H CIENCE OF PHOTOTHERAPY: AN INTRODUCTION

by

Leonard I. Grossweiner

Springer

The Science of Phototherapy: An Introduction

by

Leonard I. Grossweiner†
Illinois Institute of Technology,
Chicago, IL, U.S.A.

edited by

Linda R. Jones
College of Charleston,
Charleston, SC, U.S.A.

co-authors

James B. Grossweiner
Edward Hospital,
Naperville, IL, U.S.A.

and

B.H. Gerald Rogers
University of Chicago Pritzker School of Medicine,
Chicago, IL, U.S.A.

Springer

TABLE OF CONTENTS

Foreword vii

Preface ix

Acknowledgements xiii

In Memoriam: Leonard I. Grossweiner xv

Chapter 1
An Overview of Phototherapy 1

Chapter 2
Optical Physics and Biotechnology 9

Chapter 3
Phototherapy light sources 57

Chapter 4
Quantum Description of Light Interactions with Matter 93

Chapter 5
Tissue Optics 143

Chapter 6
Photochemical Damage to Biological Systems 171

Chapter 7
Optical Methods of Imaging and Diagnosis 197

Chapter 8
Light Dosimetry Modeling for Phototherapy 211

Chapter 9
Laser Interactions with Tissues 225

Chapter 10
Photodynamic Therapy: Science and Technology 243

Chapter 11
Photodynamic Therapy: Clinical Aspects 275

Chapter 12
Phototherapy of Skin Disease: Science and Technology 299

Chapter 13
Phototherapy of Neonatal Jaundice 329

Glossary 337

Index 367

CHAPTER 1

AN OVERVIEW OF PHOTOTHERAPY

The beneficial effects of sunlight are a ubiquitous concept in folk medicine originating from sun worship. Indian medical literature dating to 1500 BC describes a treatment of non-pigmented areas of skin by applying black seeds of the plant Bavachee or Vasuchika followed by exposure of the skin to sunlight. This plant was identified as *Psoralea corylifolia* which contains photosensitizing psoralen compounds. Similar references appear in Buddhist literature from about 200 AD and in 10th century Chinese documents from the Sung period. The clearing of psoriasis by exposing the skin to sunlight has been known for at least a century. Sunlight also has deleterious biological effects. Exposure of horses, sheep, and cattle to sunlight after ingestion of certain plants leads to ulceration, and infection, and in severe cases, convulsions and death. In some regions of Arabia white horses were not used and even non-pigmented patches on colored horses were darkened with henna or tobacco juice. In the Tarentine fields of Italy only black sheep were grown because the white ones grew ill after exposure to sunlight. The connection between plants of the genus *Hypericum* and the sunlight sensitivity of grazing animals was made in the 19th century, if not earlier. The active photosensitizer was identified with the red pigment hypericin which is found in some insects and plants including St. John's wort. The cell killing properties of ultraviolet light are well known. The use of germidical lamps for this purpose is receiving new consideration in connection with the increasing incidence of contaminated food products. The possible link between skin cancer in humans and sunlight exposure had been suspected for many years based on statistical correlations between skin cancer incidence and the average intensity of sunlight. Recent basic research provides essentially incontrovertible evidence that frequent sunlight exposure generates DNA lesions which can initiate skin cancer. This history provides ample evidence that light can act as a powerful drug, either alone or in combination with chemical agents. The objective in a phototherapy is to utilize the beneficial action of light for treatment of disease while minimizing the harmful effects.

The study of light occupies the central position in the history of science (Table 1.1). Virtually every discovery cited in the table plays a direct or indirect role in the delivery or mechanism of phototherapy. The central principle of photochemistry, first stated by Grotthus and Draper in 1818, is "only absorbed

Table 1.1 Chronology of Discoveries about Light

In the beginning God created the heaven and the earth. Now the earth was unformed and void, and darkness was upon the face of the deep; and the spirit of God appeared over the face of the waters. And God said: "Let there be light", and there was light. *Genesis 1*

Atomistic theory of nature; *430 BC*, Democritus of Abdera
Discovery of fluorescence; *1577*, Nicholas Monardes
Discovery of phosphorescence; *1603*, Vincenzo Cascariolo
Laws of light refraction; *1637*, Rene Descartes
Explanation of the rainbow; *1665*, Isaac Newton
Discovery of double refraction in calcite; *1670*, Erasmus Bartholin
Measurement of the speed of light. *1676*, Øle Roemer
Wave theory of light; *1690*, Christiaan Huygens
Discovery of diffuse light; *1760*, Johann Lambert
Discovery of infrared radiation; *1800*, William Herschel
Discovery of ultraviolet radiation; *1801*, Johann Ritter
Discovery of absorption and emission spectra; *1802*, William Wollaston
Three color theory of vision; *1802*, Thomas Young and Ludwig von Helmholtz
The first artificial light source; *1826*, Humphry Davy
Invention of photography; *1826*, Joseph Niépce
Rotation of light polarization by a magnetic field; *1845*, Michael Faraday
Theory of electromagnetic radiation; *1865*, James Clerk Maxwell
Theory of light scattering; *1871*, Lord Rayleigh
Discovery of the photoelectric effect; *1887*, Heinrich Hertz
Invention of the optical interferometer; *1881*, Albert Michelson
First controlled phototherapy; *1893*, Niels Finsen
Quantum theory of black body radiation; *1900*, Max Planck
Discovery of biological photosensitization; *1900*, Oscar Raab
Quantum theory of light; *1905*, Albert Einstein
Quantum theory of the hydrogen atom; *1913*, Niels Bohr
Wave mechanics; *1924-1925*, L. de Broglie, W, Heisenberg, E. Schroedinger and others
Theory of the chemical bond; *1927*; W. Heitler, F. London, R. S. Mulliken, F. Hund
Theory of radiative transfer; *1950*, S. Chandresekhar
Invention of the laser; *1960*, Theodore Maiman

light can produce a chemical change". This deceptively simple concept is the basis of all photobiology. Every process in photobiology has the common feature that the first light-activated step involves very fast events at the molecular level.

The photophysical act of light absorption initiates a sequence of actions and reactions that can lead to a remarkable diversity of physiological endpoints, for example, plant growth, animal vision, circadian rhythms, and sunburn. The unique feature of phototherapy is that light acts as a powerful drug. This property of light differs from the historic view that light is a benign entity that is required for vision. However, it is now well known that excess exposure to sunlight is responsible for many disorders of the skin including a risk of skin cancer. The objective in phototherapy is to utilize light to cure disease while minimizing the adverse effects. All interactions of light with biological systems utilize unique light-absorbing molecular units or *chromophores* located within the illuminated tissues. In natural photobiology the chromophores are normally present in the tissue matrix, for example, chlorophyll molecules in green plants and retinal pigments in vertebrate retina. The tissue structure is an important determinant of the overall reaction pathway. The matrix may modify the "quality" of the incident light, establish structural relationships between chromophores and reacting molecules and store intermediate products. The photosynthetic reaction center in a plant leaf performs all of these functions. Evolution has optimized the structural conditions required for efficient utilization of sunlight. In phototherapy, however, the chromophores are exogenous chemical agents or metabolic products. The optimal treatment parameters must be determined by experimentation and clinical experience. The usual research sequence starts with photochemical studies on the relevant biological molecules, followed by *in vitro* cell cultures, experimental animal models, and clinical trials on human patients. This protracted and frequently disjointed program may be expedited by utilizing the methods of tissue optics. Recent developments in this field have made it possible to model key aspects of phototherapy by using a combination of analytical techniques and experiments with *tissue phantoms*, which are inert systems that mimic the optical properties of biological tissues.

The era of controlled phototherapy started at the turn of the 20th century. Niels Finsen, the father of modern phototherapy, discovered that natural sunlight and ultraviolet radiation from a carbon arc are beneficial treatments for cutaneous tuberculosis. Finsen received the Nobel Prize in medicine for his pioneering work in 1903. The slow acceptance of phototherapy in the U.S. may have resulted from a residual distrust of a discredited blue light therapy for arthritis and other afflictions proposed A. J. Pleasonton, a Civil War general. Regulatory approvals of clinical phototherapy started in the 1960s. Although the currently approved procedures are relatively few in number, each of them is

directed to a major affliction that affects a large patient population. Clinical reports demonstrating the anti-psoriatic action of mid-range ultraviolet (UV-B) from mercury-arc lamps first appeared in the 1920s. W. H. Goeckerman found that the effectiveness of the mercury arc was enhanced by first applying and then washing off crude coal tar from the psoriatic skin. The Goeckerman regimen has been a standard phototherapy for severe plaque-type psoriasis for many decades. Remarkably, this combination of two carcinogenic agents, coal tar and UV-B, is not known to induce skin cancer in humans. The modern *photochemotherapy* of psoriasis originated from the ancient literature. This procedure is based on the combined action of an oral psoralen drug and UV-A ("black light"). It is referred to as *PUVA*. The extension of ancient sunlight treatments for vitiligo using topical plant extracts to modern PUVA therapy of psoriasis was an exceptionally insightful accomplishment. Basic research on psoralen photochemistry established the molecular action mechanism for PUVA even before the clinical procedure was developed. The dimensions of linear psoralen derivatives span the pyrimidine bases across the DNA double helix. Absorption of UV-A by the localized psoralen molecules generates covalent cross-links across the DNA chains that block DNA synthesis. PUVA exemplifies a *photosensitization* in which light absorbed by a "colored" agent present in small amounts activates physical, chemical, or biological effects in a "colorless" substrate. *Extracorporeal photophoresis* (ECP) is an extension of the PUVA concept which is being used to treat cutaneous T-cell lymphoma and inhibit acute rejection of heart transplants. ECP is performed by irradiating a small amount of blood taken from the patient with UV-A after administration of the psoralen drug and then re-infusing the treated blood. The photochemically altered white cells initiate immunological responses, the details of which are the subject of active research.

Photodynamic therapy (PDT) is a light treatment for localized tumors utilizing the combined action of light and a chemical drug targeted to malignant tissues. In general, a *photodynamic action* is defined as a biological photo-sensitization that requires the participation of molecular oxygen. PDT was first attempted in 1903 by H. von Tappenier and A. Jesionek. who treated skin cancers with eosin dye followed by exposure of the lesions to sunlight. The lack of success probably resulted from an insufficient dye concentration in the tumors and an inadequate light dose. The first effective PDT drug was synthesized by R.L. Lipson in 1960 by treating impure hematoporphyin with strong acids. The resultant brown powder termed *hematoporphyrin derivative* (HPD) is a mixture of porphyrins, including a unique active component that localizes in and is retained by tumors after intravenous injection. This active component termed *dihematoporphyrin ether* (DHE) can be identified by a reddish fluorescence excited by blue or near-ultraviolet light. HPD was used sparingly for tumor identification and a few

cancer treatments in the 1960s and 1970s. In 1978 T. J. Dougherty and co-workers reported controlled clinical trials indicating that HPD is effective for eradication of skin cancers. The procedure was first referred to as "photoradiation therapy" which was then changed to "photodynamic therapy" to avoid confusion with cancer treatments that use ionizing radiation. Subsequent studies worldwide showed that HPD is effective for PDT of other localized cancers, including tumors of the head and neck, esophagus, urinary bladder, and bronchus. PDT has received regulatory approvals in 11 countries for specific malignancies, including esophageal cancer and early-stage bronchogenic lung cancer in the U.S. The putative action mechanism in PDT involves the generation of *singlet molecular oxygen* ("singlet oxygen") by the light-activated HPD and its attack on endothelial membranes of the tumor microvasculature. The shutdown of the tumor oxygen supply induces necrosis and killing of cancer cells. Recent research indicates that the vascular pathway is augmented by direct killing of cancer cells by *apoptosis*. This process is a form of regulated cell death that is normally involved in tissue hemostasis, and regulation of the immune system. The promising results with HPD motivated a concerted search for new and possibly more effective PDT drugs. Much attention has been given to the topical PDT agent δ-aminolevulinic acid (ALA) which is converted to photosensitizing protoporphyrin after uptake by cancer cells. The ability of PDT to inactivate abnormal endothelial cells has led to experimental extensions of the procedure to light treatments for non-cancer disorders, including Barrett's esophagus, endometriosis, and atherosclerotic vascular disease. Inhibition of age-related macular degeneration is a recently approved PDT application in the U.S.

