion. For weakly guiding fibers ($\Delta \ll 1$), D may be separated into a sum $D_\lambda + D_w$ of the material and waveguide contributions.

As an example, an SMF with a light source of spectral linewidth $\sigma_\lambda = 1$ nm (from a typical single-mode laser) and a fiber dispersion coefficient $D = 1$ ps/km-nm (for operation near $\lambda_o = 1300$ nm with minimal waveguide dispersion), the response time given by (24.1-3) is $\sigma_\tau/L = 1$ ps/km. A fiber of length 100 km has a 100-ps response time.

The geometries, refractive-index profiles, and pulse broadening in multimode step-index and graded-index fibers and in single-mode fibers are schematically compared in Fig. 24.1-1.

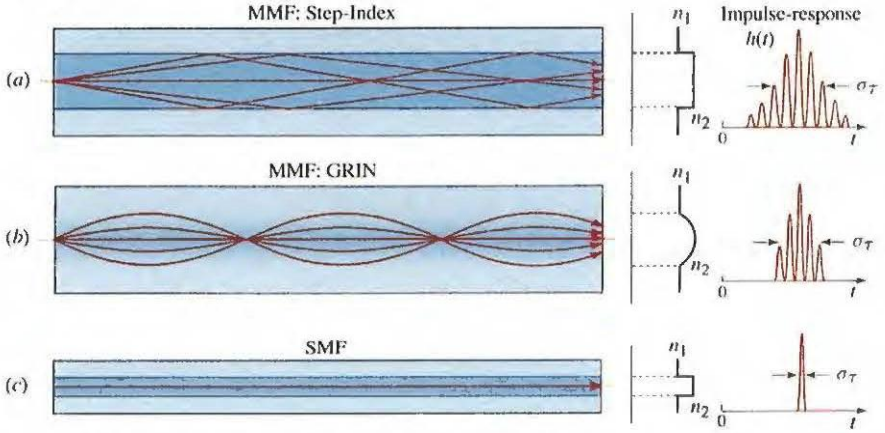


Figure 24.1-1 (*a*) Step-index multimode fibers (MMF): relatively large core diameter; uniform refractive indexes in the core and cladding; large pulse spreading due to modal dispersion. (*b*) Graded-index (GRIN) MMF: refractive index of the core is graded; there are fewer modes; pulse broadening due to modal dispersion is reduced. (*c*) Single-mode fibers (SMF): small core diameter; no modal dispersion; pulse broadening is due only to material and waveguide dispersion.

Material Attenuation and Dispersion

The wavelength dependence of the attenuation coefficients of fused-silica-glass fibers is illustrated in Fig. 24.1-2. As the wavelength increases beyond the visible band, the attenuation drops to a minimum of approximately 0.3 dB/km at $\lambda_o = 1300$ nm, increases slightly at 1.4 μm because of OH-ion absorption, and then drops again to its absolute minimum of ≈ 0.16 dB/km at $\lambda_o = 1550$ nm, beyond which it rises sharply. Fibers with suppressed OH absorption have been recently developed.

The wavelength dependence of the dispersion coefficient D_λ of fused silica glass is also illustrated in Fig. 24.1-2. It changes from negative values at short wavelengths to positive values at long wavelengths, and is zero at $\lambda_o \approx 1312$ nm. In a medium with negative dispersion, shorter-wavelength components of a pulse are slower than longer-wavelength components. This is known as **normal dispersion**. The opposite (called

anomalous dispersion) occurs in a medium exhibiting positive dispersion coefficient (see Sec. 5.6). Although the sign of the dispersion coefficient does not affect the pulse-broadening rate, the sign can play an important role in pulse propagation through media consisting of cascades of materials with different dispersion sign, as described in Sec. 24.2D (see also Sec. 22.3).

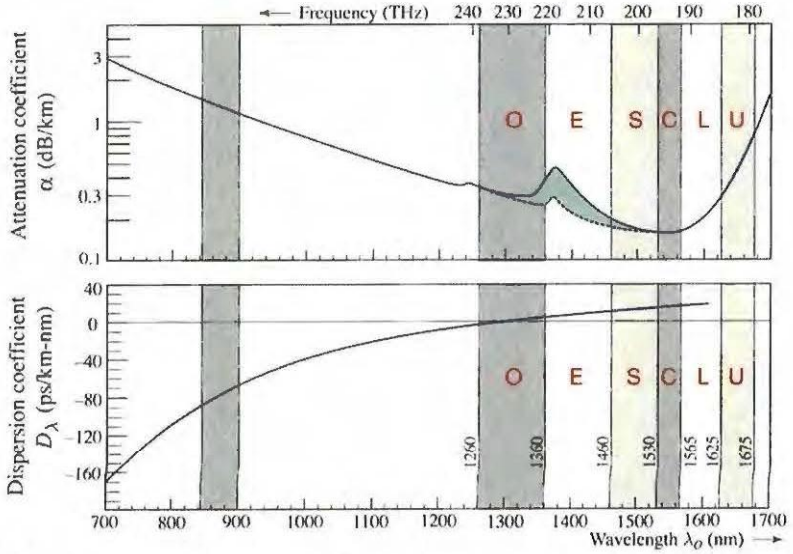


Figure 24.1-2 Wavelength dependence of the minimum attenuation α and the material dispersion coefficient D_λ of silica-glass fibers. The dashed line represents the attenuation of silica-glass fibers with suppressed OH absorption. Three spectral bands (shaded) are noted: the band centered at 870 nm, which was used in earlier systems has $\alpha = 1.5$ dB/km and $D_\lambda = -80$ ps/km-nm; the O (original) band centered at 1310 nm, for which $\alpha = 0.3$ dB/km and dispersion is minimal; and the C (conventional) band centered at 1550 nm, for which attenuation is minimal ($\alpha = 0.16$ dB/km) and $D_\lambda = +80$ ps/km-nm. Three additional bands are used in wavelength division multiplexing systems (WDM): E = extended, S = Short, L = Long; and U = Ultra-long.

Dispersion-Modified Fibers (DSF)

As described in Sec. 9.3B, advanced designs of single-mode fibers use graded-index cores with special refractive index profiles selected such that the overall chromatic dispersion coefficient D has desired values at certain wavelengths, or wavelength dependence that is useful in fiber communication systems, as in the following examples:

- In **dispersion-shifted fibers (DSF)**, D vanishes at $\lambda_o = 1550$ nm, where attenuation is minimum, rather than at 1312 nm. In **non-zero dispersion-shifted fibers (NZ-DSF)**, D is significantly reduced in the 1500–1600 nm window, but is not zero. A small amount of dispersion can be useful in alleviating nonlinear distortions encountered by narrow intense pulses. The wavelength dependence of D in DSF and NZ-DSF fibers is illustrated in Fig. 24.1-3 [see Fig. 9.3-6(a)].
- In **dispersion-flattened fibers**, D vanishes at two wavelengths and is reduced at intermediate wavelengths [see Fig. 9.3-6(b)].
- In **dispersion compensating fibers (DCF)**, D is proportional to that of the conventional step-index fiber over an extended wavelength band, but has the opposite

sign. A short fiber with a reversed large dispersion coefficient can be used to compensate the pulse spreading introduced in long conventional fibers [see Fig. 9.3-6(*c*)].

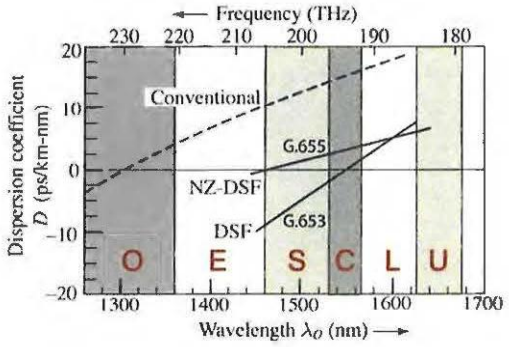


Figure 24.1-3 Wavelength dependence of the chromatic dispersion coefficient D of a conventional fiber and examples of a dispersion-shifted fiber (DSF) and non-zero-dispersion shifted fiber (NZ-DSF). The designations G.653 and G.655 are specifications of the ITU (International Telecommunications Union).

Other dispersion-modified fibers include the **holey fibers** and the **photonic-crystal fibers** (PCF), described in Sec. 9.4. In these fibers, chromatic dispersion is dominated by waveguide dispersion, which is strongly dependent on the geometry of the holes. Dispersion flattening over broad wavelength ranges can be achieved as can dispersion shifting to wavelengths lower than the zero-material-dispersion wavelength. A holey fiber may be designed to operate as a single-mode waveguide over a broad range of wavelengths (endlessly single-mode fibers). In fibers with a hollow core and a cladding with holes arranged in a periodic structure, light is guided in the core by reflection from the surrounding photonic-crystal cladding. Since the light travels in the hollow core, it suffers lower losses and reduced nonlinear effects.

Polarization-Mode Dispersion

Another form of pulse spreading, known as **polarization-mode dispersion (PMD)**, is caused by random anisotropic changes in the fiber introduced by environmental factors along its length. Random variations in the magnitude and orientation of the birefringence introduce differential delays between the two polarization modes, and as described in Sec. 9.3B, the average RMS value of the pulse broadening associated with PMD is proportional to the square root of the fiber length:

$$\sigma_{\text{PMD}} = D_{\text{PMD}}\sqrt{L}, \tag{24.1-4}$$

Polarization-Mode Dispersion

where D_{PMD} is a dispersion parameter typically ranging from 0.1 to 1 ps/$\sqrt{\text{km}}$. PMD becomes important at high data rates when other forms of dispersion are compensated.

Nonlinear Optical Effects

Silica-glass fibers exhibit two kinds of optical nonlinear effects – third-order nonlinearity, which underlies the optical Kerr effect, and nonlinear inelastic scattering, which includes stimulated Raman and Brillouin scattering. When high-power optical pulses are transmitted through single-mode fibers, which have small cross-sectional area, the optical intensity may be sufficiently high for these nonlinear interactions to occur, causing a number of deleterious effects that damage the signal integrity in communication systems:

- *Self-phase modulation (SPM)* is a form of nonlinear dispersion caused by the optical Kerr effect (the slight dependence of the refractive index, and hence the phase velocity, on the optical intensity, as described in Sec. 21.3A). Since different segments of the optical pulse travel at different velocities, pulse spreading ensues (see Sec. 22.3B). The optical Kerr effect may also result in crosstalk between counter–propagating waves in two-way communication systems.
- *Cross-phase modulation (XPM)* results from nonlinear wave mixing wherein the phase velocity of a wave at one wavelength depends on the intensities of waves at other wavelengths traveling simultaneously in the same fiber (see Sec. 21.3C). In wavelength-division multiplexing (WDM) systems, XPM can cause serious crosstalk between the different channels.
- *Four-wave mixing (FWM)* is also associated with third-order nonlinear effects (see Sec. 21.3D). It causes crosstalk between four waves of different wavelengths traveling simultaneously in the same fiber since the waves may exchange energy. This introduces an intensity-dependent gain/loss into channels of a WDM system.
- *Stimulated Raman scattering (SRS)* and stimulated Brillouin scattering (SBS) are inelastic scattering processes involving interactions between light and molecular or acoustic vibrations of the medium. In these processes, two optical waves of different wavelengths interact via a molecular vibration mode (SRS) or an acoustic wave (SBS) (see Secs. 13.5C, 14.3D, and 15.3A). Such interactions also lead to undesirable crosstalk between channels of a WDM system.

The nonlinear properties of fibers may also be harnessed for useful applications in communication systems. Nonlinear dispersion via SPM may be adjusted to compensate for chromatic dispersion in the fiber. The result is the spreadless pulses known as optical solitons (see Sec. 22.5B). Nonlinear interactions can also be used to provide useful gain via FWM or SRS. Optical Raman Amplifiers are described in Sec. 14.3D.

B. Sources for Optical Transmitters

The basic requirements for the light sources used in optical communication systems depend on the nature of the intended application (long-haul communication, local-area network, etc.). The principal features are:

- *Power*. The source power must be sufficient so that, after transmission through the fiber, the received signal is detectable with the required accuracy.
- *Speed*. It must be possible to modulate the source power at the rate desired for imparting information.
- *Linewidth*. The source must have a narrow spectral linewidth so that the effect of chromatic dispersion in the fiber is minimized.
- *Noise*. Random fluctuations in the source power must be avoided, particularly for coherent communication systems.
- *Other features*. Other important features include ruggedness, insensitivity to environmental changes such as temperature, reliability, low cost, and long lifetime.

Both light-emitting diodes (LEDs) and laser diodes (LDs) are used as sources in optical fiber communication systems. These devices are discussed in Chapter 17.

Light-emitting diodes are fabricated in two basic structures: surface emitting and edge emitting. *Surface-emitting diodes* have the advantages of ruggedness, reliability, lower cost, long lifetime, and simplicity of design. The basic limitation attendant to their use is their relatively broad linewidth, which can exceed 100 nm in the 1300–1600-nm band (see Fig. P17.1-5). When operated at maximum power, modulation frequencies up to 100 Mb/s are possible, but higher speeds (up to 500 Mb/s) can be attained at reduced powers. The *edge-emitting diode* has a structure similar to that of a laser diode without a feedback mechanism. It produces more power output with

relatively narrower spectral linewidth, at the expense of increased complexity.

Laser diodes have the advantages of high power (tens of W), high speeds (many Gb/s), narrow spectral linewidths (tens of MHz), and ease of coupling into single-mode fiber. However, they are sensitive to temperature variations. Multimode laser diodes suffer from partition noise, which is a random distribution of the laser power among the modes. When subjected to chromatic dispersion in the fiber, this leads to random intensity fluctuations and reshaping of the transmitted pulses. Laser diodes also suffer from frequency chirping, which is a change of the laser frequency as the optical power is modulated. Chirping results from modifications of the refractive index that accompany changes of the charge-carrier concentrations as the injected current is altered.

As discussed in Chapter 17, AlInGaP/InGaP LEDs are often used in inexpensive plastic-fiber communications systems that operate in the 600–650-nm region (see Fig. 17.1-19). More common, however, is $In_{1-x}Ga_xAs_{1-y}P_y$, a versatile alloy that is widely used for the fabrication of both LEDs and LDs in the near-infrared region of the spectrum. It offers a direct bandgap that is compositionally tunable over a substantial range of wavelengths, and lattice matching to an InP substrate can be maintained. InGaAsP is used to fabricate LEDs for short-haul, modest-bit-rate communications systems operating at $\lambda_o = 1.3$ μm (see Fig. 17.1-18). Long-haul high-bit-rate communication systems generally operate at $\lambda_o = 1.55$ μm and make use of laser diodes rather than LEDs since it is far easier to couple the highly collimated light emitted by an LD into a single-mode fiber.

The most common laser-diode configuration is the distributed feedback (DFB) laser (Fig. 24.1-4). As discussed in Sec. 17.4C, this device makes use of a corrugated-layer grating adjacent to the active region, which acts as a distributed reflector that substitutes for the mirrors of a Fabry–Perot laser. This design is compatible with on-chip integration. These edge-emitting lasers offer narrow spectral widths, which is critical for the efficient operation of 1.3 and 1.55 μm wavelength-division-multiplexed (WDM) optical communication systems.

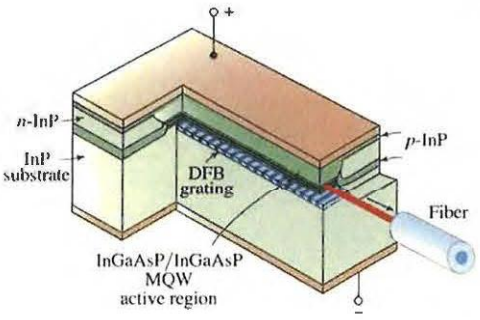


Figure 24.1-4 Buried-heterostructure multiquantum-well distributed-feedback laser used for optical fiber communications. The dielectric film provides gain guiding while the alternating p and n-type layers allow current flow only in the vicinity of the active region. Lasers such as these offer ample gain at modest current levels, and provide output powers of 1 W or more in a single spatial mode. Typical values of the threshold current and differential responsivity are $i_t < 10$ mA and $\mathcal{R}_d \approx 0.4$ W/A, respectively.

C. Optical Amplifiers

Optical amplifiers are indispensable components in modern long-haul optical fiber communication systems. They find use as postamplifiers (power amplifiers), line amplifiers, and preamplifiers. As shown in Fig. 24.1-5, power amplifiers augment the optical power before light is launched into an optical fiber, line amplifiers serve as repeaters to boost the signal in the course of transmission (see Fig. 24.0-1), and preamplifiers provide gain before photodetection.

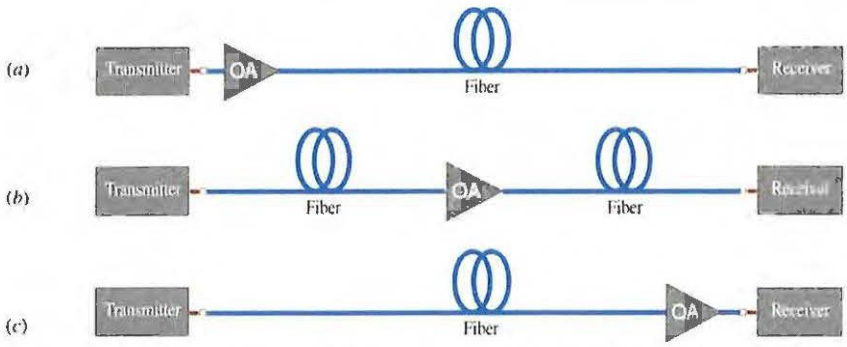


Figure 24.1-5 Optical fiber amplifiers are used in three configurations in an optical fiber communication system: (*a*) postamplifiers; (*b*) line amplifiers; and (*c*) preamplifiers.

We consider three kinds of optical amplifiers:

- Optical fiber amplifiers (OFAs). These include erbium-doped fiber amplifiers (EDFAs) (Sec. 14.3C), rare-earth-doped fiber amplifiers (REFAs) (Sec. 14.3C), and Raman fiber amplifiers (RFAs) (Sec. 14.3D).
- Semiconductor optical amplifiers (SOAs) (Sec. 17.2).
- Optical parametric amplifiers (OPAs) (Sec. 21.4C).

With the exception of the OPA, all of the optical amplifiers listed above are nonparametric devices inasmuch as they rely on an exchange of energy between the field and the amplifying medium (see introduction to Chapter 21). EDFAs and RFAs turn out to be the most suitable amplifiers for optical fiber communications, as discussed below.

Optical Fiber Amplifiers (OFAs)

OFAs comprise three varieties: EDFAs, REFAs, and RFAs:

Erbium-doped fiber amplifiers (EDFAs). Erbium-doped fiber amplifiers (EDFAs), which were the first OFAs to be developed, are widely used in optical fiber communication systems. As discussed in Sec. 14.3C, they offer high polarization-independent gain, high output power, low insertion loss, low noise, and a broad transition near $\lambda = 1.55\ \mu$m (corresponding to the wavelength of minimum loss for silica optical fibers, as shown in Fig. 24.1-2). Pumping is achieved by longitudinally coupling light into the amplifying medium, usually from strained quantum-well InGaAs laser diodes operating at $\lambda_o = 980$ nm. The pump light may be injected in the forward or backward direction, or bidirectionally.

Gains in excess of 50 dB can be achieved in EDFAs with tens of mW of pump power, and signal output powers in excess of 100 W are readily generated. The available bandwidth is $\Delta\lambda \approx 40$ nm, corresponding to $\Delta\nu \approx 5.3$ THz, which accommodates the C band. The L band is also readily covered although the optimization parameters of the EDFAs are not the same in the two bands. The large gain and bandwidth offered by these amplifiers make them highly suitable for use in wavelength-division multiplexing (WDM) systems (see Sec. 24.3C).

Rare-earth-doped fiber amplifiers (REFAs). Several ions other than Er^{3+} (e.g., Pr^{3+}, Tm^{3+}, and Nd^{3+}) are useful for making rare-earth-doped fiber amplifiers (REFAs) that cover the O/E/S/U bands (see Fig. 24.2-3). REFAs can therefore be used to extend the amplification bandwidth well beyond the 60-nm (7.5-THz) bandwidth achievable by using individually optimized EDFAs in the C and L bands. Unfortunately, however, REFAs other than Er^{3+} function far better with fluoride and tellurite

glasses than with silica glass. This materials challenge is not easily surmounted, for a number of reasons: (1) silica-glass fiber has been widely used and its manufacturing technology is entrenched; (2) each type of REFA requires its own fiber-glass matrix; (3) the splicing of different glasses is not straightforward; and (4) each type of REFA requires its own laser-diode pump at the appropriate wavelength. Nevertheless, by mixing and matching REFAs of Er^{3+} and Tm^{3+}, the available bandwidth $\Delta\lambda$ can be increased from 60 nm to $\approx$ 150 nm, corresponding to $\Delta\nu \approx 18.8$ THz at 1550 nm.

Raman fiber amplifiers (RFAs). Raman fiber amplifiers operate on the basis of stimulated Raman scattering (see Sec. 13.5C). As discussed in Sec. 14.3D, there are two standard RFA configurations: (1) distributed RFAs where the signal and pump are both sent through the transmission fiber, which serves as the gain medium, and (2) lumped RFAs in which a short length of highly nonlinear fiber serves as the amplifier and provides gain. As with the EDFA, pumping can be in the forward or backward direction, or bidirectional.

The bandwidth over which Raman gain is available in silica fiber is about 100 nm (corresponding to about 12.5 THz at 1550 nm) so RFAs typically offer greater bandwidths than EDFAs. Furthermore, multiple pumps of different frequencies can be combined to provide yet greater bandwidths; indeed, Raman amplification can, in principle, be employed over the entire region of fiber transparency. The gain of an RFA, which is $\approx$ 20 dB is substantially lower than that of an EDFA, as is the efficiency, but they can be increased by making use of dispersion-compensating fiber. Also, polarization-diverse pumping is required. The relative merits of EDFAs and RFAs have been considered in Sec. 14.3D. In spite of the shortcomings of RFAs in comparison with EDFAs, their wider bandwidths (extendable by using multiple pumps), arbitrary wavelength of operation, and compatibility with existing systems make them increasingly competitive as OFAs.

Semiconductor Optical Amplifiers (SOAs)

Semiconductor optical amplifiers (SOAs) (see Sec. 17.3) can be made to operate in any region of the optical spectrum by judiciously choosing the semiconductor material. They are compact and compatible with integrated optoelectronic circuits, particularly as postamplifiers or preamplifiers on an integrated optoelectronic circuit, and can be electrically pumped. SOAs designed for optical transmission applications in the near infrared are usually fabricated from InGaAsP, InGaAs, or InP. In the 1.3–1.6-μm communications band, achievable bandwidths using quantum-well SOAs are $\Delta\lambda \approx 50$ nm, corresponding to $\Delta\nu \approx 6.5$ THz at $\lambda_o = 1550$ nm, although quantum-dot SOAs can provide nearly 200 nm of bandwidth. However, because of their low gain ($\approx$ 15 dB), optical transmission applications have principally been limited to metropolitan optical networks where low gain suffices to overcome losses associated with multiple optical add–drop nodes.

As discussed in Sec. 17.2D, SOAs have a number of disadvantages with respect to OFAs: they are incompatible with fiber geometry, exhibit substantial interchannel and intersymbol interference, high noise, sensitivity to temperature, and residual sensitivity to signal polarization. As a result, they hold greater appeal for applications such as all-optical switching in optical networks (see Chapter 23) and wavelength conversion than as linear optical amplifiers.

Optical Parametric Amplifiers (OPAs)

The optical parametric amplifier (OPA) discussed in Secs. 21.2C and 21.4C has the merit that it offers substantial gain and broadband tunability over an extended spectral region that stretches from the infrared to the visible. However, it has a number of features that limit its deployment in wavelength-division-multiplexed (WDM) applications:

- The WDM signals must be phase matched to the pump, which requires dispersion flattening.
- Large-scale WDM implementation with equal spacing of channels is impeded by the presence of four-wave mixing.
- This amplifier is sensitive to signal polarization so that polarization-multiplexed pumping is required.

We conclude that SOAs and OPAs are less useful than OFAs in optical fiber communication systems.

D. Detectors for Optical Receivers

A comprehensive discussion of semiconductor photon detectors is provided in Chapter 18. Two types of detectors are commonly used in optical communication systems: the p–i–n photodiode and the avalanche photodiode (APD). The APD has the advantage of providing gain before the first electronic amplification stage in the receiver, thereby reducing the detrimental effects of circuit noise. However, the gain mechanism itself introduces noise and has a finite response time, which may reduce the bandwidth of the receiver. Furthermore, APDs require greater voltage and more complex circuitry to compensate for their sensitivity to temperature fluctuations. The signal-to-noise ratio and the sensitivity of receivers using p–i–n photodiodes and APDs are discussed in Sec. 18.6.

Detectors in the 870-nm Wavelength Range

Silicon p–i–n photodiodes and APDs are used at these wavelengths. In state-of-the-art preamplifiers, silicon APDs enjoy a 10-to-15-dB sensitivity advantage over silicon p–i–n photodiodes because their internal gain makes the noise of the preamplifier relatively less important.

Detectors in the 1300–1600-nm Wavelength Range

Silicon cannot be used in this wavelength region because it is transparent (see Fig. 5.5-1); this is because its bandgap wavelength lies below the wavelength of the light ($\lambda_g < \lambda_o$). Rather, InGaAs and Ge p–i–n photodiodes are used, but InGaAs is preferred because of its smaller dark noise and greater thermal stability. Typical InGaAs p–i–n photodiodes have quantum efficiencies $\eta \approx 0.8$, responsivities $\mathfrak{R} \approx 1$ A/W (see Fig. 18.3-9), bandwidths ≈ 10 GHz into 50Ω, and dark currents ≈ 0.1 nA. Waveguide structures offer larger bandwidths.

InGaAs APDs are widely used. Like all narrow-bandgap materials, however, InGaAs suffers from large tunneling leakage currents when subjected to strong electric fields. This problem is mitigated by making use of a heterostructure with a small-bandgap material for the absorption region and a larger-bandgap material for the multiplication region (SAM APD). Figure 24.1-6 illustrates a variation on this theme: a separate-absorption-grading-multiplication (SAGM APD), in which the absorption takes place in InGaAs and the multiplication in InP. The InGaAsP grading layer provides a smooth transition between the two regions.

Since holes multiply in this device, the salient ionization ratio is $1/k$ (see Fig. 18.6-4). For InP, $1/k \approx 0.5$ when the mean gain $\overline{G} = 10$, so the gain noise is substantially greater than that in Si. Nevertheless, these devices work well; they typically have efficiencies $\eta \approx 0.8$, responsivities $\mathfrak{R} \approx 10$, bandwidths ≈ 10 GHz, gain-bandwidth products ≈ 100 GHz, and dark currents ≈ 0.1 nA.

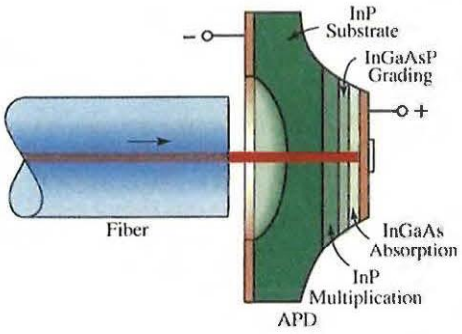


Figure 24.1-6 Structure of a separate-absorption-grading-multiplication (SAGM) APD.

24.2 OPTICAL FIBER COMMUNICATION SYSTEMS

The simplest communication system is a point-to-point link. The information is carried by a signal — a physical variable (electrical, electromagnetic, optical, etc.) *modulated* at one point and observed at the other. To transmit more than one signal simultaneously through the same link, the signals must be marked by some distinct attribute (e.g., time, frequency, or wavelength), or identified by some distinct code. The scheme is called *multiplexing*.

In an optical fiber communication system, the link is an optical fiber through which a light wave modulated by the signal is transmitted. The modulated physical variable carrying the information may be the optical intensity, amplitude, frequency, phase, or polarization. The simplest example is the intensity-modulation communication system illustrated in Fig. 24.2-1. The simplest example of optical multiplexing is wavelength-division multiplexing (WDM), in which multiple signals are transmitted through the same fiber at distinct optical wavelengths, as illustrated in Fig. 24.2-2.

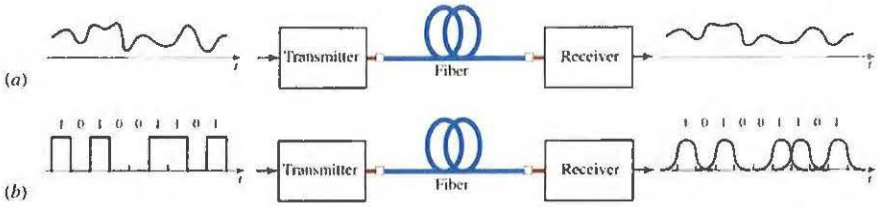


Figure 24.2-1 Optical fiber communication systems using intensity-modulation. (*a*) Analog system: the power of the light source is proportional to the signal, which is a continuous function of time representing, e.g., an audio or video waveform. (*b*) Digital ON–OFF keying: bits "1" and "0" are represented, respectively, by the presence and absence of an optical pulse.

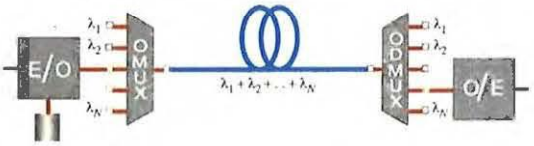


Figure 24.2-2 Wavelength-division multiplexing (WDM)

One measure of the performance of an analog communication system is the **bandwidth** B (Hz). It is the maximum frequency at which modulated optical power may be

transmitted through the link such that the received signal is detectable with a prescribed signal-to-noise ratio. The bandwidth is determined by the response time of the overall communication channel and by the attenuation and the noise level at the receiver.

Similarly, a measure of the performance of a digital communication system is the maximum **bit rate** B_0 (bits per second, or b/s) at which bits of the received signal are discernible with error rate not exceeding a prescribed value. This data rate is determined by the attenuation and the pulse spreading introduced by the system, and also by the noise level at the receiver. The following bit rates represent optical carrier (OC) levels defined by the Synchronous Optical Network (SONET), which is a standard for optical telecommunication technology:

Table 24.2-1 Approximate bit rates of the SONET standard.

OC-1	OC-3	OC-12	OC-24	OC-48	OC-192	OC-768
52 Mb/s	156 Mb/s	622 Mb/s	1.25 Gb/s	2.5 Gb/s	10 Gb/s	40 Gb/s

This section begins with an overview of the evolution of optical fiber communication systems followed by a quantitative analysis of the performance limits of simple digital and analog systems using intensity modulation.

A. Evolution of Optical Fiber Communication Systems

As illustrated in Fig. 24.1-2, the minimum attenuation in silica glass occurs at $\approx$ 1550 nm, whereas the minimum material dispersion occurs at $\approx$ 1312 nm. The choice between these two wavelengths depends on the relative importance of power loss versus pulse spreading, as explained in Sec. 24.2B. However, the availability of an appropriate light source is also a factor. First-generation optical fiber communication systems operated at $\approx$ 870 nm (the wavelength of AlGaAs light-emitting diodes and laser diodes), where both attenuation and material dispersion are relatively high. More advanced systems operate at 1300 and 1550 nm.

The various operating wavelengths, materials and types of fibers, light sources, detectors, and amplifiers that may be used for building an optical link offer many possible combinations, some of which are summarized in Fig. 24.2-3. Progress in the implementation of optical fiber systems has historically followed a path toward longer wavelengths, from multimode fibers (MMF) to single-mode fibers (SMF), from light-emitting diodes (LEDs) to laser diodes (LDs), from p–i–n (PIN) photodiodes to avalanche photodiodes (APDs), and from semiconductor optical amplifiers (SOAs) to optical fiber amplifiers (OFAs). Appropriate materials for the longer wavelengths (e.g., quaternary sources and detectors) had to be developed to make this progress possible.

The evolution of fiber components and systems has been motivated by a desire to increase the transmission bit rate B_0 (bits/s or b/s) and the length L (km) of the communication link (the repeater spacing); the product LB_0 (km-b/s) has been used as a measure of progress. The following seven systems describe this evolution and Fig. 24.2-4 depicts the increase in LB_0 over the years. The first three systems, which are often referred to as the first three generations of optical fiber systems, have achieved a 1000-fold increase in LB_0 from 1974 to the 1990s. These technologies are used as examples in the discussion of system performance in Sec. 24.2B. Subsequent progress has extended these basic systems in a number of directions, leading to an increase of LB_0 by another five orders of magnitude from 1990 to 2005. This tenfold increase every four years has been called the "**optical Moore's law**."

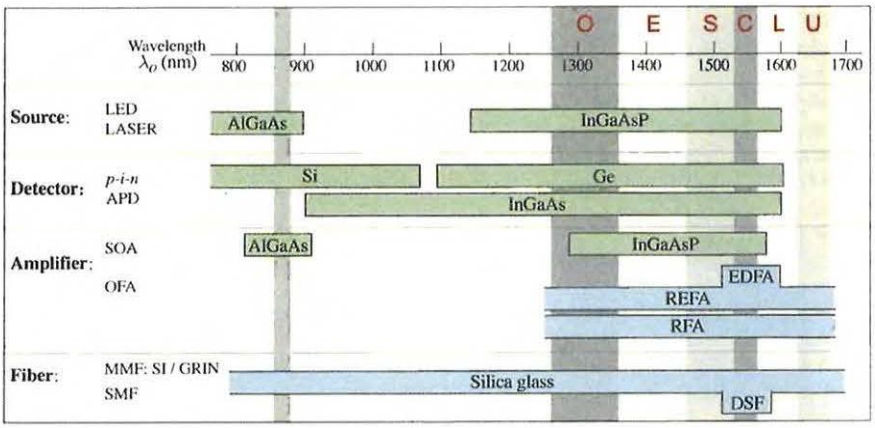


Figure 24.2-3 Types and materials of optical sources, detectors, amplifiers, and fibers used at various wavelengths. The first generations of optical links operated at wavelengths near 870 nm, 1310 nm, and 1550 nm.

System 1: *Multimode fiber (MMF) at 870 nm.* This is the early technology of the 1970s. Fibers are either step-index or graded-index. The light source is either an LED or a laser (initially GaAs and subsequently AlGaAs). Both silicon p–i–n and APD photodiodes are used. The performance of this system is limited by the fiber's high attenuation and modal dispersion. A typical intercity communication link of this era operated at B_0 = 100 Mb/s, with a repeater spacing L = 10 km, i.e., $LB_0 = 1$ km-Gb/s.

System 2: *Single-mode fiber (SMF) at 1310 nm.* The move to single-mode fibers and a wavelength region where material dispersion is minimal led to a substantial improvement in performance, limited by fiber attenuation. InGaAsP lasers are used with either InGaAs p–i–n or APD photodetectors (Ge APDs are also sometimes used). A typical long-haul link in this class operated at OC-12 (622 Mb/s) with repeater spacing L = 40 km and $LB_0 \approx 25$ km-Gb/s.

System 3: *Single-mode fiber (SMF) at 1550 nm.* At this wavelength the fiber has its lowest attenuation. Performance is limited by material dispersion, which is reduced by the use of low-chirp single-frequency distributed-feedback (DFB) lasers (InGaAsP). Subsequent use of dispersion-shifted fibers (DSF) has alleviated the dispersion problem and boosted the performance. An example of this system is a long-haul terrestrial or undersea link operating at 2.5 Gb/s (OC-48) over a distance L = 100 km, for which $LB_0 \approx 250$ km-Gb/s. Advances in transmitters and receivers have boosted this system to 10 Gb/s (OC-192), bringing LB_0 to one km-Tb/s.

System 4: *Coherent system.* As described in Sec. 24.5, rather than measuring the intensity of the signal light directly by a photodetector, a coherent system makes use of **coherent detection**, in which light from a local source (called the local oscillator) is mixed with the signal light at the detector. The use of coherent detection enhances the receiver sensitivity thus allowing greater communication distances; however, this comes at the expense of increased complexity. As a result, the commercial implementation of coherent systems has lagged behind that of direct-detection systems, particularly as a result of the emergence of optical fiber amplifiers.

System 5: *Link with optical amplifiers.* The advent of semiconductor and optical fiber amplifiers (see Sec. 24.1C) has had a dramatic impact on the performance of optical fiber communication systems. Placed periodically along the fiber, these amplifiers compensate attenuation and therefore extend the distance between electronic repeaters. An example is the transpacific TPC-5 system, which operates at bit rates of up to $B_0 = 10$ Gb/s with distances up to $L = 20{,}000$ km, i.e., $LB_0 = 200$ km-Tb/s.

System 6: *Optical soliton system.* Solitons are short (typically 1 to 50 ps) optical pulses that can travel through long optical fibers without changing the shape of their pulse envelope. As discussed in Sec. 22.5B, the effects of fiber dispersion and nonlinear self-phase modulation (arising, for example, from the optical Kerr effect) precisely cancel each other, so that the pulses act as if they were traveling through a linear nondispersive medium. Erbium-doped fiber amplifiers are effectively used in conjunction with soliton transmission to overcome absorption and scattering losses. Experimental systems have been operated at 10 Gb/s over fiber lengths in excess of 12,000 km ($LB_0 = 1200$ km-Tb/s).

System 7: *Wavelength-division multiplexing (WDM).* The introduction of WDM has provided a significant increase in the capacity of the system by the use of multiple wavelengths (channels) transmitted through the same fiber. Broadband optical amplifiers are used to provide simultaneous amplifications for all channels. An example is the TPC-6 system for which $B_0 = 100$ Gb/s, $L = 9{,}000$ km, and $LB_0 = 900$ km-Tb/s. In combination with dispersion-managed transmission and forward error correction, rates of 5–10 Tbit/s per fiber over distances of 10 000 km are now possible.