The reduction of jaundice in newborn infants by exposure to blue-light was discovered by accident. In 1957 a nursing sister in the clinic of R. J. Cremer in Essex U.K. reported that neonatal jaundice in the sunlit parts of the nursery faded after a short period of exposure. A significant fraction of healthy term newborns now receive the blue-light phototherapy. Jaundice originates from the accumulation of unbound bilirubin in the infant's serum, which is a normal breakdown product of heme generated by turnover of red blood cells. Non-pathologic hyperbilirubinemia results from delayed development of liver enzymes that conjugate bilirubin with albumin and permit its elimination. The action mechanism in this phototherapy involves the photochemical conversions of bilirubin to a different molecular shape that is more readily eliminated. The process of *photoisomerization* is relatively uncommon in natural photobiology, except in animal vision in which it is the primary photochemical reaction of the retinal pigments.

The recent developments in phototherapy, and especially PDT, motivated far-reaching interest in relevant pre-clinical areas. The rapidly-developing field of

tissue optics exemplifies this phenomenon. Crystalline and amorphous substances, such as pure liquids, glasses, and crystals, are optically homogeneous materials in which light propagation is accurately described by the laws of geometrical and physical optics Classical optics is inapplicable for most biological tissues which are "cloudy" or *turbid* substances. Tissues contain microscopic entities that scatter light rays into many different directions. This effect is responsible for much of the attenuation of an entering light beam, in addition to the loss resulting from absorption by chromophores. The most general approach to light propagation in turbid media was developed by S. Chandrasekhar in the 1930s and 1940s. This theory termed *radiative transfer* was originally developed to explain light propagation between stars and planets, and subsequently extended to such diverse applications as diffusion of neutrons in nuclear reactors, atmospheric scattering of radar, underwater acoustics, and most recently, to tissue optics. The basic objective in tissue optics is to calculate the distribution of optical energy within tissue for specified irradiation conditions. The complex mathematics of the original theory has been augmented by new digital computer techniques. Tissue optics has been used for modeling PDT light dosimetry since about 1980. Optical imaging is a more recent application of radiative transfer directed towards non-invasive diagnosis of tissue abnormalities. The resolution achieved in optical imaging has improved considerably since the mid-1990s by utilization of tomographic techniques and image reconstruction algorithms borrowed from x-ray computed tomography (CT) and magnetic resonance imaging (MRI). *Monte Carlo* simulation is an alternative and very powerful analytical technique based on statistical methods. Another imaging approach utilizes the *in situ* fluorescence of a localized PDT drug excited by external light. This technique is referred to *photodynamic diagnosis* (PDD). The light sources used for phototherapy are either strong conventional lamps or relatively low- power lasers. The apparently different modality of laser surgery requires high-power lasers in order to coagulate, vaporize, and ablate tissues. However, the initial stages in laser surgery and phototherapy are very similar. In both modalities incident light must penetrate tissue and be absorbed by the relevant chromophores. The subsequent events in phototherapy involve photochemical reactions, while in laser surgery the absorbed energy is converted to heat or photomechanical effects. Recent analytical modeling techniques combine tissue optics with heat transfer theory. The applications include modeling of cancer therapy by hyperthermia and laser treatments of port wine stain and other vascular lesions.

Radiation dosimetry is an essential aspect of modern radiotherapy. In conjunction with new imaging techniques, improvements in radiation dosimetry are responsible for the improved efficacy of radiotherapy since the 1950s. It is

generally accepted that clinicians utilizing these techniques should have a basic understanding of the underlying science and a detailed knowledge of the relevant technology. The same criteria are applicable for phototherapy. The apparatus used for phototherapy is an integral part of each procedure. Testing of a new PDT drug frequently requires the development of a new light source and delivery system. Basic information about the properties of light sources, optical fibers, and electro-optical devices can assist the phototherapy clinician in the planning and delivery of the procedure. The most up-to-date information is more likely to appear in commercial literature than in general textbooks. In principle, light dosimetry for phototherapy parallels radiation dosimetry for radiotherapy. However, the phototherapy case is more complicated owing to the weaker penetration of light compared to x-rays and gamma-rays and the greater variability of tissue optical properties. The problems are further complicated by the use of *photometric units* in the photography and lighting industries and *radiometric units* in photobiology and photomedicine. The recent advances in tissue optics and optical imaging should provide a good basis for eventual rational treatment planning based on accurate light dosimetry. Except for the light treatment of neonatal jaundice, phototherapy is not widely used at present. PUVA has an established "niche" in dermatology. The recent use of ECP to prevent acute rejection of organ transplants will almost certainly expand during the next decade. The anti-cancer applications of PDT have increased steadily since 1980. Major tasks for the future will be to determine the most advantageous PDT procedure for individual patients, and evaluate the role of PDT as a treatment of first choice instead of a last resort when all other modalities have failed. It is interesting to note that cancer treatments with x-rays and radium started almost immediately after their discovery some 100 years ago. However, the major advances in radiotherapy were made after 1950 with the availability of new radiation sources, quantitative radiation dosimetry, and advances in the knowledge of radiation damage at the cellular level. PDT is entering a comparable phase of rapid growth driven by patient interest, increased receptiveness of the medical community, and significant investments of resources by the private sector. The term "PDT" is now being used to describe virtually any phototherapy utilizing an exogenous photosensitizing agent, whether or not photodynamic action has been demonstrated. The extension of PDT to the treatment of pre-cancerous and non-cancer diseases is a recent development. Inhibition of "wet" age-related macular degeneration by PDT is a recent application which has received regulatory approvals. ALA is approved in the US for PDT of actinic keratosis. The clinical potential of PDT has engendered a significant "spin-off" in related fields. Basic research on the PDT action mechanism led to new information about cellular and tissue responses to

photosensitization. Photosensitizing agents which may not be useful for PDT are being tested for other applications, including photochemical detoxification of viral and bacterial contamination of stored blood products and photosensitization of oral pathogenic dental bacteria. PDT motivated the synthesis of a new class of "pseudo-porphyrin" photosensitizing compounds, the metallo-texaphyrins. Lutetium texaphyrin is an effective drug activated by tissue-penetrating near-infrared light. The gadolinium texaphyrin derivative shows promise as an anoxic sensitizer for radiotherapy of brain tumors, with the added feature that it localizes in tumors and can be visualized by MRI. Recent advances in tissue optics were motivated by applications to modeling of PDT light dosimetry and laser surgery.

CHAPTER 2

OPTICAL PHYSICS AND BIOTECHNOLOGY

This chapter reviews the underlying optical physics relevant for photo-therapy, including the properties of light, light sources, and the devices used to deliver, control, and measure light.

2.1 LIGHT WAVES

Light is a wave motion. In general, a wave is an oscillatory disturbance moving through space. The nature of the disturbance depends on the type of wave. The disturbance in a mechanical wave is a transient dislocation of atoms in a medium. For example, a hydrodynamic surface wave is generated by dropping a pebble into water. This motion is *transverse* because the direction of the disturbance is perpendicular to the propagation direction of the wave. The disturbance in a sound wave consists of alternating regions of high and low pressure. Sound waves are *longitudinal* because the pressure variations oscillate in the same direction as the wave motion. A *standing wave* is a wave motion in which stationary regions or nodes are fixed in space. The displacement of a violin string exemplifies transverse standing waves. Sound waves in an organ pipe are longitudinal standing waves. As the name implies, the disturbance carried by a *continuous wave* (CW) is repetitive over a relatively long time period. An ordinary light bulb emits a mixture of CW light waves. The mathematical technique of *Fourier analysis* shows that virtually any wave can be decomposed into a group of infinitely long sinusoidal waves.[1] A pure musical tone generated by a tuning fork corresponds to an approximately single-wavelength or *monochromatic* sound wave. A musical chord consists of a mixture of pure tones. An ideal monochromatic wave is described by its amplitude, frequency, and wavelength (Figure 2.1). The amplitude is the maximum strength of the disturbance. The corresponding power carried by a wave is proportional to the square of the amplitude. The wave frequency (f) is the rate of oscillation at any point along the motion. The wavelength (λ) is the shortest distance between maximum amplitudes. The parameters of an idealized sinusoidal wave are related by

$$v=\lambda f \tag{2.1}$$

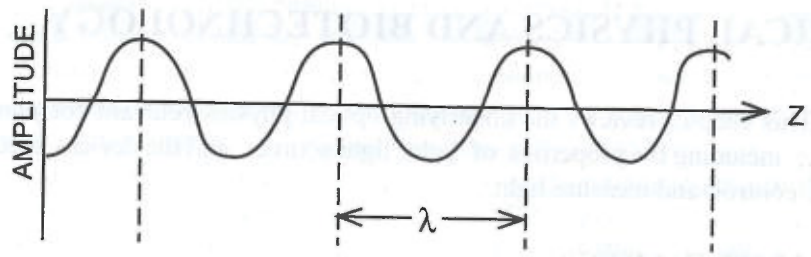


Figure 2.1 An ideal monochromatic wave is depicted as a sinusoidal disturbance. In a "running" wave the peaks and valleys propagate along the *Z*-axis with the wave velocity *V*. In a "standing" wave the points of zero amplitude or "nodes" are fixed in space and the peaks and valleys oscillate with the wave frequency *f*.

where v is the wave velocity. The *bandwidth* specifies the range of frequencies or wavelengths in a more general type of wave.

Light waves are *electromagnetic waves* (EMW). Other EMW include radio waves, microwaves, x-rays, and gamma-rays. The different forms of EMW are characterized by their range of wavelengths or frequencies. An EMW transports an electric field (E-field) and a magnetic field (B-field) at the speed of light (c) which equals 3.00×10^8 m/s in free space (Figure 2.2). The speed of an EMW in a medium depends on the type of wave and the properties of the material. The radiant power in an EMW is proportional to the square of the E-field (or B-field) amplitude. Typical optical power levels range from microwatts for small sources to gigawatts for powerful lasers.[2]

The wavelength (or frequency) is the principal determinant of how an EMW interacts with matter. The wavelength of visible light ranges from 400 nm (blue light) to 760 nm (far-red light). White light refers to a mixture of visible wavelengths that is perceived by the human eye as "white". The human eye can distinguish the order of ten million visible colors formed by adding the primary colors, blue, green, and red, in different ratios. A substance appears colored in white light because it either reflects or absorbs visible wavelengths. For example, a plant leaf is green in reflected sunlight because chlorophyll absorbs blue light and red light.

Radio waves have much longer wavelengths than light waves. Heating of biological tissues by radio waves is caused by interactions of the oscillating E-field with the dipolar water molecules in tissues.[3] Rapid back and forth rotations of the water molecules generates friction that heats the material. Similar

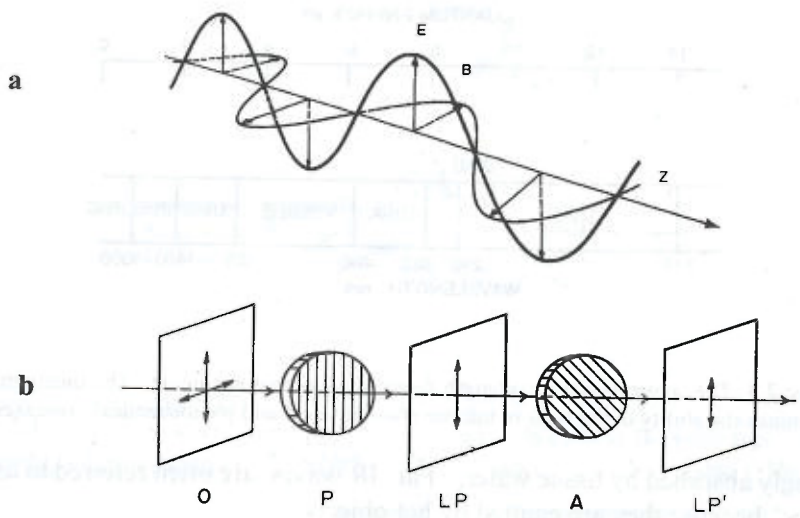


Figure 2.2 (a) The **E**-vector and **B**-vector of an electromagnetic wave are mutually perpendicular in a plane transverse to the direction of wave motions and oscillate with the wave frequency *f*. (b) The **E**-field of non-polarized light has a random orientation, as depicted by perpendicular projections of equal amplitude (O). Linear polarized light (LP) is transmitted by a polarizer (P). A second polarizer or "analyzer" (A) transmits 0-100% of the incident light depending on its relative angle to the polarizer.