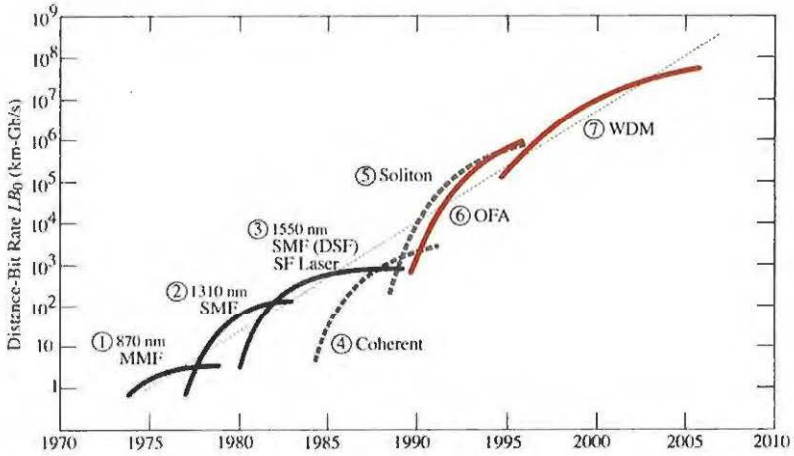


Figure 24.2-4 The history of optical fiber communication systems comprises continuous improvement of the bit rate distance product LB_0.

B. Performance of Optical Fiber Systems

The first step in assessing the performance of a fiber communication system is to come up with a mathematical model describing the effect of the various system components, principally the optical fiber, on the modulated signal. This permits us to estimate the

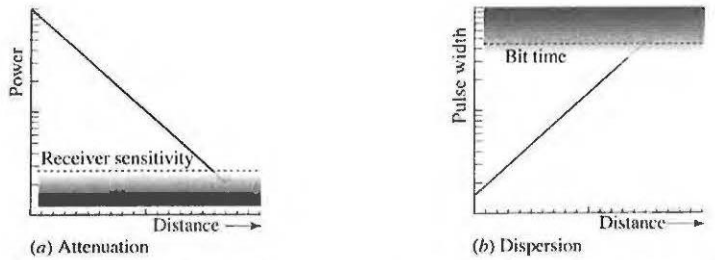
shape of received distorted signal, and hence determine the signal-to-noise ratio in analog systems and the expected bit error rate in digital systems.

In most applications, the fiber may be treated as a linear system described by an impulse response function $h(t)$ or its Fourier transform, the transfer function $H(f)$, where f is the modulation frequency. Three important parameters characterize these functions:

- *Power transmission.* This is the fraction of steady (unmodulated) input optical power received at the output. It equals the transfer function $H(f)$ at $f = 0$. Since $H(f)$ is the Fourier transform of $h(t)$, $H(0) = \int h(t)\,dt$ is the area under $h(t)$. For a fiber of length L and attenuation coefficient α (dB/km), $H(0) = \exp[-\alpha L]$, where $\alpha \approx 0.23\alpha$ is the attenuation coefficient in units of km^{-1}. Localized power losses at couplers may also be included in α in distributed units of dB/km.
- The *response time* σ_τ is the width of $h(t)$. It determines the temporal spreading of optical pulses and therefore sets the maximum data rate that can be used in digital systems. The response time is proportional to the fiber length. For example, in a single-mode fiber, $\sigma_\tau = |D|\sigma_\lambda L$, where σ_λ (nm) is the source linewidth and D (ps/km-nm) is the dispersion coefficient (ps/km-nm).
- The *bandwidth* σ_f (Hz) is the width of the transfer function $|H(f)|$. In an analog system, the bandwidth determines the maximum frequency at which the input power may be modulated and successfully detected by the receiver. Since $H(f)$ and $h(t)$ are related by a Fourier transform, the bandwidth σ_f is inversely proportional to the response time σ_τ. The coefficient of proportionality depends on the actual profile of $h(t)$ (see Appendix A, Sec. A.2). Here, we use the relation $\sigma_f = 1/2\pi\sigma_\tau$ for the purpose of illustration.

The maximum fiber length that can be used to transmit a signal with a desired performance level is set by the following principal impairments introduced by the system:

- *Attenuation* results in an exponential drop of the optical power as a function of distance [Fig. 24.2-5(*a*)]. At a distance for which the received power becomes smaller than the receiver sensitivity (the minimum power required by the receiver), the system's performance becomes unacceptable.
- *Dispersion* results in an increase of the width of the optical pulses that represent data bits in a digital system as a function of distance [Fig. 24.2-5(*b*)]. When the width exceeds the bit interval, adjacent pulses overlap, resulting in intersymbol interference (ISI), which introduces undesirable errors. In an analog system, dispersion washes out high-frequency components of the modulated signal and reduces the system's bandwidth.



(*a*) Attenuation (*b*) Dispersion

Figure 24.2-5 (*a*) Dependence of the optical power on the distance. (*b*) Dependence of the pulse width on the distance. The maximum length of the optical link is set by either (*a*) attenuation, when the received power drops below the receiver sensitivity, or (*b*) dispersion, when the pulse width exceeds the bit time.

- *Noise* added by optical components, such as optical amplifiers, and by random propagation effects, such as polarization mode dispersion, introduces additional errors.
- *Nonlinear distortion* associated with intense optical pulses results in the cross mixing of spectral components, and the introduction of interference between multiplexed signals in wavelength-division multiplexing (WDM) systems.

The communication system is more sensitive to transmission impairments at high bit rates (or high modulation frequencies) because of the following effects:

- For a fixed average power, a higher bit rate corresponds to fewer photons per bit, and therefore to greater photon noise. Other noise sources in the receiver also become more important at high data rate. The receiver sensitivity is therefore an increasing function of the bit rate [Fig. 24.2-6(*a*)].
- A higher bit rate corresponds to shorter pulses [Fig. 24.2-6(*b*)] with broader spectra and greater dispersion. Such pulses undergo greater broadening, which leads to greater intersymbol interference (ISI).
- For a fixed optical energy per bit, a higher bit rate (shorter bit time) requires greater optical power [Fig. 24.2-6(*c*)], which evokes nonlinear interactions leading to nonlinear ISI.

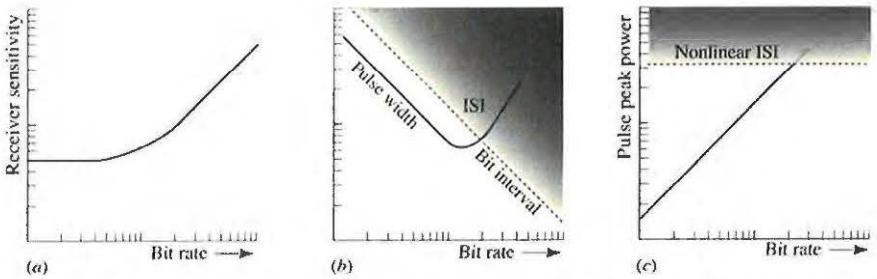


Figure 24.2-6 Effect of bit rate on (*a*) receiver sensitivity, (*b*) pulse width at the receiver, and (*c*) peak power. At higher bit rate, the communication system is more sensitive to attenuation, dispersion, and nonlinear effects.

As we will see in the remainder of this section, the design of a long-haul high-bit-rate optical fiber communication link involves the selection of fibers with the lowest attenuation and/or dispersion, and careful power and pulse width budgeting, while guarding against the deleterious nonlinear effects associated with ultra-intense pulses.

Bit-Error Rates

The performance of a *digital* communication system is measured by the **probability of error** per bit, which is referred to as the **bit error rate** (BER). For an ON–OFF keying system, such as that shown in Fig. 24.2-1, bits "1" and "0" are represented, respectively, by the presence and absence of an optical pulse. If p_1 is the probability of mistaking "1" for "0," and p_0 is the probability of mistaking "0" for "1," and if the two bits are equally likely to be transmitted, then $\text{BER} = \frac{1}{2}p_1 + \frac{1}{2}p_0$. A typical acceptable BER is 10^{-9} (i.e., an average of one error every 10^9 bits).

Errors occur as a result of noise in the received signal, or due to pulse spreading into neighboring bits, which results in intersymbol interference. Figure 24.2-7 shows an example of random realizations of the pulse corresponding to bit "1," superimposed with random realizations of the signal received from possible neighboring pulses when

the bit is "0." This diagram is known as the **eye diagram**. The more open the "eye," the more distinguishable are the "1" and "0" bits and the less the likelihood of error.

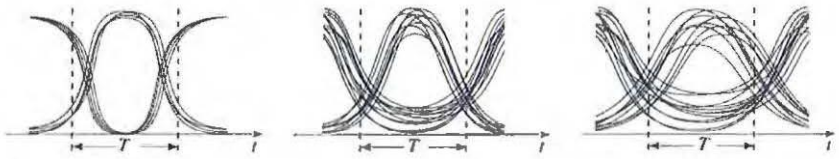


Figure 24.2-7 Closing of the eye diagram (left to right) as a result of noise and pulse broadening.

Receiver Sensitivity

The sensitivity of a digital optical receiver is defined as the minimum number of photons (or the corresponding optical energy) per bit necessary to guarantee that the rate of error (BER) is smaller than a prescribed rate (e.g., 10^{-9}). Errors occur because of the randomness of the number of photoelectrons detected during each bit, as well as the noise in the receiver circuit itself. The sensitivity of receivers using various photodetectors is discussed in Sec. 18.6E.

For example, when the light source is a stabilized laser, the detector has unity quantum efficiency, and the receiver circuit is noise-free, then an average of at least $\overline{n}_0 = 10$ photons per bit is required to ensure that BER $\leq 10^{-9}$. Therefore, *the sensitivity of the ideal receiver is 10 photons/bit.* This means that bit "1" should carry an average of at least 20 photons, since bit "0" carries no photons. In the presence of other forms of noise, a larger number of photons is required.

A sensitivity of $\overline{n}_0$ photons corresponds to an optical energy $h\nu\overline{n}_0$ per bit and an optical power $P_r = (h\nu\overline{n}_0)/(1/B_0)$,

$$P_r = h\nu\overline{n}_0 B_0, \tag{24.2-1}$$

which is proportional to the bit rate B_0. As the bit rate increases, a higher optical power is required to maintain the number of photons/bit (and therefore the BER) constant.

It is shown in Sec. 18.6E that when circuit noise is important, the receiver sensitivity $\overline{n}_0$ depends on the receiver bandwidth (i.e., on the data rate B_0). This behavior complicates the design problem. For simplicity, we shall assume in the following analysis that the receiver sensitivity (photons per bit) is independent of B_0.

C. Attenuation- and Dispersion-Limited Systems

In this section, we examine the performance limits imposed by *attenuation* and *dispersion* on a digital intensity modulation ON–OFF keying (OOK) system. Nonlinear effects are ignored and the fiber transmission system itself is assumed to introduce no noise.

Consider an optical fiber link operated as a digital communication system at a data rate of B_0 bits/s over a distance of L (km). The source has power P_s (mW), wavelength λ (nm), and spectral width σ_λ (nm). The fiber has attenuation coefficient α (dB/km) and chromatic dispersion coefficient D_λ (ps/km-ns). The receiver has a sensitivity of $\overline{n}_0$ (photons per bit), corresponding to power sensitivity $P_r = (hc/\lambda)\overline{n}_0 B_0$ (mW), which must be received for the system to operate at an acceptable error rate.

The performance limits are established by determining the maximum distance L over which the link can transmit B_0 bits/s without exceeding the prescribed bit-error

rate. Clearly, L decreases with increase of B_0. Alternatively, we may determine the maximum bit rate B_0 a link of length L can transmit with an error rate not exceeding the allowable limit. The maximum bit-rate–distance product LB_0 serves as a single number that describes the capability of the link. We shall determine the typical dependence of L on B_0, and derive expressions for the maximum bit-rate–distance product LB_0 for various types of fibers.

Two conditions must be satisfied for acceptable operation of the link:

1. The received power must be at least equal to the receiver power sensitivity P_r. This condition is met by preparing a power budget from which the maximum fiber length is determined. A margin of 6 dB above P_r is usually specified.
2. The width of the received pulses must not significantly exceed the bit time interval $1/B_0$, or else adjacent pulses overlap and cause *intersymbol interference*, which increases the error rates. This condition is met by preparing a budget for the pulse spreading resulting from the transmitter, the receiver, and various forms of dispersion in the fiber.

If the bit rate B_0 is fixed and the link length L is increased, two situations leading to performance degradation may occur: the received power becomes smaller than the receiver power sensitivity P_r, or the received pulses become wider than the bit time $1/B_0$. If the former situation occurs first, the link is said to be *attenuation limited*. If the latter occurs first, the link is said to be *dispersion limited*.

Attenuation-Limited Performance: Power Budget

Attenuation-limited performance is assessed by preparing a power budget. Since fiber attenuation is measured in dB units, it is convenient to also measure power in dB units. Using 1 mW as a reference, dBm units are defined by

$$\boxed{\mathrm{P} = 10 \log_{10} P,} \qquad P \text{ in mW}; \quad \mathrm{P} \text{ in dBm}. \tag{24.2-2}$$

For example, $P = 0.1$ mW, 1 mW, and 10 mW correspond to $\mathrm{P} = -10$ dBm, 0 dBm, and 10 dBm, respectively. In these logarithmic units, power losses are additive.

If P_s is the power of the source (dBm), α is the fiber loss (dB/km), P_c is the splicing and coupling loss (dB), and L is the maximum fiber length such that the power delivered to the receiver is the receiver sensitivity P_r (dBm), then

$$\mathrm{P}_s - \mathrm{P}_c - \mathrm{P}_m - \alpha L = \mathrm{P}_r \qquad \text{(dB units)}, \tag{24.2-3}$$

where P_m is a safety margin. The optical power is plotted schematically in Fig. 24.2-8 as a function of the distance from the transmitter.

The receiver power sensitivity $\mathrm{P}_r = 10 \log_{10} P_r$ (dBm) is obtained from (24.2-1),

$$\mathrm{P}_r = 10 \log \frac{\overline{n}_0 h\nu B_0}{10^{-3}} \text{ dBm}. \tag{24.2-4}$$

Thus, P_r increases logarithmically with B_0, and the power budget must be adjusted for each B_0 as illustrated in Fig. 24.2-9.

The maximum length of the link is obtained by substituting (24.2-4) into (24.2-3),

$$L = \frac{1}{\alpha}\left(\mathrm{P}_s - \mathrm{P}_c - \mathrm{P}_m - 10 \log \frac{\overline{n}_0 h\nu B_0}{10^{-3}}\right), \tag{24.2-5}$$

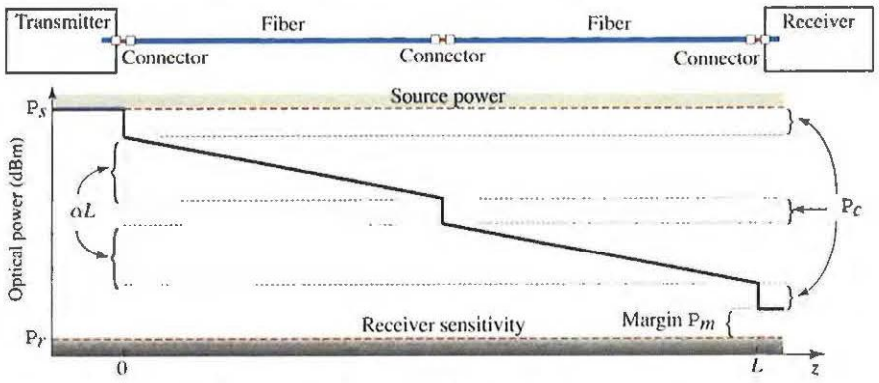


Figure 24.2-8 Power budget of an optical link.

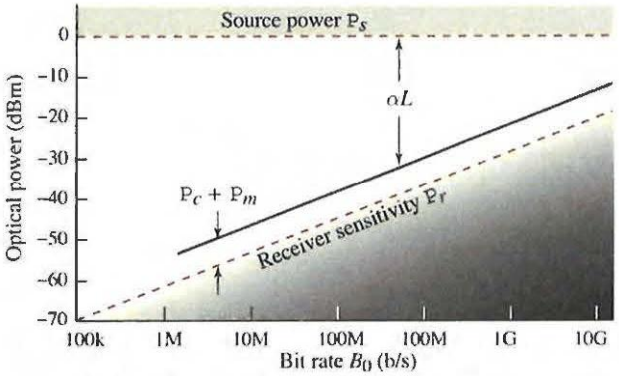


Figure 24.2-9 Power budget as a function of bit rate B_0. As B_0 increases, the power P_r required at the receiver increases (so that the energy per bit remains constant), and the maximum length L decreases.

from which

$$L = L_0 - \frac{10}{\alpha} \log B_0, \tag{24.2-6}$$

Attenuation-Limited Fiber

where $L_0 = [P_s - P_c - P_m - 30 - 10 \log(\bar{n}_0 h\nu)]/\alpha$. The length drops with increase of B_0 at a logarithmic rate with slope $10/\alpha$. Figure 24.2-10 is a plot of this relation for the operating wavelengths 870, 1300, and 1550 nm.

Dispersion-Limited Performance: Time Budget

When a pulse representing a data bit is generated by the transmitter, propagated through the fiber, and detected by the receiver, it loses power and gains width. The final pulse width σ_r depends on the original pulse width σ_s, the response time of the transmitter σ_{tx}, the response time of the fiber σ_τ, which results from various forms of dispersion, and the response time of the receiver σ_{rx}. The actual shape of the receiver pulse may be determined by convolving the original pulse profile with the impulse response

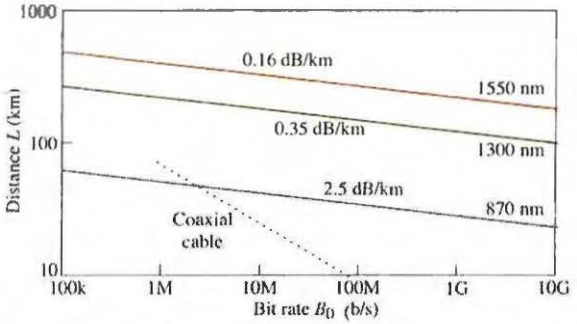


Figure 24.2-10 Maximum fiber length L as a function of bit rate B_0 under attenuation-limited conditions for a fused silica-glass fiber operating at wavelengths $\lambda_o = 870$, 1300 and 1550 nm assuming fiber attenuation coefficients $\alpha = 2.5$, 0.35, and 0.16 dB/km, respectively; source power $P_s = 1$ mW ($P_s = 0$ dBm); receiver sensitivity $\overline{n}_0 = 300$ photons/bit for receivers operating at 870 and 1300 nm and $\overline{n}_0 = 1000$ for the receiver operating at 1550 nm; and $P_e = P_m = 0$. For comparison, the L–B_0 relation for a typical coaxial cable is also shown.

functions of the transmitter, the fiber, and the receiver (assuming that all systems are linear). If all functions are Gaussian, the square of the width of the final pulse equals the sum of the squares of the widths of all constituent functions, so that

$$\sigma_o^2 = \sigma_s^2 + \sigma_{\text{sys}}^2, \tag{24.2-7}$$

where

$$\sigma_{\text{sys}}^2 = \sigma_{\text{tx}}^2 + \sigma_{\text{rx}}^2 + \sigma_\tau^2, \tag{24.2-8}$$

and σ_{sys} is the width of the response function of the communication system (transmitter + fiber + receiver). These relations are used in practical design even though the response functions are not Gaussian.

A principal design condition for the communication link ensures that the width of the received pulse does not exceed a prescribed fraction of the bit period $T = 1/B_0$, in order avoid intersymbol interference (ISI). A time budget must be prepared (Fig. 24.2-11) to ensure that this condition is met. The choice of that fraction is arbitrary and a number of ad hoc values are used. For example, some designers require that the system's response time σ_{sys} does not exceed 70 % of the bit period for non-return-to-zero (NRZ) pulses and 35 % for return-to-zero (RZ) pulses (see Fig. 24.3-4 in Sec. 24.3A for definitions of these modulation formats).

For a given receiver and transmitter, the design of the link centers around determining the maximum fiber length L. Since the only length-dependent contribution to σ_{sys} comes from the fiber σ_τ, in the following analysis we will adopt a design condition that the maximum allowed value of σ_τ be 25 % of the bit-time interval T, i.e.,

$$\sigma_\tau = \tfrac{1}{4}T = \frac{1}{4B_0}. \tag{24.2-9}$$

The choice of the factor $\frac{1}{4}$ is clearly arbitrary and serves only for comparison of the different types of fibers. We now consider the distance versus bit-rate relations that arise from this condition for the various dispersion-limited cases mentioned in Sec. 24.1A. The results are plotted in Fig. 24.2-12.

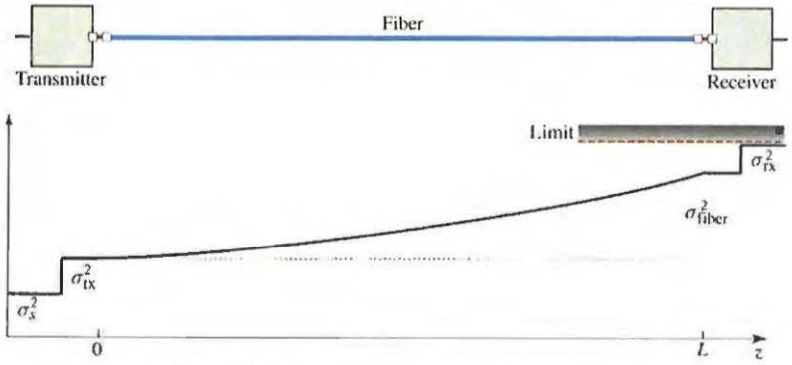


Figure 24.2-11 Budget for the pulse temporal width.

- *Multimode fiber (MMF).* For multimode fiber, the width of the received pulse after propagation a distance L is dominated by modal dispersion. For step-index fibers, (24.1-2) and (24.2-9) result in the L–B_0 relation

$$\boxed{LB_0 = \frac{c_1}{2\Delta}.} \qquad \begin{matrix}(24.2\text{-}10)\\ \text{Step-Index MMF}\end{matrix}$$

where $c_1 = c_o/n_1$ is the speed of light in the core material and $\Delta = (n_1 - n_2)/n_1$ is the fiber fractional index difference. In a graded-index fiber of optimal (approximately parabolic) refractive index profile, the pulse width is smaller by a factor $2/\Delta$, and LB_0 is greater by the same factor. For $n_1 = 1.46$ and $\Delta = 0.01$, the bit-rate–distance product $LB_0 \approx 10$ km-Mb/s for step-index fibers and $LB_0 \approx 2$ km-Gb/s for graded index fibers.

- *Single-mode fiber (SMF).* Assuming that pulse broadening in a single-mode fiber results from material dispersion only (i.e., neglecting waveguide dispersion), then for a source of linewidth σ_λ the width of the received pulse is given by (24.1-3), so that

$$\boxed{LB_0 = \frac{1}{4|D_\lambda|\sigma_\lambda},} \qquad \begin{matrix}(24.2\text{-}11)\\ \text{SMF}\end{matrix}$$

where D_λ is the dispersion coefficient of the fiber material. For operation near $\lambda_o = 1300$ nm, $|D_\lambda|$ may be as small as 1 ps/km-nm. Assuming that $\sigma_\lambda = 1$ nm (the linewidth of a single-mode laser), the bit-rate–distance product $LB_0 \approx 250$ km-Gb/s. For operation near $\lambda_o = 1550$ nm, $D_\lambda = 17$ ps/km-nm, and for the same source spectral width $\sigma_\lambda = 1$ nm, $LB_0 \approx 15$ km-Gb/s.

- *Single-mode fiber with transform-limited pulses.* To reduce chromatic dispersion, the spectral linewidth σ_λ of the source must be small. Spectral widths that are a small fraction of 1 nm are obtained with single-frequency lasers and external modulators. However, an extremely narrow spectral width is incompatible with an extremely short pulse because of the Fourier transform relation between the spectral and temporal distributions. As described in Sec. A.2 of Appendix A,

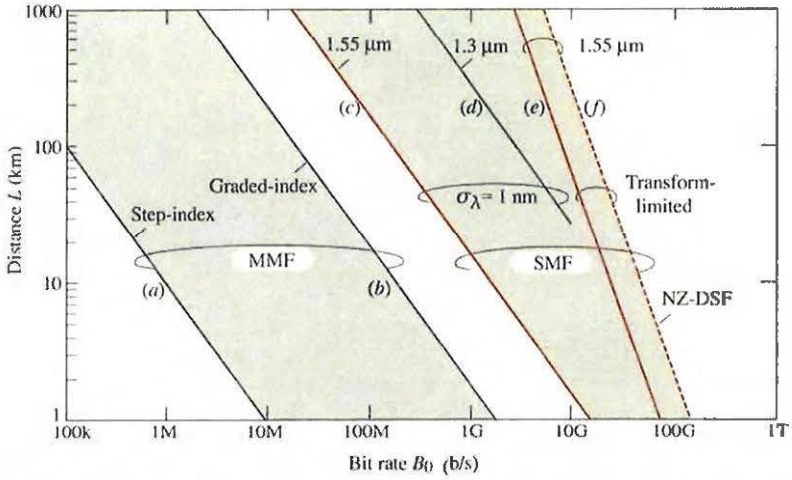


Figure 24.2-12 Dispersion-limited maximum fiber length L as a function of the bit rate B_0 for multimode fibers (MMF) and single-mode fibers (SMF). Six lines are shown (left to right): (*a*) MMF, step-index ($n_1 = 1.46, \Delta = 0.01$), $LB_0 = 10$ km-Mb/s; (*b*) MMF, graded-index with parabolic profile ($n_1 = 1.46, \Delta = 0.01$), $LB_0 = 2$ km-Gb/s; (*c*) SMF limited by material dispersion, operating at 1550 nm with $D_\lambda = 17$ ps/km-nm and $\sigma_\lambda = 1$ nm, $B_0L \approx 15$ km-Gb/s; (*d*) SMF limited by material dispersion, operating at 1300 nm with $|D_\lambda| = 1$ ps/km-nm and $\sigma_\lambda = 1$ nm, $B_0L = 250$ km-Gb/s; (*e*) SMF with transform limited pulses operating at 1550 nm with $|D_\lambda| = 17$ ps/km-nm; (*f*) same as (*e*) with non-zero dispersion-shifted fiber (NZ-DSF) with chromatic dispersion coefficient $D_\lambda = 1$ ps/km-nm.

pulses with the least product of temporal and spectral widths have a Gaussian profile. Such transform-limited pulses therefore suffer the least dispersion.

A transform-limited Gaussian pulse of width τ_0 and complex envelope $\exp(-t^2/\tau_0^2)$ has a Gaussian spectral intensity of width (FWHM) $\sigma_\nu = 0.375/\tau_0$ (see Sec. 22.1B). This corresponds to $\sigma_\lambda = |\partial\lambda_o/\partial\nu|\sigma_\nu = (\lambda_o^2/c)\sigma_\nu = 0.375\lambda_o^2/c\tau_0$. If the pulse has a width equal to half a bit period, i.e., $\tau_0 = T/2 = 1/2B_0$, then

$$\sigma_\lambda = 0.75\frac{\lambda_o^2}{c_o}B_0, \tag{24.2-12}$$

which is directly proportional to the bit rate B_0. For example, for $\lambda_o = 1550$ nm and $B_0 = 10$ Gb/s, $\sigma_\lambda = 0.06$ nm. As described in Sec. 22.3B, when a transform-limited Gaussian pulse of width τ_0 travels through a dispersive medium with dispersion coefficient D_ν it is broadened by a factor of $\sqrt{2}$ at the characteristic distance $z_0 = \pi\tau_0^2/D_\nu$. At this distance a pulse of initial width $\tau_0 = T/2$ stretches by a time $(\sqrt{2}-1)T/2 \approx 0.21T$. We may therefore take z_0 as the maximum acceptable length L of the communication link. Using the relations $L = z_0 = \pi\tau_0^2/D_\nu$, $\tau_0 = T/2 = 1/2B_o$, and $D_\nu = D_\lambda\lambda_o^2/c_o$ we obtain the distance bit-rate relation

$$LB_0^2 = \frac{\pi}{4}\frac{c_o}{|D_\lambda|\lambda_o^2}. \tag{24.2-13}$$

SMF
Transform-Limited Pulse

The maximum distance L is therefore inversely proportional to B_o^2, i.e., drops more rapidly with the data rate than in previous cases. Also, the product B_oL is inversely proportional to the data rate B_o. Figure 24.2-12 shows the $L - B_o$ relation for $\lambda_o = 1550$ nm and $D_\lambda = 17$ ps/km-nm. For example, at $B_o = 10$ Gb/s, $L = 64$ km, but at $B_o = 40$ Gb/s, L drops to 4 km. The use of transform-limited pulses therefore extends the dispersion-limited bit rate bounds substantially, although that rate drops more rapidly with further increase of bit rate.

- *Single-mode dispersion-compensated fiber with transform-limited pulses.* With single-mode fibers and transform-limited optical pulses, the maximum fiber distance for a given bit rate reaches its highest value, limited only by the dispersion coefficient. This coefficient can be reduced by use of dispersion-shifted fibers (DSF). As shown in Fig. 24.2-12, reduction of D from 17 ps/km-nm to 4 ps/km-nm, by use of DSF, increases the maximum length from 64 km to 272 km at 10 Gb/s. However, DSF fibers come with slightly higher attenuation coefficient.

Combined Attenuation- and Dispersion-Limited Performance

The attenuation-limited and dispersion-limited distance–bit-rate relations are combined in Fig. 24.2-13 by superposing Figs. 24.2-10 and 24.2-12 and selecting the smaller of the attenuation- or dispersion-limited distances. These relations describe the performance of generations of optical fibers operating at $\lambda_o = 870$ nm (multimode), at 1300 nm, and 1550 nm (single-mode). Several simplifying assumptions and arbitrary choices have been used to create this chart, and the values obtained should therefore be regarded only as indications of the order of magnitude of the relative performance of the different types of fibers. Nevertheless, a number of important conclusions can be made:

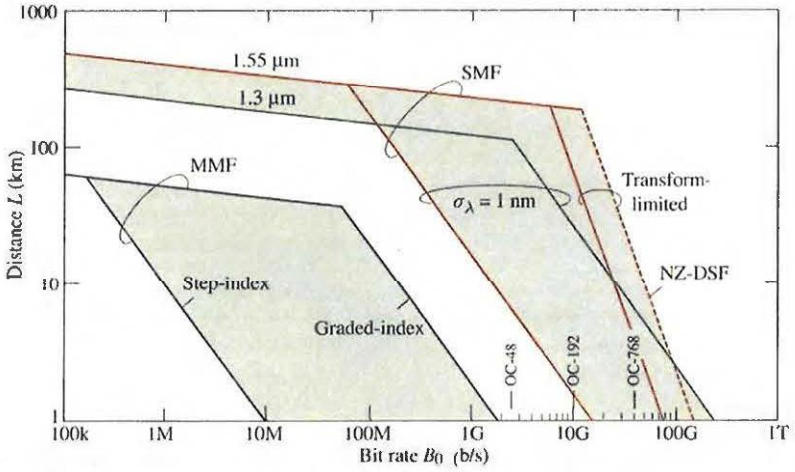


Figure 24.2-13 Maximum distance L versus bit rate B_0 for six cases of fibers. This graph is obtained by superposing the graphs in Figs. 24.2-10 and 24.2-12. Each line represents the maximum distance L of the link at each rate B_0 that satisfies *both* the attenuation and dispersion limits, i.e., guarantees the reception of the required power *and* pulse width at the receiver.

- At low bit rates, the fiber link is generally attenuation limited; L drops with B_0 logarithmically. At high bit rates, the link is dispersion limited and L is inversely

proportional to B_0 for optical pulses limited by the source linewidth and inversely proportional to B_0^2 for transform-limited.

- For high-data-rate long-haul communication links, single-mode fibers are essential. The choice between the 1300-nm and the 1550-nm wavelengths is not obvious since, for conventional fibers, chromatic dispersion is smallest at 1300 nm while the attenuation is smallest at 1550 nm. This explains the cross-over of the $L - B_o$ lines at these wavelengths.
- By use of dispersion-shifted fibers (DSF), it is possible to reduce the overall chromatic dispersion coefficient at 1550 nm, making operation at 1550 nm generally superior to operation at 1300 nm.

Performance of Analog Communication System

As in digital optical fiber communication links, the performance of an analog link is limited by the fiber attenuation and/or dispersion. Because of fiber attenuation the received signal is weakened and may not be discernible from noise. Because of fiber dispersion the transmission bandwidth $\sigma_f = 1/2\pi\sigma_\tau$ is limited so that high-frequency signal components are attenuated more than low-frequency components, resulting in signal degradation. Both of these deleterious effects increase with the increase of the fiber length L. The received optical power drops exponentially with L, whereas the fiber bandwidth is inversely proportional to L. Nonlinear effects do not play a role in analog systems since the power is distributed and not concentrated in narrow pulses.

The maximum allowable length of the analog fiber link is determined by ensuring that two conditions are met:

- The fiber attenuation must be sufficiently small so that the received power is greater than the receiver power sensitivity P_r.
- The fiber bandwidth σ_f must be greater than the spectral width B of the transmitted signal.

As discussed in Sec. 18.6, the sensitivity of an analog optical receiver is the smallest optical power necessary for the signal-to-noise ratio (SNR) of the photocurrent to exceed a prescribed value SNR_0. For an ideal receiver (with unity quantum efficiency and no circuit noise) $\mathrm{SNR} = \overline{n} = (P/h\nu)/2B$, where B is the receiver bandwidth, P the optical power (watts), and $\overline{n}$ the average number of photons received in a time interval $1/2B$, regarded as the resolution time of the system. If SNR_0 is the minimum allowed signal-to-noise ratio, the receiver sensitivity becomes $\overline{n}_0 = \mathrm{SNR}_0$ photons per resolution time and the corresponding power

$$P_r = h\nu\overline{n}_0(2B). \tag{24.2-14}$$

This is identical to the expression (24.2-1) for the power sensitivity of the digital receiver if the resolution time $1/2B$ of the analog system is equated with the bit time $1/B_0$ of the digital system.

Because of the equivalence between (24.2-14) and (24.2-1) and because of the applicability of the power budget equation (24.2-3) to analog systems as well, the L–B_0 relations determined earlier for the binary digital system are applicable to the analog system, with B_0 replaced by $2B$, provided that the acceptable performance of the analog system is $\mathrm{SNR}_0 = 10$. As an example, a 1-km fiber link capable of transmitting digital data at a rate of 2 Gb/s with a BER not exceeding 10^{-9} can also be used to transmit analog data of bandwidth 1 GHz with a signal-to-noise ratio of at least 10.

In analog systems, however, the required signal-to-noise ratio is usually much greater than 10, so that the receiver sensitivity must be much greater than 10 photons per resolution time. For high-quality audio and video signals, for example, a 60-dB signal-to-noise ratio is often required. This corresponds to $\mathrm{SNR}_0 = 10^6$, or $\overline{n}_0 =$

10^6 photons per resolution time. Additional design considerations are particularly important in analog systems. For example, the nonlinear response of the light source and photodetector cause additional signal degradation and place restrictions on the dynamic range of the transmitted waveforms.

D. Attenuation and Dispersion Compensation and Management

Attenuation Compensation

The performance of attenuation-limited fiber communication systems may be significantly enhanced by use of optical fiber amplifiers placed at appropriate distances within the fiber link, as illustrated in Fig. 24.2-14. Amplifiers elevate the diminished optical power, so that the received power remains above the receiver sensitivity for longer links. This process is ultimately limited by the noise introduced by the amplifiers themselves since, unlike repeaters, optical amplifiers do not regenerate the exact digital signal. However, before this limit is reached, dispersion often takes over and the system becomes dispersion limited. Dispersion compensation is therefore indispensable in long-haul optical fiber communication systems using optical amplifiers.

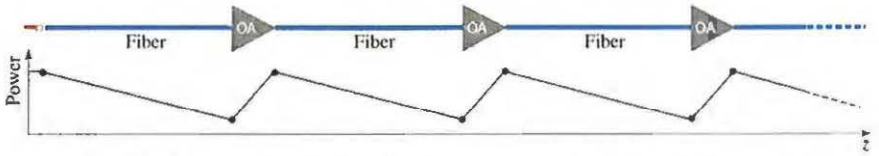


Figure 24.2-14 Compensation of attenuation by use of optical fiber amplifiers.

Dispersion Compensation

The pulse spreading introduced by propagation through an optical fiber of length L and dispersion coefficient D_λ may be reversed by use of another fiber, called a dispersion-compensating fiber, with dispersion coefficient D'_λ of opposite sign and length L' selected such that the magnitudes of the dispersion introduced by the two fibers are equal, i.e.,

$$D'_\lambda L' = -D_\lambda L. \tag{24.2-15}$$

The pulse spreading and compression introduced by an alternating sequence of such fibers is illustrated in Fig. 24.2-15. The compensating fiber is often relatively short and its dispersion coefficient must therefore be high. Since dispersion in conventional fibers is positive for wavelengths above 1310 nm, the dispersion compensating fiber must have negative dispersion in this band. This can be afforded by use of dispersion-shifted fibers (DSF).