Dielectric interactions are involved in tissue heating by microwaves. Gamma-rays are very short- wavelength EMW. Gamma-rays interact with the atomic electrons in biological matter. This process may lead to ejection of electrons (*photoelectric effect*). Optical radiations having shorter wavelengths than visible light are termed *ultraviolet* (UV) and those optical radiations having longer wavelengths than visible light are *infrared* (IR). The UV and IR bands are divided into sub-bands which are indicative of their interactions with biological macromolecules (Figure 2.3). UV-C or "far-UV" (200 nm - 290 nm) is strongly absorbed by all cellular constituents including proteins, nucleic acids, carbohydrates, and lipids. This band includes "germicidal" radiations strongly absorbed by DNA. UV-B or "mid-range UV" (290 nm - 320 nm) is absorbed mostly by aromatic amino acids and nucleic acids. The responses of skin to UV-B include vitamin-D_3 production, erythema, hyper-pigmentation, and skin cancer induction. UV-A or "black light" (320 nm - 400 nm) is weakly absorbed by "colorless" cellular constituents. Similarly, the IR spectrum is divided into sub-bands. IR-A or "near-IR" (0.76 μm - 4 mm) is an optical "window" with minimal attenuation in tissues. IR-B or "mid-range IR" (1.4 mm - 3 mm) is absorbed mostly by water and bone. IR-C or "far-IR" (3 mm - 1000 mm) is very

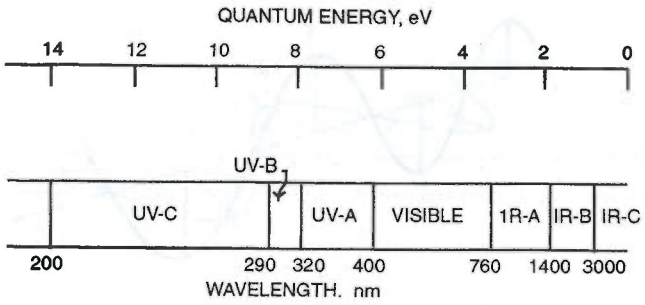


Figure 2.3 The conventional wavelength bands used in photomedicine. The quantum energy determines the ability of the light to initiate photophysical and photochemical processes.

strongly absorbed by tissue water. Far- IR waves are often referred to as "heat waves" because they are emitted by hot objects.

In addition to wavelength, frequency, and speed, EMW are characterized by the property of *polarization*. The EMW theory of Maxwell shows that E-field and B-field are located in a plane perpendicular to the direction of the wave motion or "ray" and are oriented perpendicular to each other. The vectors representing these fields oscillate with the wave frequency *f*. The orientation of the E-field in space defines the polarization. Ordinary lamps emit *non-polarized light* in which the time-averaged E-field has no preferred direction. The E-field of *linearly polarized light* maintains a fixed direction in space as the wave propagates. Linear polarization affects light absorption by long-chain molecules. This effect is utilized by Polaroid® to convert non-polarized light to linearly polarized light. This phenomenon is illustrated in Figure 2.4. The "polarizer" transmits preferentially in the direction determined by the orientation of the molecules in the film. The light intensity transmitted by the "analyzer" varies from almost zero to 100% depending on the relative orientation of the two films. In *circularly polarized light* the E-vector rotates with frequency *f* in its plane. Molecules having regions of spiral structure such as proteins and DNA interact differently with right-handed and left-handed circularly polarized light.

2.2 LIGHT QUANTA

The wave picture provides satisfactory explanations for many light properties, including mirror reflection, imaging by a lens, and light scattering by small particles. The interaction of light with a transparent object has a minute effect on the object itself. For example, a glass prism separates white light into

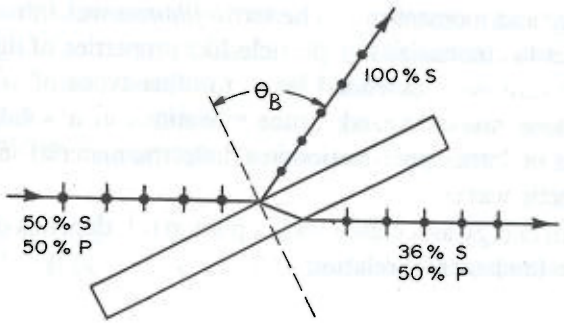


Figure 2.4 Polarization of ordinary light by a glass plate tilted at the Brewster angle. From Grossweiner, L.I., *The Science of Photobiology*, 2nd ed., Plenum Press, New York, 1989. With permission.

its constituent colors and the only effect on the prism is an extremely weak *radiation pressure* induced by reflection of light at the air-glass interface. However, light absorption may involve large interactive effects. The key role of light absorption in photobiology is expressed by extending the Grotthus and Draper central principle: *Only absorbed light can produce a physical, chemical, or physiological change in a biological system.* A proper description of light absorption and its consequences requires a quantum theory based on the molecular structure of matter.

Historically, the spectrum of thermal radiation emitted by hot objects was the first indication that classical electromagnetic theory led to drastically incorrect results. All matter emits radiation generated by the thermal motions of the constituent atoms. The red color emitted by the heating element on an electric stove is a familiar example. The classical theory of radiation predicts that thermal radiation should be increasingly intense into the UV region. However, measurements show that the thermal radiation spectrum of extra-terrestrial sunlight increases from low intensity in the IR region to a maximum near 550 nm (green light) and then decreases into the UV region. The 1901 radiation theory of Planck removed the "ultraviolet catastrophe" by discarding the classical concept that thermal radiation has a continuous spectrum of frequencies. Planck's revolutionary assumption was that thermal radiation is *quantized* in energy packets, each of which corresponds to a specific frequency. Quantization of light was extended by Einstein in 1905 to explain the *photoelectric effect* in metals and by Bohr in 1913 to explain the emission spectrum of atomic hydrogen. According to quantum theory, light and other EMW can be considered as

consisting of hypothetical zero-mass particles that move with the speed of light and carry energy and momentum. The term *photon* was introduced by G. N. Lewis much later to emphasize the particle-like properties of light quanta. The particle-like picture was extended later to other types of wave motion, for example, phonons are quantized lattice vibrations in a solid, polarons are quantized waves of lattice polarization in a dielectric material and magnons are quantized magnetic waves.

The quantum energy associated with a photon (ϵ) depends on its frequency according to the fundamental relation

$$\epsilon = hf \tag{2.2}$$

where h is Planck's constant equal to 6.62×10^{-34} J-s. The energy equivalent of a photon is usually stated in units of the electron-volt (eV) where 1 eV = 1.60×10^{-19} J.[4] In thermochemical units, 1 eV = 23.06 kcal/mol. Figure 2.3 indicates the quantum energy associated with the optical radiations. Light interactions with matter are most efficient when the photon energy is comparable to the difference between the final and initial energy states of the constituent atoms or molecules. UV photons can break chemical bonds and eject electrons from atoms and molecules. Photons of visible light "excite" valence electrons in atoms and molecules to higher energy states. IR photons interact with the vibrational motions in molecules and solids.

Example 2.1 Conversion of optical power to photons

A laser emits an optical power of 1 W at 500 nm in a 1 mm diameter collimated beam. How many photons are contained in a 1 m length of the beam?

The power per unit area in the laser beam is: $1 \text{ W}/[(\pi/4) \times (1 \times 10^{-3} \text{ m})^2] = 1.27 \times 10^6$ (J/s-m^2). The total energy in 1 m of beam length is: 1.27×10^6 (J/s-m^2) x 1 m / 3.00×10^8 (m/s) = 4.24×10^{-3} J. The photon energy at 500 nm from Equation (2.1) and Equation (2.2) is: 6.62×10^{-34} (J-s) x 3.00×10^8 (m/s) / (500×10^{-9} m) = 3.97×10^{-19} J. The number of photons in a 1 m beam length is: 4.24×10^3 J / 3.97×10^{-19} J/photon = 1.07×10^{16} photons.

It should be noted that not all radiations used in medicine are EMW. The alpha-rays and beta-rays, emitted by radioactive elements are energetic helium nuclei and electrons, respectively. *Ultrasound* is a longitudinal pressure wave, similar to ordinary sound waves but having a higher frequency and power level. Propagation of ultrasound requires a medium such as air, water, or tissue. The

wavelength and frequency of ultrasound are related by Equation (2.1) with v equal to the velocity of sound in the medium. The frequencies of ultrasound used for medical diagnosis range from 1 to 10 MHz with wavelengths in tissues about 0.15-1.5 mm. This range limits the resolution that can be achieved by ultrasound imaging. Recent interest in optical imaging was motivated by the much greater resolution feasible at shorter light wavelengths compared to ultrasound (see Chapter 7).

2.3 GEOMETRICAL AND PHYSICAL OPTICS

The properties of virtually all optical devices are based on the laws of classical optics. *Geometrical optics* provides an adequate description of those phenomena in which light rays encounter objects having dimensions much larger than the light wavelength including mirror reflection, imaging by lenses, and light transmission by optical fibers. The light rays are termed "directed" because their directions are pre-determined by the geometry of the source and interacting devices. Light interactions with objects comparable in size to the wavelength require the more complex concepts of *physical optics*, for example, lens aberrations, the spread of a laser beam, light scattering by small particles, and holographic imaging. *Interference* and *diffraction* are significant effects in physical optics. Interference occurs when two or more light waves occupy the same region of space. This effect leads to cancellation or reenforcement of intensity depending on how the waves "match up". Diffraction occurs when a physical object such a narrow slit blocks part of a propagating wave. Interference between those waves propagating past different regions of the object leads to intensity variations referred to as diffraction patterns. This section reviews the basic aspects of geometrical and physical optics which are involved in the properties of common optical devices.

2.3.1 Reflection and Refraction of Light

The speed of light is slower in a medium compared to free space owing to interactions of the E-field with atomic electrons. The velocity of light in a medium is expressed as the *refractive index* (n) according to

$$v=\frac{c}{n} \tag{2.3}$$

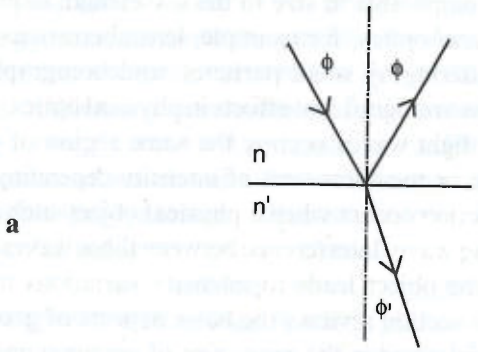
where v is the wave velocity in the medium. Refractive index depends on both the medium and the light wavelength. The variation of refractive index with wavelength is termed *dispersion* after the separation of white light into its constituent colors by a glass prism. Approximate values of n for visible light are 1.0 in air, 1.3 in water, 1.5 in glass, and 1.4 in biological tissues.

Example 2.2 Light waves in a medium

(a) Calculate the frequency of 500 nm green light in air. (b) Calculate the frequency, wavelength, and velocity of the same light in a glass of refractive index equal to 1.50.

(a) The light frequency in air from Equation (2.1) is: 3.00×10^8 (m/s)/500×10^{-9} m = 6.0×10^{14} s^{-1} = 600 THz.

(b) The velocity in the glass from Equation (2.3) is: 3.00×10^8 (m/s)/1.5 = 2.0×10^8 (m/s). The corresponding wavelength in the glass from Equation (2.1) is: 2.0×10^8 (m/s)/6.0×10^{14} s^{-1} = 3.333×10^{-7} m or 333 nm.[5]



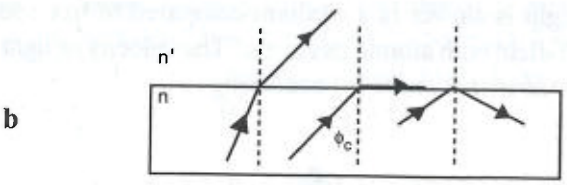


Figure 2.5 (a) Reflection and refraction of a collimated light beam at the interface between two media of refractive indices n and n' where n<n'. (b) Total internal reflection takes place for $n>n'$ and the angle of incidence exceeds the critical angle defined by Equation (2.5).

Figure 2.5a depicts a collimated beam of monochromatic light in air incident on a polished glass surface at an incidence angle ϕ from the perpendicular. A small fraction of the incident intensity is reflected at angle ϕ, typically 3% - 5%. Mirror reflection is referred to as *specular* (originally from shiny speculum metal) to distinguish it from *diffuse reflection* by a rough surface. The wave that enters the medium is bent or *refracted* at a less oblique angle ϕ'. The angle between the incident and refracted rays is determined by *Snell's law*

$$n \sin\phi = n' \sin\phi' \tag{2.4}$$

where n and n' are the refractive indices of the first medium and the second medium, respectively. The *relative refractive index* is defined as: $n_r = n'/n$. Refraction always takes place according to Equation (2.4) when a light ray enters from a medium of higher refractive index ($n' > n$). However, when a light ray encounters a medium of lower refractive index ($n' < n$) the value of $\sin\phi'$ can exceed unity for a sufficiently oblique incidence angle ϕ. This condition cannot satisfy Equation (2.4). *Total internal reflection* (TIR) is the physical consequence of this apparent paradox in which the incident ray is reflected back into the initial medium. A familiar example of TIR is the disappearance of a swimming pool bottom when viewed from outside at an oblique angle. The *Snell's law critical angle* (ϕ_c) for TIR is given by

$$\phi_c = \arcsin(n'/n) \tag{2.5}$$

where n and n' refer to the higher and lower refractive index media, respectively (Figure 2.5b). Optical fibers utilize TIR to transport light over long distances (Section 2.6.3.)

Example 2.3 Snell's law

A light beam in air enters glass at 45°. The refractive index of the glass is 1.50. Calculate: (a) the refraction angle within the glass; (b) the maximum angle at which a light ray propagating within the glass can exit to air.

(a) Refraction at the air-glass interface obeys Snell's law. From Equation (2.4.): $\sin\phi' = (1/1.50) \sin 45° = 0.471$; $\phi' = \arcsin 0.471 = 28.1°$.

(b) Total internal reflection occurs for incidence angles exceeding: ϕ_c = arc sin(1.0/1.5) = 41.8°.

Snell's law does not predict the distribution of intensity between the reflected and refracted waves. This information is provided by the *Fresnel relations* which are derived from EMW theory. The special case of perpendicular incidence leads to the following expression

$$R(\phi=0)=\frac{(n_r-1)^2}{(n_r+1)^2} \tag{2.6}$$

where R is the relative intensity of the reflected light or *reflection coefficient.*[6] For a typical glass with $n'_r = 1.5$ the value of R($\varphi = 0$) is 0.04.

Snell's law applies for "ordinary" or non-polarized light. More generally, R depends on the polarization of the incident light relative to the interface. The E-vector of linearly polarized light can be resolved into two components, which are perpendicular and parallel to the *plane of incidence.* In Figure 2.4a the plane of incidence is the plane of the page. For the component of the electric vector perpendicular to the plane of incidence (referred to as S-polarization), the reflection coefficient R starts out at the value given by Equation 2.6 and increases with increasing angle of incidence to $R = 1$ at 90° or "grazing" incidence (Figure 2.6). For the component of the electric vector parallel to the plane of incidence (referred to as P-polarization), R starts out at the same value given by Equation (2.6), falls to zero at an angle of incidence referred to as the *Brewster angle*(φ_B), and then increases to unity at grazing incidence. The Brewster angle depends on the relative refractive index according to

$$\tan \varphi_B = n_r \tag{2.7}$$

A typical value of (φ_B) is 56.3° for an optical glass with $n_r = 1.50$. The dashed line in Figure 2.6 shows the total reflection coefficient for non-polarized light with equal contributions from each polarization component. The reflection and refraction of polarized light have important practical applications It is shown in

elementary optics that *the reflected ray and transmitted ray are perpendicular to each other at the Brewster angle.* When ordinary light is incident at φ_B , the reflected ray is 100% linearly polarized in the S-direction and the refracted ray is partially polarized in the P-direction (Figure 2.6). A high degree of polarization can obtained by passing a light beam through a stack of plates tilted at φ_B . Each successive plate increases the P-polarization of the transmitted beam without any reflection losses. The end widows of a laser cavity are tilted at j_B. Multiple traverses of the laser light through the Brewster window are equivalent to a single passage through a stack of plates.

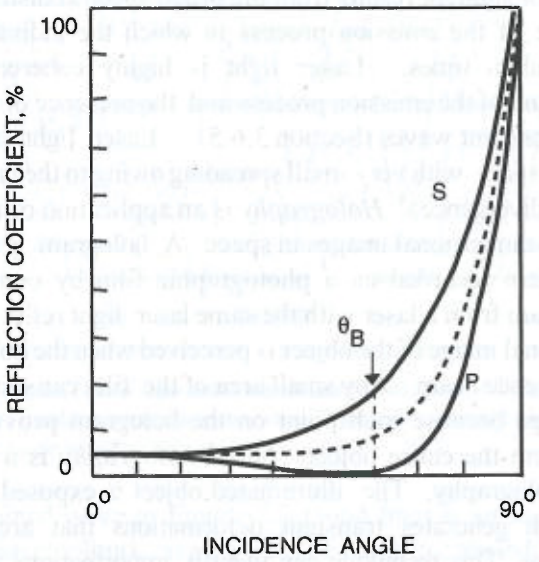


Figure 2.6 The reflection coefficient depends on the angle of incidence and the polarization. The figure shows the dependence of *R* on j for the **E**-vector parallel to the plane of incidence (P-polarization) and perpendicular to the plane of incidence (S-polarization). The dashed line applies for non-polarized light. The reflected and refracted rays are perpendicular at the Brewster angle φ_B and the reflected ray is 100% S-polarized.

2.3.2 Interference

If an approximately monochromatic light source is divided into two beams which are then brought together after traveling different optical paths, the optical image of the combined beam consists of an array of bright and dark fringes. This effect is *interference.* The fringes originate because the amplitudes in different regions of the two beams either add or subtract. Interference is possible only if the two beams are at least partially *coherent.* In coherent light the individual waves

move together in space and time. Coherent light is analogous to a platoon of soldiers marching together in the same direction at exactly the same speed. However, if there are small variations in either speed or direction, the individual marchers will separate after some distance. Ideal coherence can never occur because the output of even the most highly monochromatic light source has a finite spectral bandwidth and beam divergence. The maximum path difference over which two light beams can form an interference pattern is the *coherence length*. Some typical values of coherence length are 1000 m for a helium-neon laser, 6 mm for a low-pressure sodium arc, and 0.001 mm for sunlight. The weak coherence of sunlight results from the broad spectral distribution and also from the nature of the emission process in which the individual atoms emit radiation at random times. Laser light is highly coherent owing to the synchronous nature of the emission process and the presence of an optical cavity that reinforces coherent waves (Section 3.6.5). Laser light can propagate for long distances in space with very small spreading owing to the narrow bandwidth and small beam divergence. [7] *Holography* is an application of interference used to generate three-dimensional images in space. A hologram is two-dimensional interference pattern recorded on a photographic film by combining a highly coherent light beam from a laser with the same laser light reflected by an object. A three-dimensional image of the object is perceived when the hologram is viewed by the same reference beam. Any small area of the film can recreate a complete holographic image because each point on the hologram provides interference contributions from the entire object. *Shock-holography* is a recent practical application of holography. The illuminated object is exposed to simultaneous ultrasound which generates transient deformations that are visible on the holographic image. This technique can identify imperfections in manufactured objects and biological tissues.

Example 2.4 Narrow bandwidth light

An approximately monochromatic light beam has a wavelength of 500 nm and a wavelength spread of 1 nm. Calculate the frequency bandwidth.

The wavelength and frequency are related by: $f = c/\lambda$. Elementary differentiation leads to: $\delta f = (c/\lambda^2)\delta\lambda$, where δf and $\delta\lambda$ are the frequency bandwidth and wavelength bandwidth respectively. Substituting the quantities gives: $\delta f = [3 \times 10^8 \text{ m/s}/ (500 \times 10^{-9} \text{ m})^2] \times (1 \times 10^{-9} \text{ m}) = 1.2 \times 10^{12} \text{ s}^{-1} = 1{,}200$ GHz. Note that an approximately monochromatic light wave has a very large frequency bandwidth.

2.3.2.1 More Details About Interference

Figure 2.7 depicts interference between two ideally monochromatic waves having the same wavelength, frequency, and amplitude (dashed lines). The relative displacement of two waves in distance or time is referred to as a *phase shift* because one wave reaches its maximum amplitude before or after the other

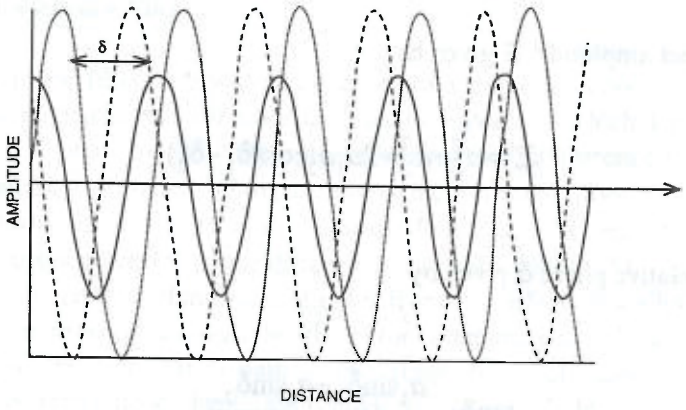


Figure 2.7 Superposition of two sinusoidal waves having the same wavelength and amplitudes. The resultant wave (solid line) has the same wavelength and an amplitude that depends on the relative phase difference between the two waves.

wave.The combined wave in Figure 2.7 (solid line) is sinusoidal at the same frequency, with an amplitude ranging from zero to twice the individual amplitudes depending on the relative phase shift. Phase shift is frequently expressed in units of the wavelength. The maximum reinforcement occurs when the waves are "in phase" by a multiple of the wavelength ($\Delta\lambda = 0, \lambda, 2\lambda$, etc.) and maximum cancellation occurs when the waves are "out of phase" by a multiple of one half-wavelength: $\Delta\lambda = \lambda/2, 3\lambda/2$, etc. The effect of phase on interference can be analyzed by considering the simpler case of two oscillations fixed in space having the same frequency f, with different maximum amplitudes ψ_1 and ψ_2 and different phases δ_1 and δ_2 . The instantaneous amplitudes ψ_1 and ψ_2 are given by

$$\psi_1 = a_1 \sin(2\pi ft - \delta_1)$$

$$\psi_2 = a_2 \sin(2\pi ft - \delta_2)$$

For $\delta_2 > \delta_1$, amplitude ψ_1 lags behind amplitude ψ_2 by a time difference $\Delta t = (\delta_2 - \delta_1)/2\pi$ The algebraic sum of the two amplitudes leads to the combined oscillation:

$$\psi_1 + \psi_2 = \Xi \cos(2\pi ft - \delta) \tag{2.8}$$

with the net amplitude Ξ given by

$$\Xi^2 = a_1^2 + a_2^2 + 2a_1 a_2 \cos(\delta_1 - \delta_2)$$

and the relative phase δ given by

$$\tan\delta = \frac{a_1 \sin\delta_1 + a_2 \sin\delta_2}{a_1 \cos\delta_1 + a_2 \cos\delta_2}$$

Equation (2.8) shows that the combined oscillation is sinusoidal at the same frequency as the constituent oscillations with a resultant amplitude Ξ and phase δ. The equivalent analysis is applicable for light waves occupying the same region of space. The radiant power carried by the resultant wave is proportional to the Ξ^2 in Equation (2.8). The key results are most easily seen for equal initial amplitudes $a_1 = a_2 \equiv a$.

Case (1) The combining waves are in phase. In this case $\delta_1 = \delta_2$ and $\Xi^2 = 4a^2$. Constructive interference results. The resultant intensity is twice as high as the sum of the intensities of the two constituent waves.

Case (2) The combining waves are 180° out of phase: In this case $\delta_2 - \delta_1 = 180°$ and $\Xi = 0$. Destructive interference results. There is no resultant wave.

Case (3) The combining waves are incoherent: In this case there is a random relationship between δ_1 and δ_2. The value of $\cos(\delta_2 - \delta_1)$ averages to zero over many oscillations and $\Xi^2 = 2a^2$. In this case the total intensity equals the sum of the individual intensities.