Other optical components may be used in place of DCFs. As described in Sec. 22.2, the propagation of an optical pulse through a dispersive medium is equivalent to a quadratic chirp filter, which is a phase-only filter with a phase proportional to the square of the frequency. A fiber of length L and dispersion coefficient D_λ is a quadratic chirp filter with chirping coefficient $b = D_\lambda L$. The effect of this filter may be completely eliminated by use of an inverse compensation filter — another quadratic chirp filter with chirping coefficient of equal magnitude and opposite sign, $b' = -b$. The dispersion compensating fiber plays such a role, but other optical components, such as gratings and interferometers, may also be used (see Sec. 22.2).

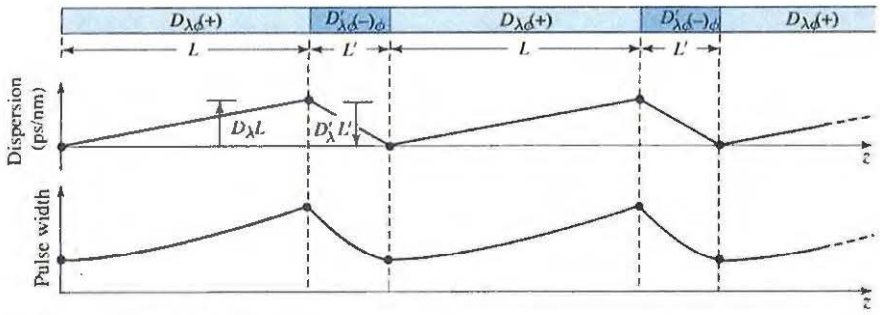


Figure 24.2-15 Dispersion compensation by use of fiber segments of opposite dispersion.

The compensation filter may be placed at the transmitter end of the link, thus precompensating the dispersion that is subsequently introduced by the fiber. Alternatively, it may be placed at the receiver end, thus postcompensating the broadened pulses immediately before they are detected. More commonly, multiple compensation filters are placed periodically within the link, providing distributed compensation. Under linear propagation conditions, the actual locations of the compensation filters is not important. However, in order to avoid the deleterious nonlinear effects, compensation filters are placed such that short pulses are avoided over extended distances within the fiber.

Broadband Dispersion Compensation: Dispersion Management

For broadband communication systems, such as WDM, the condition for dispersion compensation, (24.2-15), must be satisfied at all wavelengths within the spectral band; i.e., the error $e_\lambda = D'_\lambda L' - D_\lambda L$ must be zero everywhere. Since the dispersion coefficients are wavelength dependent, this condition is difficult to satisfy. Fig. 24.2-16 illustrates a situation for which $e_\lambda = 0$ at a wavelength λ_1 in the middle of the band, wherein the compensation is perfect, a positive e_λ at a wavelength λ_2 corresponding to a net positive dispersion, and a negative e_λ at another wavelength λ_3 with net negative dispersion.

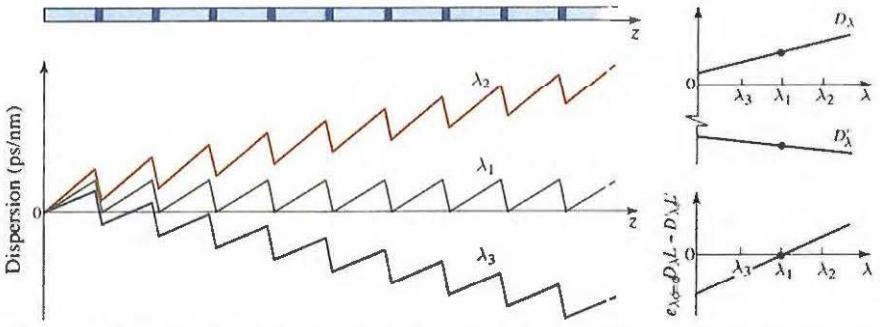


Figure 24.2-16 Perfect dispersion compensation at λ_1, and imperfect dispersion compensation with net positive and negative dispersion at λ_2 and λ_3, respectively. The error e_λ vanishes if the slopes of D_λ and D'_λ are equal.

If both D_λ and D'_λ are approximately linear functions of λ with the same slope, and if $e_\lambda = 0$ at the central wavelength λ_1, then $e_\lambda \approx 0$ everywhere. The design of a

dispersion compensation filter with the appropriate value of the dispersion coefficient and the slope of its wavelength dependence is known as **dispersion management**.

E. Soliton Optical Communications

The ultimate dispersion compensation occurs naturally in optical solitons. These non-spreading pulses have an intensity that is sufficiently high for the nonlinear optical properties of the fiber to play a principal role in their formation. As described in Sec. 22.5B, optical solitons are pulses for which nonlinear dispersion (the dependence of the phase velocity on the intensity via the optical Kerr effect) completely compensates linear chromatic dispersion — the net result is that the pulse travels without altering its width or shape. The gain provided by a fiber amplifier can also be used to compensate for fiber attenuation so that the pulse maintains its peak intensity and continues to travel as a soliton.

As expressed in (22.5-13), for a pulse of width τ_0, peak intensity I_0, and free-space wavelength λ_o at its central frequency, the condition for soliton formation is

$$\frac{2\pi}{\lambda_o} n_2 I_0 = \frac{-\beta''}{\tau_0^2}, \qquad \text{(24.2-16)}$$

Soliton Condition

where n_2 is the Kerr coefficient and $-\beta'' = (\lambda_o^2/2\pi c_o) D_\lambda$ is proportional to the dispersion coefficient D_λ. The intensity profile of the soliton is described by $I(t) = I_0 \operatorname{sech}^2(t/\tau_0)$, which is a bell-shaped function of FWHM $1.76\,\tau_0$.

In a digital optical communication system, a soliton of width τ_0 much smaller than the bit interval T represents bit "1," while bit "0" is represented by the absence of a soliton. This is necessarily a return-to-zero (RZ) modulation format.

Soliton communication systems are neither attenuation-limited nor dispersion-limited. Instead, they are limited by nonlinear intersymbol interference that occurs as a result of the nonlinear interaction between the tails of solitons representing neighboring bits. For example, when two identical solitons separated by the bit interval T travel a sufficiently long distance through the same fiber, they eventually collapse and merge into a single pulse, which subsequently separates back into the original two pulses. As mentioned in Sec. 22.5B, this process is repeated periodically with a period [see (22.5-32)]

$$L_p = \pi e^{r/2} z_0, \qquad \text{(24.2-17)}$$

where $r = T/\tau_0$ is the ratio of the separation to the soliton width and $2z_0 = -\tau_0^2/\beta'' = 2\pi c_o \tau_0^2/\lambda_o^2 D_\lambda$ is the fiber dispersion distance.

The period L_p increases exponentially with the ratio r. If $r \gg 1$, i.e., if the bit interval T is much greater than the soliton width τ_0, then L_p can be made much longer than z_0. If the fiber length L is much smaller than L_p, then the interaction between neighboring bits is minimal. For a fixed ratio r, the condition $L \ll L_p$ may be written in terms of the bit rate $B_0 = 1/T$ as

$$L B_0^2 \ll \frac{\pi^2 c_o}{\lambda_o^2 D_\lambda} \frac{e^{r/2}}{r^2}. \qquad \text{(24.2-18)}$$

This places a limit on the ultimate distances and transmission rates allowed.

EXAMPLE 24.2-1. ***Soliton Communication System.*** A soliton communication system transmits data at 10 Gb/s through a single-mode dispersion-shifted fiber at $\lambda_0 = 1550$ nm using

10-ps (FWHM) soliton pulses. At this wavelength, the dispersion coefficient $D_\lambda = 1$ ps/nm-km and the nonlinear coefficient $n_2 = 2.6 \times 10^{-20}$ m^2/W. The fiber effective cross-sectional area is $A_{\text{eff}} = 60\ \mu\text{m}^2$. We proceed to determine the source optical power and the maximum length of the link.

The 10-ps FWHM pulse width corresponds to a time constant $\tau_0 = 10/1.76 = 5.7$ ps. To satisfy the soliton condition (24.2-16), the peak intensity is $I_0 = 3.75 \times 10^8$ W/m^2, corresponding to a peak power $I_0 A_{\text{eff}} = 22.5$ mW, which must be delivered by the source. The fiber dispersion distance $2z_0 = (2\pi c_o/\lambda_o^2)\tau_0^2/D_\lambda \approx 25$ km.

Since the bit interval $T = 1/B_0 = 100$ ps, the ratio $r = T/\tau_0 = 17.6$, and the interaction period given in (24.2-17) is $L_p \approx 2.1 \times 10^4 z_0$. The fiber length must be much shorter than this length. In this example, (24.2-18) provides $LB_0^2 \ll 26$ (Tb/s)^2 km.

24.3 MODULATION AND MULTIPLEXING

A. Modulation

Optical communication systems are classified in accordance with the optical variable modulated by the transmitted signal. Two principal types are used: field modulation and intensity modulation.

Field modulation. The field of a monochromatic optical wave serves as a sinusoidal carrier of very high frequency (200 THz at $\lambda_o = 1500$ nm, for example). In amplitude modulation (AM), phase modulation (PM), and frequency modulation (FM) systems, the amplitude, phase, or frequency is varied in proportionality to the signal (Fig. 24.3-1). Because of the extremely high frequency of the optical carrier, a very wide spectral band is available, and large amounts of information can, in principle, be transmitted. Although modulation of the optical field is an obvious extension of conventional radio

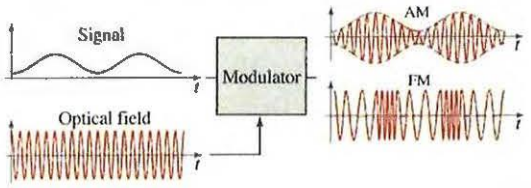


Figure 24.3-1 Amplitude and frequency modulation of the optical field.

and microwave communication systems to the optical band, it is rather difficult to implement, for several reasons:

- It requires a source whose amplitude, frequency, and phase are stable and free from fluctuations, i.e., a highly coherent laser.
- Direct modulation of the phase or frequency of the laser is usually difficult to implement. An external modulator using the electro-optic effect, for example, may be necessary.
- Because of the assumed high degree of coherence of the source, multimode fibers exhibit large modal noise; a single-mode fiber is therefore necessary.
- Unless a polarization-maintaining fiber is used, a mechanism for monitoring and controlling the polarization is needed.
- The receiver must be capable of measuring the magnitude and phase of the optical field. This is usually accomplished by use of a heterodyne detection system.

Because of the requirement of coherence, optical communication systems using field modulation are called **coherent communication systems**. These systems are discussed in Sec. 24.5.

Intensity modulation. In an intensity modulation system, the optical intensity (or power) is proportional to the signal, or a coded version thereof, as illustrated in Fig. 24.3-2). The majority of commercial fiber communication systems at present use intensity modulation. The power of the source is modulated by varying the injected current in an LED or a laser diode. The fiber may be single-mode or multimode and the optical power received is measured by use of a **direct-detection receiver**. The high-frequency optical field oscillations play no role in the modulation and demodulation; only *power* is varied at the transmitter and detected at the receiver. However, the wavelength of light may be used to identify different signals traveling through the same link — a process known as wavelengtb division multiplexing (WDM).

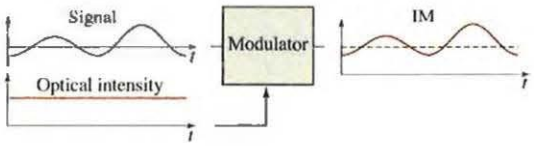


Figure 24.3-2 Intensity modulation.

Modulation format. Once the modulation variable is chosen (intensity, frequency, or phase), any of the conventional modulation formats (analog, pulse, or digital) can be used. An important example is **pulse code modulation** (PCM). In PCM the analog signal is sampled periodically at an appropriate rate and the samples are quantized to a discrete finite number of levels, each of which is binary coded and transmitted in the form of a sequence of binary bits, "1" and "0," represented by pulses transmitted within the time interval between two adjacent samples (Fig. 24.3-3).

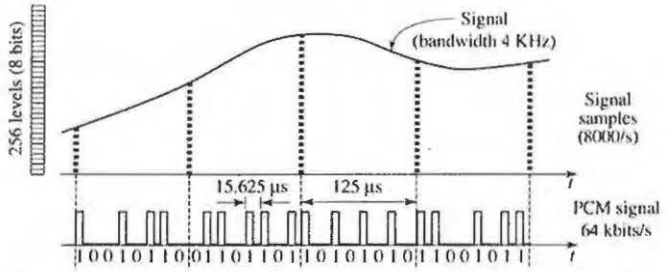


Figure 24.3-3 An example of PCM. A 4-kHz voice signal is sampled at a rate of 8×10^3 samples per second. Each sample is quantized to $2^8 = 256$ levels and represented by 8 bits, so that the signal is a sequence of bits transmitted at a rate of 64 kb/s.

If intensity modulation is adopted, each bit is represented by the presence or absence of a pulse of light. This type of modulation is called **ON–OFF keying** (OOK). For frequency or phase modulation, the bits are represented by two values of frequency or phase. The modulation is then known as **frequency shift keying** (FSK) or **phase shift keying** (PSK). These modulation schemes are illustrated in Fig. 24.3-4. It is also possible to modulate the intensity of light with a harmonic function serving as a subcarrier whose amplitude, frequency, or phase is modulated by the signal (in the AM, FM, PM, FSK, or PSK format).

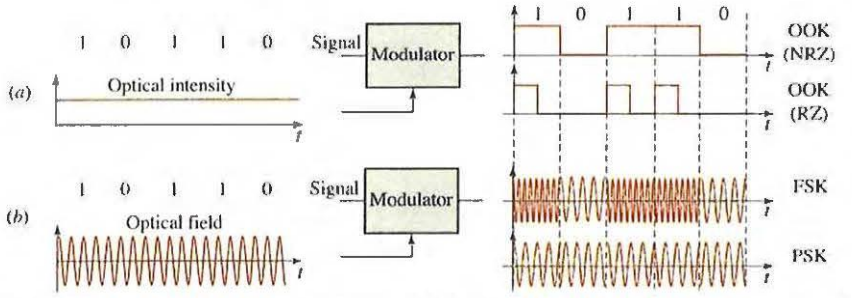


Figure 24.3-4 Examples of binary modulation of light: (*a*) ON–OFF keying intensity modulation (OOK/IM); (*b*) frequency shift keying (FSK) and phase shift keying (PSK) field modulation.

B. Multiplexing

Multiplexing enables the transmission and retrieval of more than one signal through the same communication link, as illustrated in Fig. 24.3-5. This is accomplished by marking each signal with a distinct physical label or a code that may be identified at the receiver.

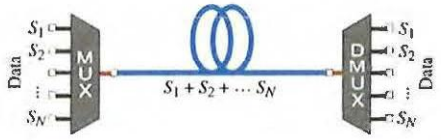


Figure 24.3-5 Transmission of N signals through the same channel by use of a multiplexer (MUX) and demultiplexer (DMUX).

There are three standard multiplexing systems: **frequency-division multiplexing, (FDM), time-division multiplexing (TDM),** and **code-division multiplexing (CDM).**

FDM. In FDM, carriers of distinct frequencies are modulated by the different signals. At the receiver, the signals are identified by the use of filters tuned to the carrier frequencies, as illustrated in Fig. 24.3-6(*a*).

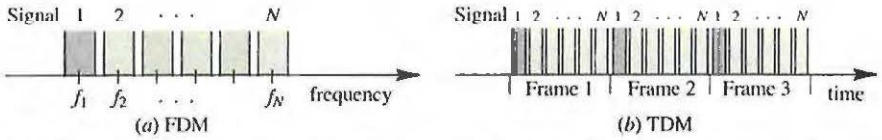


Figure 24.3-6 (*a*) In frequency-division multiplexing (FDM), a spectral band centered about a distinct frequency is allocated to each signal. (*b*) In time-division multiplexing, a sequence of time slots is allocated to each signal. The time slots of different signals are interleaved.

TDM. In TDM, data is transmitted in a sequence of time frames, each with a set of time slots allocated to bits or bytes of the different signals, as illustrated in Fig. 24.3-6(*b*). These bits must be synchronized to the same clock. At the receiver, each signal is identified by its location within the frame. An example of a hierarchical TDM system is the T-system illustrated in Fig. 24.3-7.

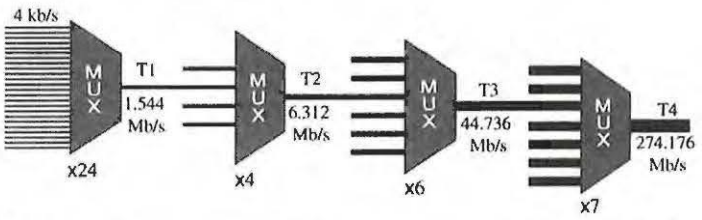


Figure 24.3-7 The T-system. A set of 24 4-kb/s signals are multiplexed by a TDM generating a T1 composite signal at 1.544 Mb/s. Four such signals are multiplexed to generate a T2 signal, and so on.

CDM. In CDM, each signal is assigned a code (or key) in the form of a unique function of time defined within the bit period. The code can be a sequence of one/zero bits at a much higher rate than that of the original data. Codes of different signals must be uncorrelated (orthogonal) so that they can be separated at the receiver by use of a correlator. In one encoding scheme, shown in Fig. 24.3-8, each bit "1" of the original data is replaced with the code sequence. Each receiver correlates its own code with the received signal. This locks it to only the bits associated with its own code, disregarding all other bits.

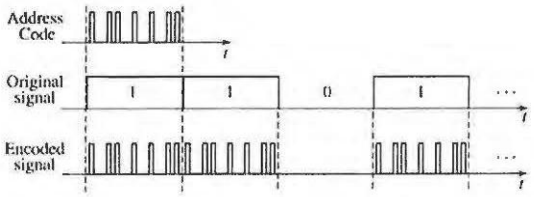


Figure 24.3-8 CDM encoding.

Multiplexing may be electronic or optical. In **electronic multiplexing**, the signals are multiplexed (FDM or TDM) to generate a composite electronic signal that is used to modulate the light source in any of the optical modulation schemes discussed in Sec. 24.3A. For example, an FDM electronic signal may be generated by use of a set of carrier frequencies, called "subcarriers," to modulate the intensity of the light source (IM modulation). At the receiver, the light is detected and the demultiplexing is accomplished by use of electronic filters. Another example is a TDM electronic signal, such as the T4 signal shown in Fig. 24.3-7, used to intensity modulate the light source; the demultiplexing of the detected signal is accomplished electronically.

In **optical multiplexing**, the labels distinguishing the multiplexed signals are optical in nature. For example, in optical FDM, different optical frequencies are used as the carriers of the various signals. These frequencies are separated at the receiver by use of optical filters.

When the frequencies of the carriers in optical FDM are widely spaced (say, greater than 20 GHz) this form of optical FDM has become known as **wavelength-division multiplexing** (WDM). WDM systems are popular since they can be used to expand the capacity of an existing fiber network without laying more fiber.

C. Wavelength-Division Multiplexing (WDM)

A WDM system uses light sources of different wavelengths, each intensity modulated by a different electrical signal. The modulated light beams are mixed into the fiber

using an optical multiplexer (OMUX). Demultiplexing is implemented at the receiver end by use of an optical demultiplexer (ODMUX), which separates the different wavelengths and directs them to different detectors. Optical multiplexers and demultiplexers are described in Sec. 23.2A. The electronic signal associated with each wavelength is often an electronically multiplexed set of other signals, and electronic demultiplexing is then necessary at the receiver end. The overall system is illustrated in Fig. 24.3-9.

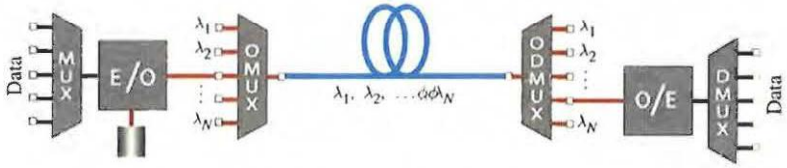


Figure 24.3-9 Wavelength division multiplexing (WDM).

The spectral bands used in modern optical fiber communication systems are shown in Fig. 24.3-10, along with the attenuation of silica-glass fibers. WDM systems use any combination of wavelengths within these bands. The spacing between the wavelengths of the different channels must be sufficiently greater than the spectral widths of the modulated light in each channel, which is determined by the linewidth of the light sources, and also by the spectral widths of the data carried by the channel. The channel spacing must also be sufficiently large to permit optical multiplexing and demultiplexing with minimal crosstalk.

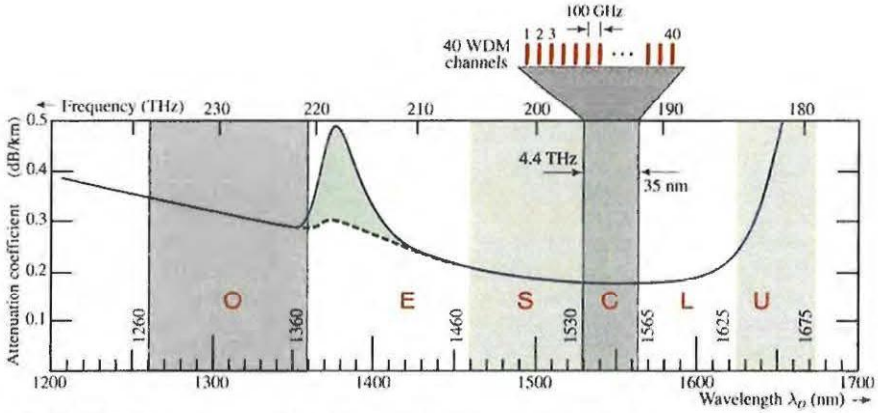


Figure 24.3-10 A 40-channel WDM system in the C spectral band, where fiber attenuation is minimal. The channel spacing is 100 GHz.

WDM systems are classified into two categories, coarse and dense, depending on the number of channels and the channel spacing.

Coarse WDM (CWDM) systems use a few channels with widely spaced wavelengths (20 nm or more). An example is a system with two wavelengths, one at 1310 nm and another at 1550 nm. CWDM is used in cable television networks, wherein different wavelengths are used for the downstream and upstream signals. The Ethernet LX-4 physical layer standard is another example in which four wavelengths near 1310 nm

are used, each carrying a 3.125 Gb/s data stream. Metropolitan networks use CWDM systems with a 20 nm wavelength spacing.

Dense WDM (DWDM) systems have a large number of channels (generally more than 8) with closely spaced wavelengths. At a wavelength of 1550 nm in the C band, a frequency spacing $\Delta\nu = 200$ GHz corresponds to a wavelength spacing $\Delta\lambda = (\lambda_o^2/c_o)\Delta\nu = 1.6$ nm. DWDM systems use channel spacings as small as 50 GHz or even 25 GHz, corresponding to wavelength spacings of 0.4 nm and 0.2 nm, respectively. As shown in Fig. 24.3-10, the width of the C band is 35 nm, or approximately 4.4 THz. This can accommodate 40 channels with a 100-GHz (0.8 nm) spacing. DWDM tends to be used at a higher level (and higher data rates) in the communications hierarchy, for example, on the Internet backbone. Design of DWDM systems is significantly more difficult than CWDM because the lasers need to be significantly more stable, and precision temperature control is often required to prevent wavelength drift.

24.4 FIBER-OPTIC NETWORKS

A communication network comprises a set of communication links connecting multiple users (terminals) distributed within some geographical area. Messages or data may be passed from one terminal to another by transmission through one or several links along paths controlled by routers and switches. A **local-area network (LAN)**, for example, connects terminals such as computers, printers, video monitors, or faxing and copying machines in a restricted region such as a building, a campus, or a manufacturing plant. Larger networks include the telephone network, the global Telex network, and the Internet. The network may use electrical cables, optical fibers, or satellite links. Fiber-optic networks use fiber-optic links together with electronic or optical routers and switches (see Chapter 23).

A. Network Topologies and Multiple Access

A network of N nodes may be constructed by use of a dedicated point-to-point link between each node and all other nodes. This requires $N(N-1)$ duplex (i.e., bidirectional) point-to-point links, and employs $2N(N-1)$ transmitters and $2N(N-1)$ receivers.

Topologies using fewer point-to-point links, and fewer transmitters and receivers, include the bus, the star, and the ring topologies, illustrated in Fig. 24.4-1, as well as a system for accessing the shared links. In these networks, only N transmitters and

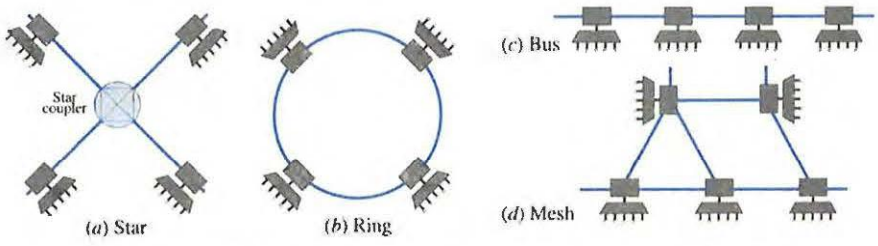


Figure 24.4-1 Network topologies: (*a*) star, (*b*) bus, (*c*) ring, and (*d*) mesh.

N receivers are necessary. In the star network, each node is connected to all other nodes via the star coupler at the center of the network; power transmitted by one node

is distributed equally among the other nodes. In the bus and ring networks, the fiber passes through the nodes, and data may be extracted from, or added to, the optical signal by any node. Since light transmitted by each node travels different distances to different nodes, the receivers must be able to process received power at various levels, i.e., must have a large dynamic range. A more general configuration is the mesh network.

Several networks of the same or different topologies are often connected to create a larger network, as illustrated by the example in Fig. 24.4-2.

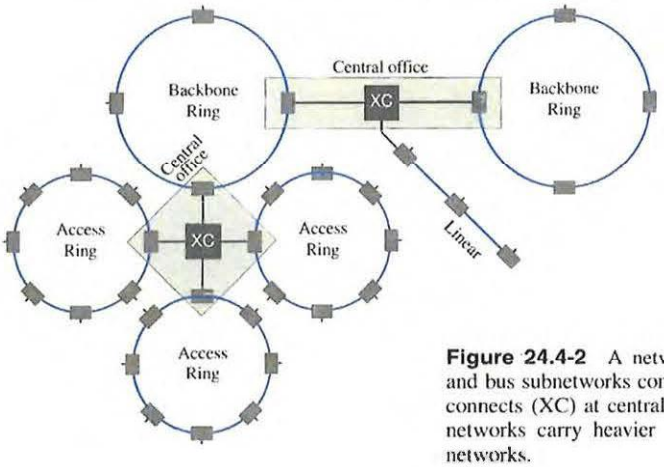


Figure 24.4-2 A network comprised of ring and bus subnetworks connected by digital cross-connects (XC) at central offices. Backbone ring networks carry heavier traffic and feed access networks.

Interface

The interface between the terminal and the fiber network at each node includes a receiver, a transmitter and an electronic add-drop multiplexer (ADM), as illustrated in Fig. 24.4-3(*a*). The receiver detects the optical signal, and the ADM extracts data and adds new data that modulate a source and transmit a new optical signal through another fiber. Such interface is said to be *opaque* since the light is detected and regenerated at the node.

A *transparent* interface is coupled to the fiber network optically, as illustrated in Fig. 24.4-3(*b*). Optical directional couplers are described in Sec. 23.1 and Sec. 23.3. An optical interface to a bidirectional (duplex) fiber uses two directional couplers to transmit and receive in either direction, as shown in Fig. 24.4-3(*c*).

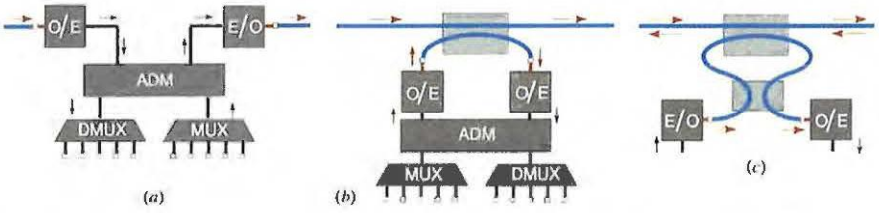


Figure 24.4-3 Interface between the node and the fiber network. (*a*) Opaque interface. The signal is converted from optical to electronic (O/E) and the ADM extracts data and adds new data, which is used to generate a new optical signal (E/O). (*b*) Optically coupled (transparent) interface using a directional coupler. (*c*) Optically coupled interface to a duplex fiber using two directional couplers.

Multiple Access

The signals transmitted by the network nodes share the same fiber (the **medium**). To avoid confusion, a scheme for **multiple access** or **medium access** is necessary. Time-domain, frequency-domain, and code-domain multiple access systems are in use:

- **Time-division multiple access (TDMA)** is similar to time-division multiplexing (TDM), which is used in conventional point-to-point communication systems (see Sec. 24.3B). The nodes send their data through the shared medium during interleaved time slots. Buffers may be used to store data until the appropriate time. Since it is not possible to synchronize the timing of all nodes, guard times separating consecutive slots are necessary.

- **Frequency-division multiple access (FDMA)** is similar to frequency-division multiplexing (FDM) (see Sec. 24.3B). Here, the nodes send their data through the shared medium in preassigned spectral bands, and there is no need to synchronize the bit clocks of the input signals. In optical networks, FDMA is called **wavelength-division multiple access (WDMA)** and is the counterpart to wavelength-division multiplexing (WDM).

- **Code-division multiple access (CDMA)** is similar to code-division multiplexing (CDM) (see Sec. 24.3B). In CDMA, each node is preassigned a unique address code. Data transmitted by a node is encoded with the address code of the destination node. Each node correlates its own address code with the incoming signal. This locks it to only the bits associated with its own address, disregarding all other bits. The data come in a sequence of packets each with the address of its destination (see Sec. 23.3F).

Synchronous Optical Network (SONET)

SONET [and its international version, the Synchronous Digital Hierarchy (SDH)] is a TDM standard for transmission over optical fibers. It addresses the difficulty of time-division multiplexing of signals with slightly different clock rates by embedding these signals within frames of longer duration. The payload (the signal bits) are allowed to float within the frames, but the frames are perfectly synchronous. SONET provides a hierarchy of multiplexed signals in which the basic unit, known as the STS-1 signal or the optical carrier-1 (**OC-1**), transports data at 51.84 Mb/s. Combining N such signals generates the **OC-N** signal, which has an N times greater rate, as listed in Table 24.4-1. For example, OC-192 and OC-768 operate at approximately 10 Gb/s and 40 Gb/s, respectively.

Table 24.4-1 Transmission rates (Mb/s) in the STS hierarchy used in the SONET network.

OC-1	OC-3	OC-12	OC-24	OC-48	OC-192	OC-768
51.84	155.52	622.08	1,244.16	2,488.32	9,995.33	39,813.12

EXAMPLE 24.4-1. ***Ring Network.***
An example of a fiber-optic 4-node ring network operating at different data rates is shown in Fig. 24.4-4. Each of the four nodes transmits data to the other three nodes at either the OC-12 ($\approx$ 622 Mb/s) or the OC-24 ($\approx$ 1.24 Gb/s) rate, as shown. The fiber segment connecting nodes 1 and 2 carries the heaviest traffic at a combined rate OC-12+OC-12+OC-24 = OC-48 ($\approx$ 2.5 Gb/s). The 2–3 and 3–4 segments carry lighter combined traffic at the OC-24 rate.

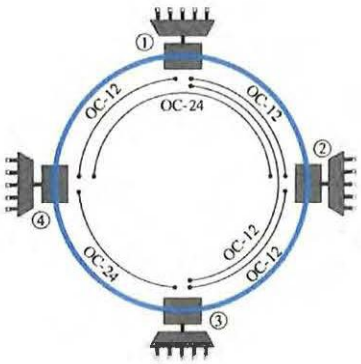


Figure 24.4-4 A 4-node ring network.

B. Wavelength-Division Multiplexing (WDM) Networks

A wavelength-division multiplexing (WDM) fiber-optic network uses coarse or dense WDM for communication along its links and WDMA for medium access. The nodes are connected in some topology (e.g., star, ring, bus, or mesh), and each node transmits into one or several wavelength channels and receives from one or several wavelength channels. The existence of multiple wavelength channels for each physical connection adds another dimension to the network and offers some flexibility, at the expense of some complexity.

Broadcast-and-Select WDM Network

The simplest WDM network is the **broadcast-and-select network**. Each node transmits at a unique fixed wavelength and *broadcasts* its transmission to all other nodes via a passive optical couplers. The receiver in each node *selects* the one wavelength addressed to it by use of a tunable filter. As an example, in the 5-node network shown in Fig. 24.4-5(a), nodes $1, 2, \cdots, 5$ transmit at wavelengths $\lambda_1, \lambda_2, \cdots, \lambda_5$, respectively. An optical star coupler broadcasts each transmission to all other nodes. In the shown state, for example, node 1 is tuned to channel λ_5; nodes 2, 3, and 4 are tuned to channel λ_1; and node 5 is tuned to channel λ_2. As illustrated in the equivalent connection diagram in Fig. 24.4-5(b), node 2 transmits to node 5, node 5 transmits to node 1, and node 1 multicasts its transmission to nodes 2, 3, and 4.

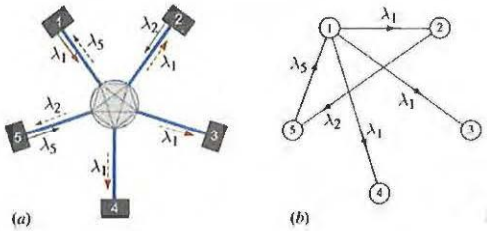


Figure 24.4-5 A WDM broadcast-and-select network (a) and its equivalent logical connections (b).

In another example, shown in Fig. 24.4-6(a), the receiver of each node is tuned to

the wavelength transmitted by its next neighbor. Thus, the network, which has a *star* physical topology, is equivalent to a *ring* logical topology, as illustrated in Fig. 24.4-6 (*b*).

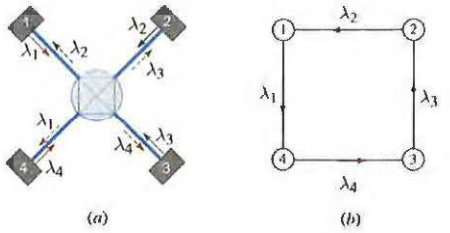


Figure 24.4-6 A WDM network in a star physical topology (*a*) is equivalent to the ring logical topology (*b*).

The network changes its state, i.e., the wavelengths to which each node is tuned, as desired. Dynamic coordination is required in order to avoid conflict and collisions.

Multi-Hop Broadcast-and-Select WDM Network

The requirement that each of the nodes of the broadcast-and-select network be capable of selectively detecting any of the wavelengths transmitted by the other nodes can be demanding. This requirement is alleviated in a multi-hop network, for which each node is allocated two different wavelength channels for transmission and only two different channels for reception. At any time, a node may transmit at one of its two allocated wavelengths and may receive by tuning to one its two allocated wavelengths. The channels are allocated to the nodes in such a way that a node may access any other node by following either a single-hop (i.e., direct) connection or a two-hop connection via an intermediate node. For example, in the network shown in Fig. 24.4-7(*a*), node 2 can transmit to node 1 directly via channel λ_3. Although node 1 cannot transmit to node 2 directly, since they share no common wavelength, this transmission may occur in two hops: node 1 transmits to node 3 on the λ_1 channel, and node 3 subsequently transmits to node 2 on the λ_6 channel. This configuration is therefore called the **multi-hop broadcast-and-select network**.

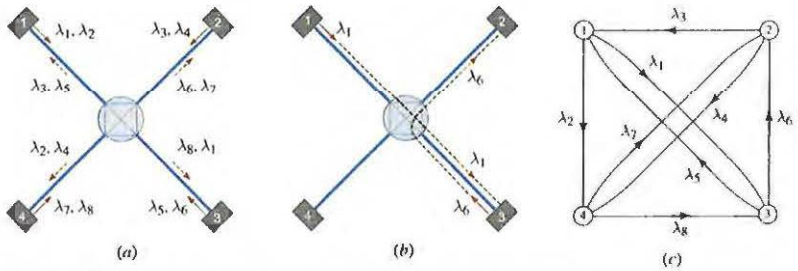


Figure 24.4-7 (*a*) A WDM multi-hop broadcast-and-select network. (*b*) A two-hop connection from node 1 to node 2 via node 3. (*c*) Logical topology of the network.

The broadcast-and-select configuration, single-hop or multi-hop, is not suitable for networks with large number of nodes. Since the power transmitted by each node must reach all other nodes, the system becomes inefficient for a large number of nodes. Also the number of channels used, which must equal or exceed the number of nodes, becomes prohibitive for large networks.

Wavelength-Routed Networks

In a **wavelength-routed network**, a pair of nodes communicates by use of one of the wavelength channels following some connection path. Another pair of nodes may use the same wavelength channel if their connection path does not share a common link with the path of the first pair. For example, in the network shown in Fig. 24.4-8(*a*), nodes 1 and 2 communicate on channel λ_1, and so do nodes 2 and 3. However, nodes 1 and 3 must use a different wavelength λ_2 if they use the path connecting them via node 2. Similarly, nodes 4 and 1 communicate via a third channel λ_3 since their path contains links that use the λ_1 and λ_2 channels.

In this network, each link carries one or more wavelengths (but not necessarily all of the wavelengths, as is the case in the broadcast-and-select network). For example, the link between nodes 4 and 5 carries traffic at three wavelength channels, but each of the other four links carry only two channels. Also, each node transmits and receives data at one or more wavelengths. For example, node 5 receives data from node 4 at λ_1 and from node 3 at λ_2; it transmits data to node 1 at λ_1; data carried by channel λ_3 pass through this node without being detected. The logical connections in this network are shown in Fig. 24.4-8(*b*).