Coherent light can attain extremely high power levels in regions of space

where the individual waves are in phase. For example, if 10 equal monochromatic waves of unit amplitude are added incoherently the total power is $10 \times 1^2 = 10$. This situation applies to the combination of ordinary light beams. However, if the waves are added coherently the total power is $(1+1+1+1+1+1+1+1+1+1)^2 = 100$. The high optical power levels in a laser cavity result from the superposition of coherent waves.

2.3.2.2 Interference Filters

Interference filters are widely used for wavelength selection. The basic operation of an interference filter is illustrated in Figure 2.8 which depicts a thin glass film illuminated with white light. Constructive interference takes place when the phase difference between successive reflected rays (a,b,c, ..) or successive transmitted rays (a',b',c', ..) equals 0, λ , 2λ , 3λ, etc. Destructive interference takes place for phase differences of $\lambda/2$, $3\lambda/2$, $5\lambda/2$, etc. The phase shift induced by a thin film originates from two effects. Parallel rays that travel different distances through the film before combining are phase-shifted by the differences in their optical paths. The other source of phase shift is less obvious. Electromagnetic theory shows that reflection of light at a medium of higher refractive index (the top face in Figure 2.8) leads to an additional $\lambda/2$ phase shift that does not occur for reflection at a lower refractive index medium (the bottom face in Figure 2.8).

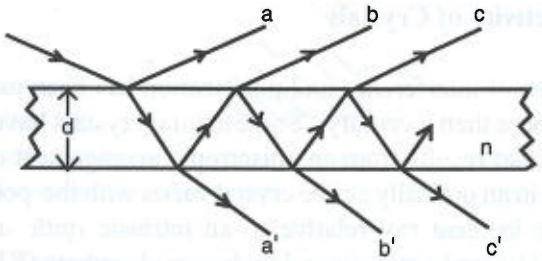


Figure 2.8 Interference of monochromatic light by a thin dielectric film. Constructive interference takes place when the phase between parallel rays equals an integral multiple of the wavelength.

Thus, constructive interference occurs for reflected light when the path difference equals $(p + .5)\lambda / n_r$ for $p = 0, 1, 2$, etc., while constructive interference of the transmitted rays occurs when the path difference is $p\lambda/n_r$. The minimum path difference for perpendicular incidence is 2d. Therefore, if a thin film is illuminated by white light and viewed by reflection, bright fringes are observed

only for those wavelengths that satisfy the reinforcement condition. An additional requirement for fringe formation is that the film thickness must be less than the coherence length of the incident light. Thin-film interference of sunlight requires a very thin film owing to the short coherence length. Interference of laser light is feasible with relatively thick films.

Example 2.5. Thin film interference

Consider a 0.3 μm film of soapy water in air, for example, as might be formed by dipping a wire ring. If the film is illuminated with white light at perpendicular incidence, which colors of fringes are seen by reflection and transmission?

Bright reflection fringes will be seen for wavelengths that satisfy the reinforcement condition: $2d = (p + 0.5)\,\lambda / n_r$. Substituting $d =$ 350 nm and $n_r = 1.33$ for water gives: $\lambda = (2 \times 350 \times 1.33)/(p + 0.5)$. Setting $p = 0, 1, 2$, etc. leads to $\lambda =$ 1862 nm, 621 nm, 372 nm, etc. Only the red light fringes at 621 nm fringes will be observed by the human eye. For the same soap film, interference of transmitted light satisfies: $2d = p\lambda/n_r$. The wavelengths of reinforcement are: $2 \times 350 \times 1.33/p =$ 931 nm, 466 nm, 310 nm, etc. This film can act as a selective transmission filter for blue light at 466 nm. However, the spectral bandwidth would be broad because only a few waves can combine owing to the low reflection coefficient of the soap-air interface.

2.3.2.3 Optical Activity of Crystals

The combination of interference and polarization has been used in passive optical devices for more than a century. Some natural crystals have the property of *optical activity* that results from an anisotropic arrangement of the atoms. The refractive index in an optically active crystal varies with the polarization and the direction of the incident ray relative to an intrinsic *optic axis*. Calcite ($CaCO_3$), quartz (SiO_2), and potasssium dihydrogen phosphate (KH_2PO_4, KDP) are *uniaxial* or *birefringent* crystals. They have a single optic axis and two refractive indices. A *biaxial* crystal such as mica has two optic and three refractive indices. Calcite is described as "doubly refracting" because an entering light beam separates into two beams having perpendicular polarizations and different propagation velocities. Figure 2.9a depicts double refraction by a calcite crystal. The crystal plate was cut with the optic axis tilted at 45° to the faces. The incident beam splits into an "ordinary" ray (O-ray) and an "extraordinary" ray (E-ray) with perpendicular polarizations. The faster O-ray has a constant velocity and obeys Snell's law according to Equation (2.4). The velocity of the

slower E-ray depends on its direction relative to the optic axis. A birefringent crystal can be fabricated as a high-quality linear polarizer. The *Nicol prism* consists of a split and cemented calcite rhomb (Figure 2.9b). The E-ray is non-deviated while the O-ray is deflected out of the crystal by internal reflection.

A *retardation plate* is a classical application of optical activity that has received recent impetus with the development of lasers. A *half-wave plate* is cut from an optically active crystal at a thickness of $\lambda/2$. This device rotates the plane of linearly polarized light by an angle 2θ, where θ is the angle between the incident polarization and the optic axis (Figure 2.10). A half-wave plate can be used as a fine-tuning element in a laser by rotating the plate around the Brewster angle. The thickness of the plate selects the wavelength of polarized light amplified by the laser cavity. A *birefringent filter* consists of one or more half-wave plates separated by sandwiches of polarizing films. The transmitted spectrum has a series of narrow bands, widely spaced in wavelength, which can be shifted by rotating the unit. A *quarter-wave plate* is cut at a thickness of $\lambda/4$. When linearly polarized light is incident on a quarter wave plate with the electric vector oriented at 45°to the optic axis, the horizontal and vertical projections

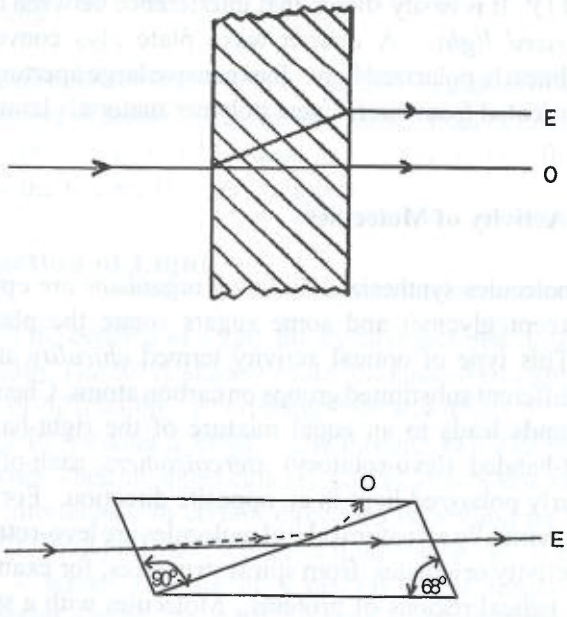


Figure 2.9 (a) A birefringent crystal plate separates non-polarized light into two rays of linear polarized light with perpendicular polarizations. The 45° lines depict the direction of the optic axis. (b) A calcite Nicol proem transmits only the linear polarized E-ray. The O-ray is reflected out of the crystal at the cemented interface.

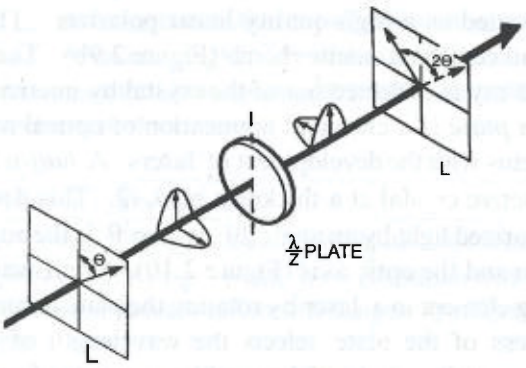


Figure 2.10 A half-wave plate rotates the plane of monochromatic linear polarized light (L). The vertical component of the E-vector propagating parallel to the optic axis (dashed line) is unchanged, while the phase of the horizontal component is retarded by a half-wavelength. The linear polarization of the emergent wave is rotated from θ to 2θ.

Of the electric vector emerge with thee same amplitude and a relative phase shift of λ/4 (Figure 2.11)[8]. It is easily shown that interference between these rays lead to *circular polarized light*. A quarter-wave plate also converts circularly polarized light to linearly polarized light. Inexpensive large aperture retarders are now available fabricated from birefringent polymer materials laminated between glass plates.

2.3.2.4 Optical Activity of Molecules

Nearly all molecules synthesized by living organisms are optically active. Amino acids (except glycine) and some sugars rotate the plane of linearly polarized light. This type of optical activity termed *chirality* arises from the presence of four different substituted groups on carbon atoms. Chemical synthesis of chiral compounds leads to an equal mixture of the right-handed (dextro-rotatory)and left-handed (levo-rotatory) *stereoisomers*, each of which rotates the plane of linearly polarized light in an opposite direction. For unknown and profound reasons, virtually all natural chiral molecules are levo-rotatory. Another form of optical activity originates from spiral structures, for example, the DNA double helix and helical regions of proteins. Molecules with a spiral structure interact differently with right-handed and left-handed circular polarized light. In *circular dichroism* (CD) the absorption spectrum is different for the two directions of circular polarization. In *optical rotatory dispersion* (ORD) the dependence of the refractive index on wavelength is different for the two

directions of circular polarization. CD and ORD spectra are complementary measurements. CD is measured in spectral regions of strong absorption in order

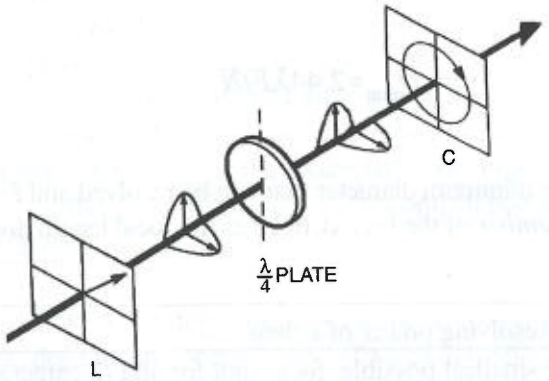


Figure 2.11 A quarter-wave plate converts linear polarized light (L) to circular-polarized light. The incident E-vector is 45° to the optic axis (dashed line). The phase of the horizontal component is retarded by a quarter-wavelength. The transmitted transverse components combine to give circular-polarized light.

to promote the differences in absorption. ORD is measured in spectral regions of weak absorption in order to emphasize the differential refractive index. ORD spectra can be converted into CD spectra and *vice versa* by a fundamental pair of integrals called the Kronig-Kramers relations.

2.3.3 Diffraction of Light

Diffraction is the bending of a light ray by an object of dimensions comparable to the wavelength. The fuzzy shadow of an illuminated straight edge on a screen is an example of diffraction. Alternating light and dark diffraction fringes are easily observed by viewing a bright light through two closely spaced fingers. Diffraction occurs when an object cuts off part of an incident wave front, leading to destructive interference in certain directions depending on the shape of the object. The diffraction pattern of a thin slit illuminated with monochromatic light as viewed on a screen consists of a central bright fringe located between increasingly weaker and narrower bright fringes on each side separated by dark regions. The diffraction pattern of a circular aperture illuminated by a point source of monochromatic light consists of a central bright disc surrounded by concentric fainter rings. The spacing of diffraction fringes is a classical problem in physical optics. Diffraction limits the ability of a lens to produce a point image of a point

source and to focus a collimated laser beam to a point. The resolving power of a uniformly illuminated simple lens is given by:

$$D_{\min}=2.44\lambda F/N \tag{2.9}$$

where $D_{\min}$ is the minimum diameter that can be resolved and F/N is the relative aperture or *F-number* of the lens, defined as the focal length divided by the lens diameter.

Example 2.6 Resolving power of a lens

Calculate the smallest possible focal spot for an F/2 camera lens at 500 nm.

According to Equation (2.9): D_{min} = 2.44 x 0.0005 mm x 2 =0.0024 mm. This result applies for an ideal lens and a perfectly collimated, uniform circular beam.