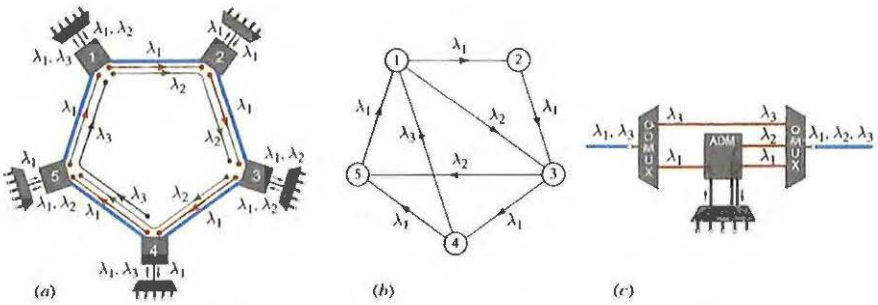


Figure 24.4-8 (*a*) A 5-node 3-channel wavelength-routed ring network. (*b*) Its logical topology. (*c*) An optical-add-drop multiplexer (OADM) used at node 5.

The key component in a wavelength-routed WDM network is the optical add-drop multiplexer (OADM) (see Sec. 23.2A). Each node has an OADM that extracts (drops) data from certain wavelength channels on the incoming fiber, adds data to certain channels on the outgoing fiber, and lets data on certain channels of the incoming fiber pass through without change to the outgoing fiber. An OADM is made of an optical demultiplexer (ODMUX), an add-drop multiplexer (ADM), and an optical multiplexer (OMUX). As an example, the OADM used in node 5 of the network in Fig. 24.4-8(*a*) is shown in Fig. 24.4-8(*c*). Agile networks use reconfigurable OADMs (denoted by the acronym ROADM).

Wavelength-routed networks with configurations other than the ring configuration have nodes with multiple incoming and outgoing fibers. At these nodes, more complex routers are necessary. For example, a node with two incoming and two outgoing fibers, as shown in Fig. 24.4-9, employs an **optical cross-connect (OXC)** that receives data from selected incoming fibers/channels, adds data to selected outgoing fibers/channels, and routes data on selected incoming channels to selected outgoing channels. The OXC uses multidimensional space-domain wavelength-domain switches and ADMs (see Sec. 23.3D). A wavelength-routed network also uses a hub node which uses a server to process data at all wavelength channels.

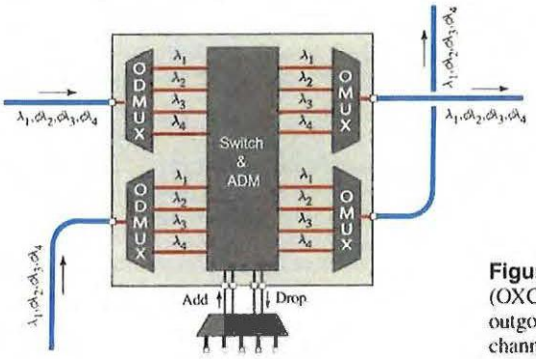


Figure 24.4-9 An optical cross-connect (OXC) at a node with two incoming and two outgoing fibers, each with four wavelength channels.

EXAMPLE 24.4-2. ***WDM Upgrade of a Ring. Network.***

A 4-node wavelength-routed WDM ring network operates at 3 channels of wavelength λ_1, λ_2, and λ_3 at the rates shown in Fig. 24.4-10. This network is an upgraded version of the network in Fig. 24.4-4. Nodes 1 and 3 access the wavelengths λ_1 and λ_2; node 4 accesses the wavelengths λ_1 and λ_3; and node 2 accesses all three wavelengths. In the upgraded network, the nodes communicate at twice the rates of the original network, but the highest rate in any of the WDM channels does not exceed that of the original network. The fiber segment connecting nodes 1 and 2 carries the heaviest traffic at a combined OC-96 rate ($\approx$ 5 Gb/s), but the highest rate at any given wavelength is OC-48 ($\approx$ 2.5 Gb/s).

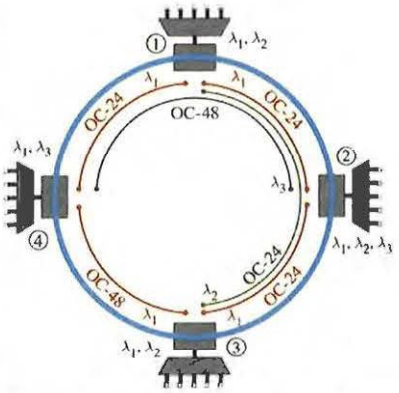


Figure 24.4-10 A schematic of a 4-node 3-channel WDM ring network.

24.5 COHERENT OPTICAL COMMUNICATIONS

Coherent optical communication systems use field modulation (amplitude, phase, or frequency) instead of intensity modulation. They employ coherent light sources, single-mode fibers, and heterodyne receivers. In this section we examine the principles of operation of these systems, determine their performance advantage, and briefly discuss the requirements on the components of the system.

Heterodyne and Homodyne Receivers

Photodetectors are responsive to the photon flux and, as such, are insensitive to the optical phase. It is possible, however, to measure the complex amplitude (both magnitude and phase) of the signal optical field by mixing it with a coherent reference optical field of stable phase, called the **local oscillator**, and detecting the superposition using a photodetector, as illustrated in Fig. 24.5-1. As a result of interference (beating)

between the two fields, the detected electric current contains information about both the amplitude and phase of the signal field.

This detection technique is called **optical heterodyning, optical mixing, photomixing, light beating** (see Sec. 2.6B), or **coherent optical detection** (as opposed to **direct detection**. The coherent optical receiver is the optical equivalent of a superheterodyne radio receiver. The signal and local-oscillator waves usually have different frequencies (ν_s and ν_L). When $\nu_s = \nu_L$ the detector is said to be a **homodyne detector**.

Let $\mathcal{E}_s = \text{Re}\{A_s \exp(j2\pi\nu_s t)\}$ be the signal optical field, with $A_s = |A_s| \exp(j\varphi_s)$ its complex amplitude and ν_s its frequency. The magnitude $|A_s|$ or the phase φ_s are modulated with the signal at a rate much slower than ν_s. The local oscillator field is described similarly by $\mathcal{E}_L$, A_L, ν_L, and φ_L. The two fields are mixed using a beamsplitter or an optical coupler, as illustrated in Fig. 24.5-1. If the incident fields are perfectly parallel plane waves and have precisely the same polarization, the spatial dependence may be suppressed and the total field is the sum of the two constituent fields $\mathcal{E} = \mathcal{E}_s + \mathcal{E}_L$. Taking the absolute square of the sum of the complex waves, we obtain

$$\begin{aligned} &|A_s \exp(j2\pi\nu_s t) + A_L \exp(j2\pi\nu_L t)|^2 \\ &\quad = |A_s|^2 + |A_L|^2 + 2|A_s||A_L| \cos[2\pi(\nu_s - \nu_L)t + (\varphi_s - \varphi_L)]. \end{aligned} \tag{24.5-1}$$

Since the intensities I_s, I_L, and I are proportional to the absolute-square values of the complex amplitudes,

$$I = I_s + I_L + 2\sqrt{I_s I_L} \cos[2\pi\nu_I t + (\varphi_s - \varphi_L)], \tag{24.5-2}$$

where $\nu_I = \nu_s - \nu_L$ is the difference frequency (also called the **intermediate frequency**).

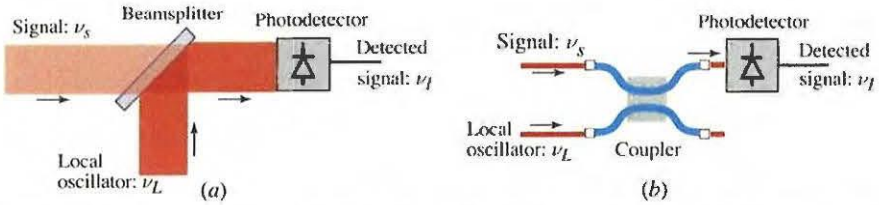


Figure 24.5-1 Optical heterodyne detection. A signal wave of frequency ν_s is mixed with a local oscillator wave of frequency ν_L using (*a*) a beamsplitter, and (*b*) an optical coupler. The photocurrent varies at the frequency difference $\nu_I = \nu_s - \nu_L$.

The optical power P collected by the photodetector is the product of the intensity and the detector area, so that

$$P = P_s + P_L + 2\sqrt{P_s P_L} \cos[2\pi\nu_I t + (\varphi_s - \varphi_L)], \tag{24.5-3}$$

where P_s and P_L are the powers of the signal and the local oscillator beams, respectively. The third term of (24.5-3) varies with time at the difference frequency ν_I with a phase $\varphi_s - \varphi_L$. If the signal and local oscillator beams are close in frequency, their difference ν_I can be many orders of magnitude smaller than the individual frequencies.

Misalignment between the directions of the two waves reduces or washes out the interference term [the third term of (24.5-3)], since the phase $\varphi_s - \varphi_L$ then varies

sinusoidally with position within the area of the detector. As is readily understood from Fig. 2.5-4, this can be avoided by keeping the angle θ between the wavefronts sufficiently small, such that $\theta \ll \lambda/a$ where a is the size of the photodetector aperture.

The photocurrent i generated in a semiconductor photon detector is proportional to the incident photon flux Φ (see Sec. 18.1B). When ν_I is much smaller than ν_s and ν_L, the superposed light is quasi-monochromatic and the total photon flux $\Phi = P/h\overline{\nu}$ is proportional to the optical power, where $\overline{\nu} = \frac{1}{2}(\nu_s + \nu_L)$. The mean photocurrent is therefore $\bar{i} = \eta e\Phi = (\eta e/h\overline{\nu})P$, where e is the electron charge and η the detector's quantum efficiency, so that

$$\bar{i} = \bar{i}_s + \bar{i}_L + 2\sqrt{\bar{i}_s\bar{i}_L}\cos\big[2\pi\nu_I t + (\varphi_s - \varphi_L)\big], \tag{24.5-4}$$

where $\bar{i}_s = \eta e P_s/h\overline{\nu}$ and $\bar{i}_L = \eta e P_L/h\overline{\nu}$ are the photocurrents generated by the signal and local oscillator individually. The local oscillator is usually much stronger than the signal, so that the first term in (24.5-4) is negligible and

$$\boxed{\bar{i} \approx \bar{i}_L + 2\sqrt{\bar{i}_s\bar{i}_L}\cos\big[2\pi\nu_I t + (\varphi_s - \varphi_L)\big].} \tag{24.5-5}$$

Photomixing Current

The time dependence of the detected current $\bar{i}$ is sketched in Fig. 24.5-2(a). The second term in (24.5-5), which oscillates at the difference frequency ν_I, carries the useful information. With knowledge of $\bar{i}_L$ and φ_L, the amplitude and phase of this term can be determined, and $\bar{i}_s$ and φ_s estimated, from which the intensity and phase (and hence the complex amplitude) of the measured optical signal can be inferred. The information-containing signal variables $\bar{i}_s$ or φ_s are usually slowly varying functions of time in comparison with the difference frequency ν_I, so they act as slow modulations of the amplitude and the phase of the harmonic function $2\sqrt{\bar{i}_L}\cos(2\pi\nu_I t - \varphi_L)$. This amplitude- and phase-modulated current can be demodulated by drawing on the conventional techniques used in AM and FM radio receivers.

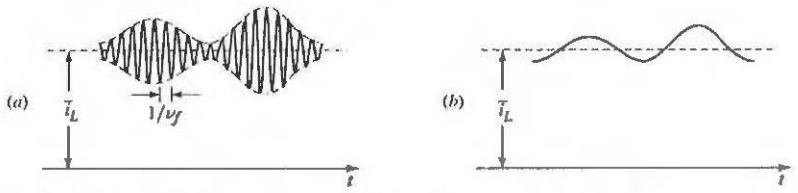


Figure 24.5-2 (a) Photocurrent generated by the heterodyne detector. The envelope and phase of the time-varying component carries complete information about the complex amplitude of the optical field representing the signal. (b) Photocurrent generated by the homodyne detector.

From a photon-optics point of view, this process can he understood in terms of the detection of polychromatic (two-frequency) photons (see Prob. 12.1-11).

The homodyne system is a special case of the heterodyne system for which $\nu_s = \nu_L$ and $\nu_I = 0$. The demodulation process is different. A phase-locked loop is used to lock the phase of the local oscillator so that $\varphi_L = 0$ and (24.5-5) yields

$$\bar{i} = \bar{i}_L + 2\sqrt{\bar{i}_s\bar{i}_L}\cos\varphi_s. \tag{24.5-6}$$

Amplitude and phase modulation is achieved by varying $\bar{i}_s$ and φ_s, respectively.

Advantages of Heterodyne/Homodyne Receivers

In comparison with the direct-detection receiver, the heterodyne receiver has the following advantages:

- It is capable of measuring the optical phase and frequency.
- It permits the use of wavelength-division multiplexing (WDM) with smaller channel spacing ($\approx$ 100 MHz). In conventional direct-detection systems the channel spacing is of the order of 100 GHz.
- It permits the use of electronic equalization to compensate for pulse broadening in the fiber. Pulse broadening is a result of the dephasing of the different wavelength/frequency components because of differences in group velocities. Since the receiver monitors the phase, this dephasing may be removed by proper electronic filtering.
- By use of a strong reference field, the heterodyne receiver has an inherent noiseless gain conversion factor that effectively amplifies the signal above the circuit noise level.
- It provides a 3-dB SNR advantage over even the noiseless direct-detection receiver, as shown in Sec. 24.5.
- It is insensitive to unwanted background light with which the local oscillator does not mix. Heterodyning is one of the few ways of attaining photon-noise-limited detection in the infrared, where background noise is so prevalent.

The cost of these advantages is an increase in the system's complexity since heterodyning requires a stable local oscillator, an optical coupler in which the mixed fields are precisely aligned, and complex circuits for phase locking.

Coherent Systems

An essential condition for the proper mixing of the local oscillator field and the received optical field is that they must be locked in phase, be parallel, and have the same polarization in order to permit interference to take place. This places stringent requirements on the two lasers and on the fiber. The lasers must be single-frequency and have minimal phase and intensity fluctuations. The local oscillator is phase-locked to the received optical field by means of a control system that adjusts the phase and frequency of the local oscillator adaptively (using a phase-locked loop). The fiber must be single-mode (to avoid modal noise). The fiber must also be polarization-maintaining, or the receiver must contain an adaptive polarization-compensation system.

A schematic diagram of a coherent optical fiber communication system using two lasers and phase modulation is shown in Fig. 24.5-3. The local oscillator field is mixed with the received optical field using an optical directional coupler. One branch of the coupler output contains the sum of the two optical fields and the other branch contains the difference. Using (24.5-4), the detected currents

$$\bar{i}_{\pm} = \bar{i}_s + \bar{i}_L \pm 2\sqrt{\bar{i}_s \bar{i}_L}\cos[2\pi\nu_I t + (\varphi_s - \varphi_L)] \tag{24.5-7}$$

are subtracted electronically, yielding $4\sqrt{\bar{i}_s \bar{i}_L}\cos[2\pi\nu_I t + (\varphi_s - \varphi_L)]$, which is demodulated to recover the message. This type of coherent receiver is known as a **balanced mixer**. It has the advantage of canceling out intensity fluctuations of the local oscillator.

A number of coherent optical fiber communication systems have been implemented at $\lambda_o = 1550$ nm (where fiber attenuation is minimal) with bit-rate–distance products matching theoretical expectations. One example is provided by a system operating at a bit rate $\approx$ 4 Gb/s. A DFB laser with a 15-MHz CW linewidth was directly modulated in an FSK signal format. The local oscillator was a tunable DBR laser (see Sec. 17.3C). This system exhibited a receiver sensitivity $\approx$ 190 photons/bit and was used for transmission over a 160-km length of fiber.

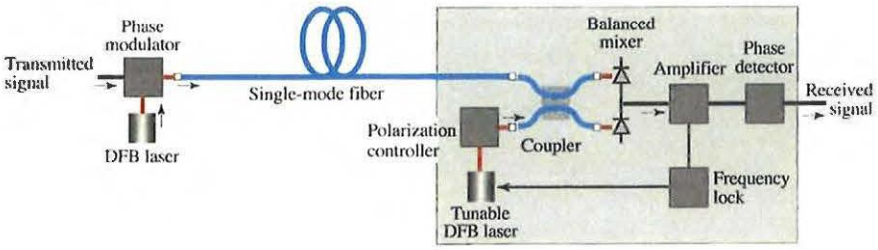


Figure 24.5-3 Coherent optical fiber communication system.

* *Performance of Analog Coherent Communication Systems*

Heterodyne detection is necessary whenever the phase of the optical field is to be measured. However, heterodyne detection can also be useful for measuring the optical intensity since it provides gain through the presence of the strong local oscillator. As such, it offers an alternative to both optical amplification (see Chapter 14 and Sec. 17.2) and APD amplification (see Sec. 18.4). This can provide a signal-to-noise ratio advantage over direct detection, as we show in this section.

The mean photocurrent $\bar{\imath}$ generated by a photodiode is accompanied by noise of variance

$$\sigma_i^2 = 2e\bar{\imath}B + \sigma_r^2, \tag{24.5-8}$$

where B is the receiver's bandwidth; the first term is due to photon noise and the second represents circuit noise (see Sec. 18.6). The intensity of the local oscillator can be made sufficiently large so that even if the signal is weak, the total current $\bar{\imath}$ is such that the circuit noise σ_r^2 is negligible in comparison with the photon noise $2e\bar{\imath}B$.

Assuming that $\bar{\imath}_L \gg \bar{\imath}_s$ and $2e\bar{\imath}_L B \gg \sigma_r^2$, we use (24.5-5) and approximate (24.5-8) by

$$\bar{\imath} \approx \imath_L + 2\sqrt{\bar{\imath}_s\bar{\imath}_L}\cos[2\pi\nu_I t + (\varphi_s - \varphi_L)] \tag{24.5-9a}$$

$$\sigma_i^2 \approx 2e\bar{\imath}_L B. \tag{24.5-9b}$$

In the case of amplitude modulation, the signal is represented by the RMS value of the sinusoidal waveform in (24.5-9a), with the phase ignored. The electrical signal power is therefore $\frac{1}{2}[2\sqrt{\bar{\imath}_s\bar{\imath}_L}]^2 = 2\bar{\imath}_s\bar{\imath}_L$ and the noise power is $\sigma_i^2 = 2e\imath_L B$, so that the power signal-to-noise ratio is

$$\text{SNR} = \frac{2\bar{\imath}_s\bar{\imath}_L}{2e\bar{\imath}_L B} = \frac{\bar{\imath}_s}{eB}. \tag{24.5-10}$$

If $\overline{m} = \bar{\imath}/2Be$ is the mean number of photoelectrons counted in the resolution time interval $T = 1/2B$, then

$$\boxed{\text{SNR} = 2\overline{m}.} \tag{24.5-11}$$

Signal-to-Noise Ratio
Heterodyne Receiver

In comparison, the SNR of the direct-detection photodiode receiver measuring the

same signal current $\bar{\imath}_s$ without the benefit of heterodyning is

$$\text{SNR} = \frac{\bar{\imath}_s^2}{2e\bar{\imath}_s B + \sigma_r^2} = \frac{\overline{m}^2}{\overline{m} + \sigma_q^2}, \tag{24.5-12}$$

where $\sigma_q^2 = (\sigma_r/2Be)^2$ is the circuit-noise parameter discussed in Sec. 18.6C. The principal advantage of the heterodyne system is now apparent. For strong light or low circuit noise ($\overline{m} \gg \sigma_q^2$), the direct-detection result is SNR $= \overline{m}$. The heterodyne receiver, which yields SNR $= 2\overline{m}$, offers a factor-of-2 improvement (3-dB advantage). But for weak light (or large circuit noise) the advantage can be even more substantial, since the heterodyne receiver has SNR $= 2\overline{m}$, whereas the SNR of the direct-detection receiver is reduced by circuit noise to SNR $= \overline{m}/(1 + \sigma_q^2/\overline{m})$.