The complete theory of diffraction is described in standard optics texts. The angular spread of the dark diffraction fringes for a single slit illuminated by a collimated, monochromatic light beam is given by

$$\sin\theta=p\lambda/a \tag{2.10}$$

where a is the slit width, θ is the angular deviation of the diffracted ray, and p is an integer specifying the "order" of the diffraction. The actual spacing of the fringes on a viewing screen increases with increasing distance of the screen from the slit. The intensity of each bright fringe is given by

$$I=I_m[\frac{\sin\alpha}{\alpha}]^2 \tag{2.11}$$

where $\alpha \equiv (\pi a/\lambda)\sin\theta$ and I_m is the maximum intensity of the central bright fringe. The bright fringes do not lie exactly halfway between the adjacent dark fringes. A numerical calculation leads to the values of a (in radians) for the first few maxima: 0, 1.4303 p, 2.4590 p, 3.4707 p, etc.

Example 2.7 Single-slit diffraction

Consider diffraction of 500 nm green light by a 2 mm width slit. Calculate: (a) the angular deflection of the first three dark fringes; (b) the relative intensity of the first three bright fringes.

(a) From Equation (2.10): $\sin\theta = (500 \times 10^{-9} / 2 \times 10^{-6})p = 0.25\,p$. Substituting $p = 1$, 2, and 3 gives $\theta_1 = 14.48°$, $\theta_2 = 30.00°$, and $\theta_3 = 48.59°$. The angular spread between the successive orders is quite large

(b) From Equation (2.11) and the values of a for bright fringes: $I_1 = 1$, $I_2 = 0.0472$, and $I_3 = 0.0165$. More than 93% of the optical power is concentrated in the broad central band.

A diffraction pattern is readily visualized by the famous two-slit experiment of Thomas Young depicted in Figure 2.12a . Two closely spaced slits S_1 and S_2 are illuminated by monochromatic light from a single slit S. The largest separation between the single slit and double slits that leads to interference fringes on the screen provides an experimental determination of the coherence length.

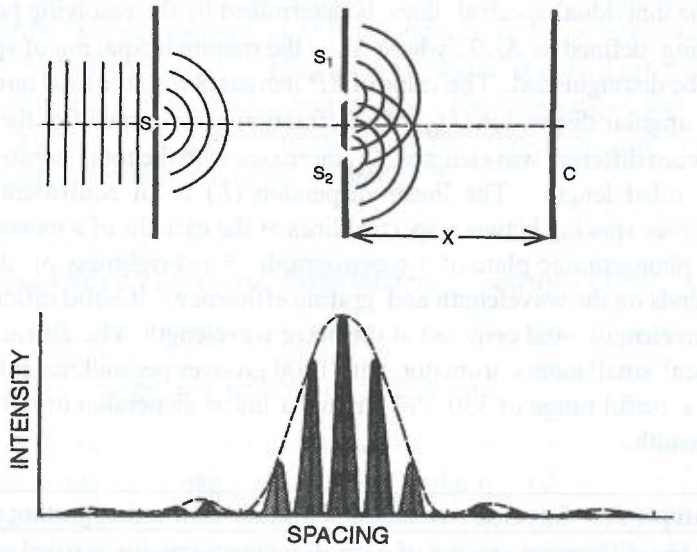


Figure 2.12 Two-slit diffraction. (a) Approximately coherent light emerging from the single slit S is incident on parallel slits S_1 and S_2 separated by d. The interference pattern is viewed on a screen C located at a distance X from the two slits. (b) The diffraction pattern consists of alternating light and dark fringes separated by $X\lambda/d$. The envelope of the diffraction pattern corresponds to the diffraction pattern of one slit. Multiple slits lead to sharper lines having the same spectral distribution.

The two-slit diffraction pattern consists of alternating bright and dark fringes within an intensity envelope corresponding to the single slit diffraction pattern (Figure 2.12b). When the distance of the slits to the screen (X) is much larger than the slit spacing (d), the alternating bright and dark fringes are equally separated by $\lambda X/d$. The diffraction pattern for white light consists of overlapping sets of fringes belonging to the different orders. The spacing of fringes within each order is proportional to the wavelength. A double-slit can be used for wavelength separation. The fringe pattern is sharper and more intense when the two slits are replaced by a large number of closely-spaced parallel slits. This array forms a transmission *diffraction grating*. For practical reasons reflection-type diffraction gratings are used in monochromators. In a reflection grating the slits are replaced by parallel grooves cut into a metallic surface. The groove faces are tilted or "blazed" relative to the base in order to shift the brightest fringe to the first order. The first precision diffraction gratings were cut on a metal base by a ruling engine. However, ruling engines always have minute imperfections and generate errors in the diffraction pattern call "ghosts". (Note that resealed red blood cell membranes are also called "ghosts".) The best available gratings at present are fabricated from images of optical interference patterns. The bandwidth of the individual spectral lines is determined by the resolving power (RP) of the grating defined as $\Delta\lambda/\lambda$, where $\Delta\lambda$ is the minimum spacing of spectral lines that can be distinguished. The value of RP increases with the total number of grooves. The angular dispersion (D_θ) of a diffraction grating specifies the angular spread between different wavelengths. D_θ increases with the total number of grooves per unit ruled length. The linear dispersion (L) is an equivalent parameter that specifies spacing between spectral lines at the exit slit of a monochromator or on the photographic plate of a spectrograph. The brightness of the spectral lines depends on the wavelength and grating efficiency. Useful efficiency occurs for a wavelength band centered at the blaze wavelength. The diffraction grating in a typical small monochromator with 1200 grooves per millimeter blazed at 500 nm has a useful range of 330-750 nm and a linear dispersion of 6.4 nm per mm. of slit width.

Example 2.8 Spectral resolution of a small diffraction grating monochromator.

The diffraction grating of a small monochromator is ruled with 24,000 lines over a 20 mm length. Calculate the most closely-spaced spectral lines that can be resolved in the first order spectrum at 500 nm.

RP for this grating equals 24,000. Therefore, $\Delta\lambda$ =500 nm/24,000 = 0.021 nm. The resolution is two times larger in the second-order spectrum or 0.042 nm. However, the brightness of the lines is much weaker.

2.4 OPTICAL RADIOMETRY

Radiometry is the science of optical power measurements. This subject is complicated by several systems of units. *Photometric units* are based on the spectral response of an idealized human eye. Consequently, photometric units are inapplicable for IR and UV radiations. However, photometric units are still widely used in the lighting and photographic industries. *Radiometric units* are based on radiant power using SI units established by the International Commission on Illumination (CIE). This system of units is used in photobiology and photomedicine. *Actinometric units* are based on the photon content of light. Photons are generally used in photophysics and photochemistry for events that take place at atomic and molecular levels. *Radiant flux* (F) is the basic unit of optical power in the SI system. Since power equals energy per unit time, radiant energy (Q) equals the product of the flux in watts and the time in seconds. *Flux density* is a general term for flux per unit area. However, *energy density* usually refers to radiant energy per unit volume. *Luminous flux* in units of the lumen (lm) is the SI photometric equivalent of radiant flux. One lumen equals 683 W at 555 nm, which is the peak of the *visual photopic luminous efficiency function* $V(\lambda)$. The relative value of $V(\lambda)$ equals unity at 555 nm and falls to about 0.001 at 410 nm and 720 nm. Numerical values of $V(\lambda)$ are provided in standard references. All radiometric units have photometric equivalents. The two sets of units are compared in Table 2.1. A general relationship for converting radiometric quantities to the corresponding photometric quantity is

$$Photometric\ Quantity = \int (Spectral\ Radiometric\ Quantity) V(\lambda) d\lambda \quad (2.12)$$

where the limits of the integration are 380 nm to 770 nm.

Radiometric properties are employed to specify the power and directionality of light emitted by a source, incident on a receiver, and reflected by a surface. In order to do this, it is necessary to describe directions in space. *Solid angle* (Ω) is the three-dimensional equivalent of *polar angle* in a plane. Polar angle in units of radians (rad) is defined as the length of a circular arc centered around a point in a plane divided by the radius (Figure 2.13a). The total polar angle subtended by a circle of radius r about its center is $2\pi\ r/r = 2\pi$ rad or 360°. Any simple closed loop in a plane subtends 2π rad. The solid angle subtended by a spherical surface around a point in space is defined as the area of the spherical surface divided by

Table 2.1 Radiometric and Photometric Quantities

Radiometric Quantity	Definition	Equation	S. I. Units	Photometric Equivalent
Radiant energy		Q	J	Luminous energy (lm-s)[a]
Radiant flux	Radiant power	F = dQ/dt	W	Luminous flux (lm)[b]
Radiance	Radiant flux in a given direction per unit solid angle and per unit area perpendicular to the direction of propogation	L= dI/dA cos θ	W/m^2-sr	Luminance (cd/m^2)[c]
Intensity	Radiant flux emitted by a source in a given direction per unit solid angle	I = dF/dΩ	W/sr	Candlepower (cd; lm/sr)
Irradiance	Radiant flux arriving per unit area	E = dF/dA	W/m^2	Illuminance (lux; lm/m^2)[d]
Exitance	Radiant flux emmitted or reflected per unit area	M = dF/dA	W/m^2	Luminous exitance (lm/m^2)
Radiant exposure	Radiant energy arriving per unit area	H = Et	J/m^2	Luminous exposure ($lm-s/m^2$)

[a]The *talbot* is a unit of luminous energy equal to 1 lm-s.

[b]One *lumen* is equivalent to approximately 683 W at 555 nm and to fewer watts at other wavelengths (see text).

[c]Other luminance units are the *nit* (1 cd/m^2); the *stilb (1 cd/cm2);* the *lambert* [$(1/\pi)$ cd/cm^2].

[d]Another illuminance unit is the *foot-candle* (1 lm/ft^2).

the square of the radius (Figure 2.13b). The units of solid angle are steradians (sr). The total solid angle subtended by a spherical shell about its center is $4\pi r^2 / r^2 = 4\pi$ sr. Any simple closed surface on space subtends 4π sr.

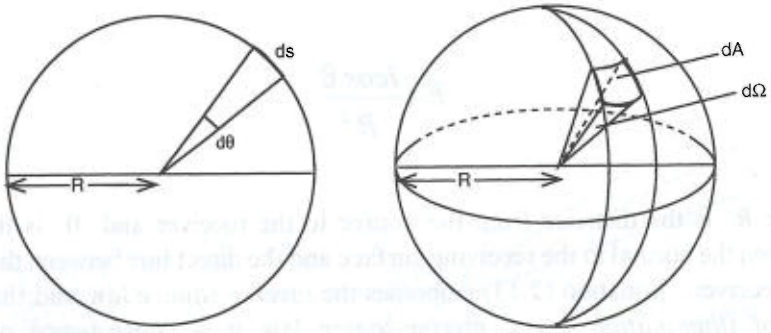


Figure 2.13 (a) The differential polar angle dθ at a point in a plane subtended by a circular arc d*s* is given by: dθ = d*s*/*R* radians where *R* is the radius of the circle. (b) The differential solid angle dΩ at a point in space subtrended by spherical surface dA is given by: dΩ at a point in space subtended by spherical surface dA is given by: dΩ = d*A*/*R*² steradians where *R* is the radius of the spherical shell.

Example 2.9 Solid angle

Show that the integration of the differential solid angle dΩ over the total solid angle is 4π sr for a spherical shell of constant radius *r*.