The performance of a direct-detection *avalanche* photodiode receiver is also inferior to that of a heterodyne photodiode receiver. In accordance with (18.6-36), the SNR obtained when the APD gain is sufficiently large to overcome circuit noise is

$$\text{SNR} = \frac{\overline{m}}{F}, \tag{24.5-13}$$

where F is the APD excess noise factor ($F > 1$). Therefore, even a noiseless APD receiver ($F = 1$) is a factor of 2 inferior to the heterodyne receiver.

* Performance of Digital Coherent Communication Systems

In this section the performance and sensitivity of a digital coherent communication system are determined in the cases of amplitude and phase modulation.

ON–OFF keying (OOK) homodyne system. Consider an ON–OFF keying (OOK) system transmitting data at a rate B_0 bits/s and using a homodyne receiver. Bits "1" and "0" are represented by the presence and absence of the signal $\bar{\imath}_s$ during the bit time $T = 1/B_0$, respectively. Assuming that $\varphi_s = \varphi_L = 0$ and $\nu_I = \nu_s - \nu_L = 0$, the measured current has the following means and variances obtained from (24.5-9a) and (24.5-9b):

$$\begin{aligned} &\text{mean } \mu_1 \approx \bar{\imath}_L + 2\sqrt{\bar{\imath}_L \bar{\imath}_s}, \quad \text{variance } \sigma_1^2 \approx 2\bar{\imath}_L eB \quad \text{for bit "1"} \\ &\text{mean } \mu_0 \approx \bar{\imath}_L, \qquad\qquad\quad \text{variance } \sigma_0^2 \approx 2\bar{\imath}_L eB \quad \text{for bit "0."} \end{aligned} \tag{24.5-14}$$

The receiver bandwidth $B = B_0/2$ since the bit time $T = 1/B_0$ is the sampling time $1/2B$ for a signal of bandwidth B.

The performance of the binary communication system under the Gaussian approximation has been discussed in Sec. sec22-4. The bit error rate is given by (18.6-55), where

$$Q = \frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0} = \sqrt{\frac{\bar{\imath}_s}{2eB}} = \sqrt{\overline{m}}, \tag{24.5-15}$$

and $\overline{m} = \bar{\imath}_s/2eB$ is the mean number of detected photoelectrons in bit 1. For a bit error rate BER $= 10^{-9}$, $Q \approx 6$ and therefore $\overline{m} = 36$, corresponding to a receiver sensitivity $\overline{m}_a = \frac{1}{2}\overline{m} = 18$ photoelectrons per bit (averaged over both bits).

Table 24.5-1 Receiver sensitivity for different receivers and modulation systems under ideal conditions (photons per bit).

	Direct Detection	Homodyne	Heterodyne
OOK	10	18	36
PSK	—	9	18
FSK	—	—	36

Phase-shift-keying (PSK) homodyne system. Here, bits "1" and "0" are represented by a phase shift $\varphi_s = 0$ and π, respectively. Assuming that $\varphi_L = 0$, the means and variances of the photocurrent for bits "1" and "0" are, from (24.5-9),

$$\text{mean } \mu_1 = \bar{i}_L + 2\sqrt{\bar{i}_L \bar{i}_s}, \quad \text{variance } \sigma_1^2 = 2e\bar{i}_L B \quad \text{for bit "1"}$$

$$\text{mean } u_0 = \bar{i}_L - 2\sqrt{\bar{i}_L \bar{i}_s}, \quad \text{variance } \sigma_0^2 = 2e\bar{i}_L B \quad \text{for bit "0"}$$

and therefore

$$Q = \frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0} = 2\sqrt{\frac{\bar{i}_s}{2eB}} = 2\sqrt{\overline{m}}. \tag{24.5-16}$$

For a BER $= 10^{-9}, Q = 6$, from which $\overline{m} = 9$. Since *each* of the two bits must carry an average of nine photoelectrons in this case, the average number of photoelectrons per bit is $\overline{m}_a = \overline{m} = 9$. It follows that the receiver sensitivity is 9 photoelectrons/bit. The PSK homodyne receiver is twice as sensitive as the OOK homodyne receiver because it requires half the number of photoelectrons.

Comparison. The sensitivity of the heterodyne digital receiver can be determined by following a similar analysis. Table 24.5-1 lists the receiver sensitivities of several digital modulation systems, assuming $\eta = 1$. Although it appears that the direct-detection OOK system has about the same performance as the best coherent system (homodyne PSK), in practice this is not so. In the homodyne system, circuit noise is overcome, whereas in the direct-detection system, circuit noise cannot be ignored, unless an APD is used. When an APD is used in a direct-detection receiver, circuit noise is overcome, but the APD gain noise raises the receiver sensitivity from 10 to at least $10F$, where F is the excess-noise factor. Direct-detection systems would have performance comparable to coherent-detection systems if a perfect APD with $F = 1$ (no excess noise) were available.

READING LIST

Books on Optical Fiber Communications

See also the reading lists in Chapters 8, 9, 17, 18, 22, and 23.

G. P. Agrawal, *Nonlinear Fiber Optics*, Academic Press, 1991, 4th ed. 2006.

S. Ramachandran, *Fiber-Based Dispersion Compensation*, Springer-Verlag, 2006.

J. N. Damask, *Polarization Optics in Telecommunications*, Springer-Verlag, 2005.

K.-P. Ho, *Phase-Modulated Optical Communication Systems*, Springer-Verlag, 2005.

J. C. Palais, *Fiber Optic Communications*, Prentice Hall, 5th ed. 2005.

A. Galtarossa and C. R. Menyuk, eds., *Polarization Mode Dispersion*, Springer-Verlag, 2005.

C.-F. Lin, *Optical Components for Communications: Principles and Applications*, Springer-Verlag, 2004.

H. Kolimbiris, *Fiber Optics Communications*, Prentice Hall, 2004.

T. Schneider, *Nonlinear Optics in Telecommunications*, Springer-Verlag, 2004.

M. Cvijetic, *Optical Transmission Systems Engineering*, Artech, 2004.

G. P. Agrawal, *Fiber-Optic Communication Systems*, Wiley, 3rd ed. 2002.

I. P. Kaminow and T. Li, eds., *Optical Fiber Telecommunications IVA: Components*, Academic Press, 2002.

I. P. Kaminow and T. Li, eds., *Optical Fiber Telecommunications IVB: Systems and Impairments*, Academic Press, 2002.

C. DeCusatis, ed., *Fiber Optic Data Communication: Technological Trends and Advances*, Academic Press, 2002.

C. DeCusatis, *Handbook of Fiber Optic Data Communication*, Academic Press, 2nd ed. 2002.

R. L. Freeman, *Fiber-Optics Systems for Telecommunications*, Wiley, 2002.

E. W. Van Stryland and M. Bass, eds., *Fiber Optics Handbook: Fiber, Devices, and Systems for Optical Communications*, McGraw–Hill, 2002.

N. Grote and H. Venghaus, eds., *Fibre Optic Communication Devices*, Springer-Verlag, 2001.

D. K. Mynbaev and L. L. Scheiner, *Fiber-Optic Communications Technology*, Prentice Hall, 2001.

G. Mahlke and P. Gössing, *Fiber Optic Cables: Fundamentals, Cable Design, System Planning*, Wiley–VCH, 4th revised and enlarged ed. 2001.

G. Keiser, *Optical Fiber Communications*, McGraw–Hill, 3rd ed. 2000.

R. Sabella and P. Lugli, *High Speed Optical Communications*, Kluwer, 1999.

G. Guekos, ed., *Photonic Devices for Telecommunications: How to Model and Measure*, Springer-Verlag, 1999.

H. J. R. Dutton, *Understanding Optical Communications*, Prentice Hall, 1999.

G. Lachs, *Fiber Optic Communications: Systems, Analysis, and Enhancements*, McGraw–Hill, 1998.

I. P. Kaminow and T. L. Koch, eds., *Optical Fiber Telecommunications IIIA: Components*, Academic Press, 1997.

I. P. Kaminow and T. L. Koch, eds., *Optical Fiber Telecommunications IIIB: Systems and Impairments*, Academic Press, 1997.

S. B. Alexander, *Optical Communication Receiver Design*, SPIE Optical Engineering Press, 1997.

L. Kazovsky, S. Benedetto, and A. E. Willner, *Optical Fiber Communication Systems*, Artech, 1996.

R. M. Gagliardi and S. Karp, *Optical Communications*, Wiley, 1976, 2nd ed. 1995.

S. Ryu, *Coherent Lightwave Communication Systems*, Artech, 1995.

S. Shimada, ed., *Coherent Lightwave Communications Technology*, Chapman & Hall, 1995.

J. M. Senior, *Optical Fiber Communications: Principles and Practice*, Prentice Hall, 2nd ed. 1992.

J. E. Midwinter, *Optical Fibers for Transmission*, Wiley, 1979; Krieger, reissued 1992.

S. E. Miller and I. P. Kaminow, eds., *Optical Fiber Telecommunications II*, Academic Press, 1988.

C. K. Kao, *Optical Fibre*, Institution of Electrical Engineers, 1988.

S. D. Personick, *Fiber Optics: Technology and Applications*, Plenum, 1985.

C. K. Kao, *Optical Fiber Systems*, McGraw–Hill, 1982.

S. E. Miller and A. G. Chynoweth, eds., *Optical Fiber Telecommunications*, Academic Press, 1979.

B. E. A. Saleh, *Photoelectron Statistics with Applications to Spectroscopy and Optical Communication*, Springer-Verlag, 1978.

Books on Fiber-Optic Networks

P. R. Prucnal, ed., *Optical Code Division Multiple Access: Fundamentals and Applications*, Taylor & Francis/CRC Press, 2006.

B. Mukherjee, *Optical WDM Networks*, Springer-Verlag, 2006.

E. Desurvire, *Wiley Survival Guide in Global Telecommunications: Broadband Access, Optical Components and Networks, and Cryptography*, Wiley, 2004.

E. Desurvire, *Wiley Survival Guide in Global Telecommunications: Signaling Principles, Protocols, and Wireless Systems*, Wiley, 2004.

G. Bernstein, B. Rajagopalan, and D. Saha, *Optical Network Control: Architecture, Protocols, and Standards*, Addison–Wesley, 2004.

D. Greenfield, *The Essential Guide to Optical Networks*, Prentice Hall, 2002.

R. J. Bates, *Optical Switching and Networking Handbook*, McGraw–Hill, 2001.

A. Jukan, *QoS-Based Wavelength Routing in Multi-Service WDM Networks*, Springer-Verlag, 2001.

R. Ramaswami and K. N. Sivarajan, *Optical Networks: A Practical Perspective*, Morgan Kaufmann, 2nd. ed. 2002.

M. T. Fatehi and M. Wilson, *Optical Networking with WDM*, McGraw–Hill, 2002.

C. D. Chaffee, *Building the Global Fiber Optics Superhighway*, Plenum, 2001.

T. E. Stern and K. Bala, *Multiwavelength Optical Networks: A Layered Approach*, Prentice Hall, 1999.

A. Bononi, *Optical Networking*, Springer-Verlag, 1999.

P. E. Green, *Fiber Optic Networks*, Prentice Hall, 1993.

Articles

Issue on optical communications, *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 12, no. 4, 2006.

Issue on optical communications, *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 10, no. 2, 2004.

Millennium issue, *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 6, no. 6, 2000.

D. L. Begley, ed., *Selected Papers on Free-Space Laser Communications II*, SPIE Optical Engineering Press (Milestone Series Volume 100), 1994.

E. G. Rawson, ed., *Selected Papers on Fiber Optic Local Area Networks*, SPIE Optical Engineering Press (Milestone Series Volume 91), 1994.

L. D. Hutcheson and S. C. Mettler, eds., *Selected Papers on Fiber Optic Communications*, SPIE Optical Engineering Press (Milestone Series Volume 88), 1993.

S. F. Jacobs, Optical Heterodyne (Coherent) Detection, *American Journal of Physics*, vol. 56, pp. 235–245, 1988.

P. R. Prucnal, M. A. Santoro, and T. R. Fan, Spread Spectrum Fiber Optic Local Area Network Using Optical Processing, *IEEE Journal of Lightwave Technology*, vol. LT-4, pp. 547–554, 1986.

M. C. Teich, Laser Heterodyning, *Optica Acta (Journal of Modern Optics)*, vol. 32, pp. 1015–1021, 1985.

PROBLEMS

24.1-1 **Optical Fiber Systems.** Discuss the validity of each of the following statements and indicate the conditions under which your conclusion is applicable.

(a) The wavelength $\lambda_o = 1300$ nm is preferred to $\lambda_o = 870$ nm for all optical fiber communication systems.

(b) The wavelength $\lambda_o = 1550$ nm is preferred to $\lambda_o = 1300$ nm for all optical fiber communication systems.

(c) Single-mode fibers are superior to multimode fibers because they have lower attenuation coefficients.

(d) There is no pulse spreading at $\lambda_o \approx 1312$ nm in silica-glass fibers.

(e) Compound semiconductor devices are required for optical fiber communication systems.

(f) APDs are noisier than p–i–n photodiodes and are therefore not useful for optical fiber systems.

24.1-2 **Components for Optical Fiber Systems.** The design of an optical fiber communication system involves many choices of sources, fibers, amplifiers, and detectors, some of which are shown in Fig. 24.2-3. Make appropriate choices for each of the applications listed below. More than one answer may be correct. Some choices, however, may be incompatible.

(a) A transoceanic cable carrying data at a 2.5 Gb/s rate with 100-km repeater spacings.

(b) A 1-m cable transmitting analog data from a sensor at 1 kHz.

(c) A link for a computer local-area network operating at 500 Mb/s.

(d) A 1-km data link operating at 100 Mb/s with ±50°C temperature variations.

24.2-1 **Performance of a Plastic Fiber Link.** A short-distance low-data-rate communication system uses a plastic fiber with attenuation coefficient 0.5 dB/m, an LED generating 1 mW at a wavelength of 870 nm, and a photodiode with receiver sensitivity −20 dBm. Assuming a power loss of 3 dB each at the input and output couplers, determine the maximum length of the link. Assume that the data rate is sufficiently low so that dispersion effects play no role.

24.2-2 **Maximum Length of Attenuation-Limited System.** An optical fiber communication link is designed for operation at 10 Mb/s. The source is a 100-μW LED operating at 870 nm and the fiber has an attenuation coefficient of 3.5 dB/km. The fiber is made of 1-km segments and connectors between segments have a loss of 1 dB each. Input and output couplers each introduce a loss of 2 dB. The safety margin is 6 dB. Two receivers are available, a Si p–i–n photodiode receiver with sensitivity 5000 photons per bit, and a Si APD with sensitivity 125 photons per bit. Determine the receiver sensitivity p_r (dBm units) and the maximum length of the link for each receiver.

24.2-3 **Maximum Data Rate of Attenuation-Limited System.** A 50-km optical fiber link is operated at a wavelength of 1550 nm. The source is a 2-mW InGaAsP laser and the fiber has attenuation coefficient of 0.2 dB/km. Connectors and couplers introduce a total loss of 8 dB and the safety margin is 6 dB. The receiver is an InGaAs APD with a sensitivity of 1000 photons per bit for a bit error rate of 10^{-9}. Determine the maximum data rate that can be used assuming an attenuation-limited system. If the required error rate is 10^{-11}, what is the maximum data rate?

24.2-4 **Maximum Length of an Analog Link.** An optical fiber communication link uses intensity modulation to transmit data at a bandwidth $B = 10$ MHz with a signal-to-noise ratio of 40 dB. The source is a $\lambda_o = 870$ nm light-emitting diode producing 100 μW of average power with a maximum modulation index of 0.5. The fiber is a multimode step-index fiber with an attenuation coefficient of 2.5 dB/km. The detector is an avalanche photodiode with mean gain $\overline{G} = 100$, excess noise factor $F = 5$, and responsivity 0.5 A/W (excluding the gain). Assuming that the circuit noise is negligible, use the theory presented in Sec. 18.6D to calculate the optical power sensitivity of the receiver and the attenuation-limited maximum length L of the fiber.

24.2-5 **Time Budget for Dispersion-Limited System.** A 100-km single-mode fiber link operating at a wavelength of 1550 nm. The source is an InGaAsP laser diode of spectral width 0.2 nm and response time 20 ps. The fiber has a dispersion coefficient 17 ps/km-nm. The receiver uses an InGaAs APD and has a response time of 0.1 ns. Determine the maximum data rate based on the criterion that the response time of the fiber does not exceed 25 % of the bit time. Also, determine the maximum data rate using the criterion that the response time of the overall system does not exceed 70 % of the bit time. If the a dispersion-shifted fiber is used instead, so that the dispersion coefficient is reduced to 1 ps/km-nm, what are the maximum data rates under the two criteria?

24.3-1 **Number of WDM Channels.** How many WDM channels fit in the C band (1530–1565 nm) and in the O band (1260–1360 nm) if the channel spacing is 75 GHz?

24.3-2 **Number of Nodes in a Broadcast-and-Select WDM Network.** The maximum number of nodes N that can be used by a broadcast-and-select WDM network is often limited by the available optical power. Determine N for a local area network using an optical star coupler connected to each of the nodes by a fiber of 2-km length, 0.3 dB/km attenuation coefficient, and 1dB of connector loss. The star coupler distributes the power equally among its outputs and introduces an additional loss of 3 dB. Each node uses a 1-mW optical source, the receiver sensitivity is -35 dBm, and a 5-dB safety margin is assumed.

24.3-3 **Wavelength-Routed WDM Ring Network.** Consider a 4-node 6-channel WDM network. Each node uses an add-drop multiplexer to transmit or receive at any of three different wavelengths assigned to it, but passes through the other three wavelengths. For example, node 1 may add or drop data at channels λ_1, λ_2, or λ_3, but passes through data at λ_4, λ_5, and λ_6. Allocate sets of three add-drop channels to each of the nodes 2, 3, and 4, such that any node on the ring may communicate with any of the other nodes. The idea is that each node must have one add-drop channel common with each of the other three nodes, but this channel must *not* be common with nodes in-between.