It is convenient to use spherical coordinates with θ as the polar angle, ϕ as the azimuthal angle, and *r* as the radius (Figure 2.14b). The differential surface element is: $da = (r\,d\theta)(r \sin\theta\, d\phi)$ and $\cos\theta = 1$ for this geometry. Performing the integration:

$$\Omega = r^2 \int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} d\phi / r^2$$

Intensity (*I*) is the radiometric property describing the directionality of radiant power emitted by a light source. For a small source, or a surface element of a large source, intensity is defined as the flux emitted in a given direction per unit solid angle in that direction (Figure 2.14a). If a source of intensity *I* emits uniformly in all directions, the total emitted flux is $4\pi I$. The corresponding photometric quantity of luminous intensity is *candlepower* with the SI unit of *candela* (cd). *Irradiance* (*E*) is the fundamental unit of illumination, defined as

the incident flux on a unit area of receiving surface (Figure 2.14b). The corresponding quantity for an emitting or reflecting surface is the *exitance*. The irradiance at a small surface illuminated by a point source of uniform intensity I is given by

$$E=\frac{I\cos\theta}{R^2} \tag{2.13}$$

where R is the distance from the source to the receiver and θ is the angle between the normal to the receiving surface and the direct line between the source and receiver. Equation (2.13) embodies the *inverse-square law* and the *cosine law of illumination.* The inverse-square law is a consequence of three-dimensional geometry because the area of a spherical shell increases as the square of the radius. The cosine law of illumination is important for phototherapy. Light delivery systems are designed to maximize the perpendicular incidence of the therapeutic radiation. The light intensity emitted by a practical source depends on the angle between the emitting surface and the receiver. The *radiance* (L) specifies the intensity emitted in a given direction per unit projected area in that direction (Figure 2.14c). The equivalent definition of radiance applies for reflected light. The photometric equivalent of radiance is *luminance* (lm/-sr-m^2). An object is "bright"when its image on the retina or other receiver has a high flux density in a narrow angular spread. A low-power light source can be very bright if the emitted light is confined to a narrow solid angle. For example, a 1 mW helium-neon laser viewed along the axis is much brighter than a 300-W projection lamp. The radiance of a practical light source depends on the wavelength. *Spectral radiance* is defined as the radiance per unit bandwidth. The spectral radiance of the extraterrestrial sun at 633 nm is 0.0002 (W/m^2-sr-nm) compared to 10,000 (W/m^2-sr-nm) for a small helium-neon laser.

In 1760 Lambert observed that sunlight reflected by a white painted wall appeared equally bright at all viewing directions. He correctly deduced that the scattered intensity distribution does not depend on the direction of the incident light and decreases with the cosine of the viewing angle from the vertical. *Lambert's cosine law* has been verified for many types of paper and powder layers. Such materials consist of many small reflectors oriented in random directions. The reflected light is referred to as *diffuse light* to distinguish it from directed light. Light sources that obey Lambert's cosine law are referred to as *lambertian*. The Sun is a lambertian source.

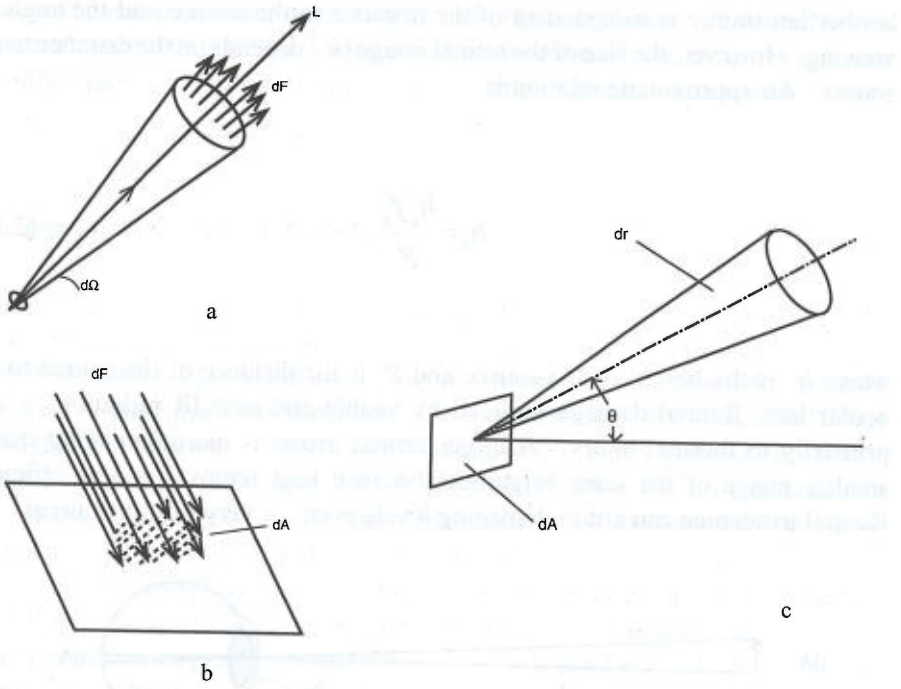


Figure 2.14 (a) Intensity (I) is a measure of the radiant flux per unit solid angle emitted in a given direction. (b) Irradiance (E) is a measure of the radiant flux per unit area incident on a planar surface. (c) Radiance (L) is a measure of the intensity per unit projected area or "brightness".

The radiometry associated with the human retina is important in connection with optical hazards. Figure 2.15 depicts the imaging of a distant lambertian source on the retina, where dA is a surface element of the source, dA' is a surface element of the retinal image, and A_p is the area of the pupil. A straightforward calculation of the irradiance at the retina (E_{ret}) leads to

$$E_{ret} = \frac{T_e L_s A_p}{f_e^2} \tag{2.14}$$

where T_e is the overall transmission coefficient of the eye, L_s is the source radiance, and f_e is the focal length of the eye. The subjective brightness of a

lambertian source is independent of the distance to the source and the angle of viewing. However, the size of the retinal image (h_r) depends on the distance to the source. An approximate relation is

$$h_r = \frac{h_s f_e}{R} \tag{2.15}$$

where h_s is the height of the source and R is the distance of the source to the ocular lens. Retinal damage induced by visible and near-IR radiations is due primarily to thermal injury. A larger retinal image is more damaging than a smaller image of the same brightness because heat removal is less efficient. Retinal irradiance can attain damaging levels even for very distant sources.

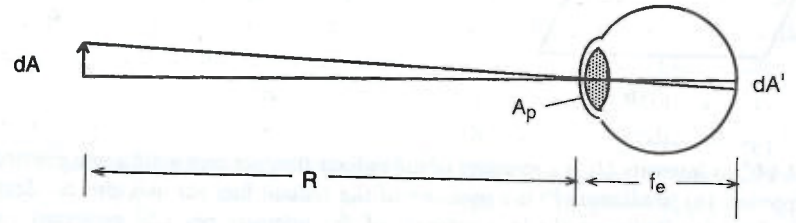


Figure 2.15 Retinal imaging. A lambertian source of area dA located a distance R from the ocular lens forms an image on the retinal of area dA'.

Example 2.10 Direct viewing of the sun

The sun is 1.5 x 10^{11} m from the earth with a diameter of 1.4 x 10^9 m. The approximate radiance of the noon sun from 400-1400 nm is 1000 W/cm ^{2}-sr.

(a) Calculate the size of the sun's image on the retina. (b) Calculate the irradiance of the sun on the retina for a 2 mm pupil and T_e = 1. (c) For a thermal injury damage threshold of 100 J/cm^2, calculate the maximum time of direct viewing of the sun. The focal length of the normal human eye is 1.7 cm.

(a) The size of the sun's image on the retina from Equation (2.15) is: h_r = 1.4 x 10^9 m x 1.7 cm/1.5 x 10^{11} m = 0.015 cm.

(b) The retinal irradiance from Equation (2.14) is: E_r = [(π /4) x (0.2 cm) 2] x 1000 W/cm^2 -sr/(1.7 cm)^2 = 11 W/cm^2

(c) The maximum safe time of direct viewing is: 100 J/cm^2 /11 W/cm^2 = 9s. The safe viewing time is much briefer with a binocular or telescope owing to the magnified retinal image. A 10 times magnification reduces the safe viewing time to less than 1 sec!

2.5 LIGHT SCATTERING

Light scattering is one of the most fascinating of all optical phenomena. The white appearance of cumulus clouds against the blue sky and the red sunset are manifestations of light scattering. Light scattering also is one of the most important factors involved in light dosimetry because biological tissues and fluids are highly inhomogeneous substances. Microscopic entities in tissues induce multiple scattering of incident light leading to the appearance of turbidity or "cloudiness". In many cases the poor penetration of visible light into a tissue or fluid results from its light scattering and not its absorptive properties.

The correct explanation for atmospheric light scattering was provided by Lord Rayleigh in 1871, three decades before the first elucidation of atomic structure. Rayleigh proposed that the oscillating E-field of a light wave initiates synchronous vibrations of electrons in the scattering particles. The vibrating electrons radiate scattered light at the same wavelength as the incident light. Owing to the transverse character of EM waves, the intensity and polarization of scattered light in a given direction depend on the incident polarization and the angle between the incident and scattered directions. Using only dimensional arguments, Rayleigh deduced that the scattered intensity in a given direction decreases with the inverse-fourth power of the wavelength. The strong dependence of *Rayleigh scattering* on wavelength explains the blue daytime sky, because molecules and fine dust in the atmosphere preferentially scatter blue light to the earth's surface. However, when the sun is close to the horizon, the more weakly scattered red light is viewed preferentially by line-of-sight.

Example 2.11 Rayleigh scattering

Compare the total scattered intensity of blue light (400 nm) and red light (700 nm) for Rayleigh scattering by small particles.

The scattered intensity for blue light is: $(700/400)^4 = 9.4$ times stronger than red light.

Light scattering by very small particles is adequately explained by Rayleigh's theory. More generally, light is scattered by local variations of the refractive

index. For example, the white appearance of milk is due to light scattered by butterfat particles. Gans and Debye explained light scattering for larger particles as arising from interference between rays originating from different scattering sites. This theory was applied to ideal spherical and cylindrical macromolecules and the random coil structures of globular proteins. The rigorous theory of light scattering is based on Maxwell's electromagnetic equations. The exact solution for an isolated sphere is referred to as Mie scattering, after Gustave Mie who was one the first to solve the problem in 1908. A key result of Mie theory is that particles much larger than the light wavelength scatter predominantly in the forward direction. Figure 2.16a depicts the angular scattering pattern for Rayleigh scattering of ordinary light by a small particle, which has equal lobes in the forward and backward directions. Mie scattering by an insulating sphere is highly forward-directed. The angular pattern in Figure 2.16b was calculated for scattering of green light by a 0.5 μm diameter sphere. The scattered intensity in the forward direction is 200 times stronger than the backward direction.

Light scattering within a turbid material leads to spreading of an incident light beam and loss of directionality. The directionality of the individual scattering events is lost in multiple scattering after several steps and the rays acquire approximately random directions. Light within the material is diffuse. The *energy fluence rate* (Φ) is a useful radiometric quantity for diffuse light within a turbid medium. (Figure 2.17). This definition of flux density is defined as the total flux incident on a small spherical surface centered around a point in the medium divided by the cross-sectional area of that sphere. The energy fluence rate is equivalent to the average intensity, calculated by integrating the radiance over the entire solid angle.

$$\Phi = \int_{4\pi\ sr} L d\Omega \tag{2.16}$$

The units of energy fluence rate are the same as irradiance. However, E and Φ measure flux density in different geometries. A small planar detector measures E, while an isotropic probe measures Φ. For perfectly isotropic light $\Phi = 4E$; for perfectly collimated light $\Phi = E$. The energy fluence rate is also referred to as *spherical irradiance* and *space irradiance*. The term *photon fluence rate* is used when the optical power is expressed as photons per unit area per unit time.

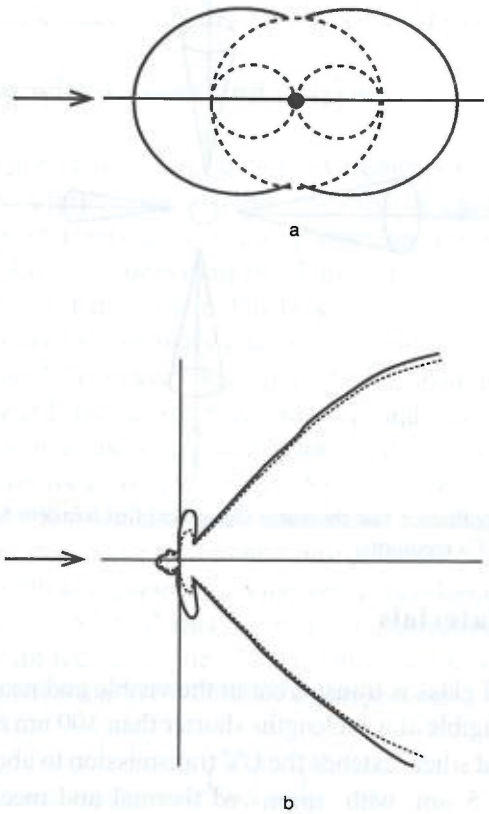


Figure 2.16 An angular scattering diagram depicts the directionality of light scattering by a single particle. (a) Rayleigh scattering. The dashed lines indicate the scattering directions for two perpendicular components of linearly polarized light. The solid line applies for nonpolarized light. (b) Mie scattering by a spherical dielectric particle. The calculation is for $2\pi a n_r/\lambda=4$ and $n_r=n_1/n_2=1.25$, where a is the particle radius, n_1 is the refractive index of the sphere and n_2 is the refractive index of the medium. The solid line and dashed line apply for vertical and horizontal polarization components, respectively.

2.6 PRACTICAL OPTICAL DEVICES

Virtually all of the passive components used in optical systems were invented in the 19th century or earlier. However, the current demands of precise control and high-power applications have led to new implementations of these venerable devices.

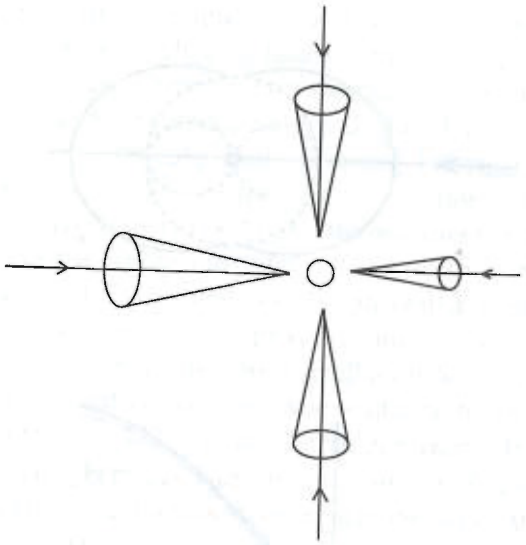


Figure 2.17 The energy fluence rate measures the optical flux incident from all directions at a small spherical region of a medium.

2.6.1 Optical Materials

Ordinary optical glass is transparent in the visible and near-IR regions. The transmission is negligible at wavelengths shorter than 300 nm and longer than 2.8 μm. Synthetic fused silica extends the UV transmission to about 200 nm and the IR transmission to 5 μm, with improved thermal and mechanical properties compared to optical glass. Optical windows and lenses for the far-IR region are of necessity fabricated from inorganic crystals that are brittle and hygroscopic. There are no practical optical fiber materials for the far-IR region at present. Carbon dioxide surgical lasers emitting at 10.6 μm are directed by systems of mirrors. The reflectivity of a freshly deposited aluminum film is 85-95% from 0.2 μm to 30 μm. The silver or aluminum coating in an ordinary mirror is located on the rear surface of a glass plate to reduce oxidation, in which case the spectral range of reflection is limited by the transmission of the glass. The full spectral range of the mirror coating can be utilized with a front-surface mirror coated with a transparent protective layer. Gold mirrors are good reflectors from 0.8 μm to 30 μm and do not require a protective coating. Laser mirrors must be highly reflecting only over a narrow wavelength range, but they must withstand very high optical power levels. They are usually fabricated from dielectric coatings on a hard glass or fused-silica substrate. The basic construction of a laser mirror consists of alternating quarter-wave stacks of high refractive index and low

refractive index materials. Constructive interference between reflected rays differing by successive $\lambda/2$ path differences leads to high reflectivity in a narrow bandwidth, typically exceeding 99.5% for normal incidence.

2.6.2 Imaging with Lenses and Mirrors

Lens design for cameras and optical instruments is a highly specialized subject which the "end-user" is seldom required to master. However, the practitioner of phototherapy may be required to design or modify a simple optical system for light delivery or measurement. This section provides a brief review of optical imaging that may assist in this task.

Simple lenses used for ordinary imaging have spherical surfaces. They are described as "convex", "concave", and "plano". For example, a plano-convex lens has one outward facing spherical surface and one flat surface. The elementary theory of imaging is applicable for *paraxial rays* which are close to the optical axis. The *focal length* of a lens or mirror (F) is the distance from a plane through the optical center or *principal plane* at which parallel rays either converge or diverge, depending on the curvature of the surfaces. In physiological optics the focal length is expressed in *diopters* defined as the reciprocal of the focal length in meters. A "thin" lens can be treated as having a single principal plane with equidistant foci each side. The distances of the object (p) and image (q) from the principal plane are related by the expression

$$\frac{1}{p}+\frac{1}{q}=\frac{1}{F} \tag{2.17}$$

The effective focal length of two thin lenses separated by an air distance d is given by:

$$\frac{1}{F}=\frac{1}{F_1}+\frac{1}{F_2}-\frac{d}{F_1F_2} \tag{2.18}$$

Equation (2.17) applies also for spherical mirrors. The numerical applications of these formulas requires sign conventions: F is taken as positive for a lens that forms a real image and negative for a lens that forms a virtual image. A *real image* can be projected on a screen. A *virtual image* is perceived at those points in space from which divergent rays appear to originate. For example, a camera

lens forms a real image on the film, while a plane mirror forms a virtual image. Assuming that rays travel from left to right, the object distance p is taken as positive if the object is to the left of the lens or mirror surface and negative if the object is to the right of the lens or mirror surface and vice-versa for the image distance q. The ratio of image size to object size or magnification (M) is given by

$$M = -\frac{q}{p} \tag{2.19}$$

The sign of M is positive for an upright image and negative for an inverted image.

Ray tracing is a graphical approach to image analysis for simple optical systems that avoids the requirement for sign conventions. For paraxial rays the basic principles of ray tracing are: (1) A ray parallel to the optic axis is refracted through a focal point. (2) A ray through a focal point is refracted parallel to the optic axis. (3) A ray through the lens center is non-deviated. In Figure 2.18a, a double-convex lens forms an inverted real image of an object located outside the focal plane, for example, a simple camera lens. Figure 2.18b shows that an upright virtual image is formed by an object located inside the focal plane by a double-convex lens. This arrangement is used for a hand magnifier. Figure 2.18c shows an upright, virtual image formed by a double-concave lens for an object located outside the focal plane. Light from a small source can be collimated by collecting the light with a convex lens followed by convex lens to form a collimated beam.[9] The light-gathering efficiency of a lens system is specified by the F-number (F/N). A "fast" lens has a low F/N. Numerical aperture ($N.A.$) is an equivalent parameter defined by

$$N.A. = n_0 \sin\alpha = \frac{1}{(2F/N)} \tag{2.20}$$

where α is the half-acceptance angle of the lens and n_0 is the refractive index of the external medium.

The high-quality lenses used in microscopes and other precision optical instruments are fabricated of high-refractive index optical glass. By judicious use of multiple elements and different types of glasses it is possible to minimize imaging errors, including chromatic aberration (change of focal length with

wavelength), spherical aberration (failure to form a perfect image of a point source on the axis), and astigmatism (off-axis rays are focused as an ellipse instead of a circle). A variety of special lenses are employed in modern optical systems. A *Fresnel lens* consists of concentric stepped rings, each of which bends light as if it were a small portion of a conventional spherical lens. With this concentric groove structure there is no inherent physical limit to the *F/N*. Fresnel lenses have been fabricated from acrylic plastic with very large diameters and high optical speed, typically *F/1*. They are used primarily for light collection in the visible and near-IR regions. A *gradient index lens* (GRIN lens) uses continuously variable refractive index for focusing. GRIN lenses are available in the form of cylindrical rods in which the refractive index increases with the radial distance from the axis. The focusing properties depend on the length of the rod. A GRIN lens can form a real image directly on or very close to the lens surface with very low spherical aberration. GRIN lenses are used to couple a narrow laser beam into an optical fiber or a small photodetector. A *ball lens* is a sphere with a uniform refractive index. Optical fibers are often coupled with a ball lens because the alignment is easier than with a spherical thin lens and they have high coupling efficiency. A *cylindrical lens* has one or two cylindrical surfaces and, therefore, it refracts in only one plane. They can be used to image a point source on a line and to change the height of an image without changing the width. An *aspheric lens* uses one or more non-spherical faces to provide a low F/*N* typically below *F*/1. In many applications it is desirable to use mirrors instead of lenses because mirrors have no chromatic aberration. Special mirror shapes are available in addition to standard plane and spherical mirrors. A parabolic mirror brings a parallel incident beam to a focus and delivers collimated light from a point source. A line source is focused on a parallel line with an ellipsoidal mirror. Ellipsoidal reflectors are employed with flash-lamp pumped lasers to achieve very efficient light delivery.

2.6.3 Optical Fibers

Long optical fibers are used in communications to transport light signals over long distances. The short optical fibers designed for phototherapy applications provide for efficient coupling to lasers and other small, high-power light sources. The basic operating principles are the same for the two types of optical fibers. Light is transmitted down the fiber by a succession of internal reflections until it exits at the far end. The basic design of a *step-index fiber* consists of an inner core and a concentric outer cladding of slightly lower refractive index. (Figure 2.19). A *graded-index fiber* has a core in which the refractive index decreases gradually from the center. The "speed" of an optical

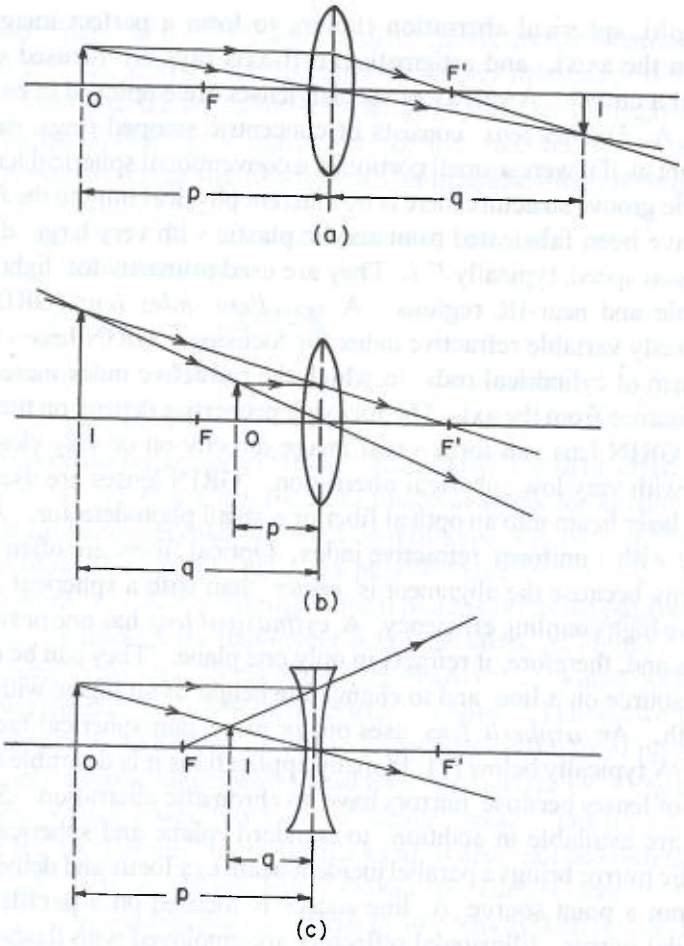


Figure 2.18 Ray-tracing diagrams for spherical thin lenses. (a) A convex lens forms a real, inverted image of an object outside the focal plane. (b) A convex lens forms a virtual, upright image of an object inside the focal plane. (c) A concave lens forms a virtual, upright image of an object inside the focal plane.

fiber can be specified by the *F/N*, the numerical aperture *N.A*, or the acceptance angle α, which are related by Equation (2.20). A simple calculation of the *N.A.* of a step-index fiber based leads to

$$N.A. = \frac{\sqrt{(n_1^2 - n_2^2)}}{n_0} \tag{2.21}$$

where n_1 and n_2 are the refractive indices of the core and cladding, respectively, and n_0 is the refractive index of the external medium.

Example 2.12 Properties of a step-index optical fiber

A fused-silica optical fiber has an inner core refractive index of 1.457 at 633 nm and a plastic outer core with a refractive index of 1.432. Calculate: (a) the numerical aperture; (b) the *F/N*; (c) the full acceptance angle in air.

(a) From Equation (2.15): $N.A. = \sqrt{(1.457^2 - 1.432^2)/1} = 0.27$

(b) $F/N = 1/(2 \times 0.27) = F/1.9$

(c) $\alpha = 2 \arcsin(0.27) = 31.3°$

The acceptance angle for this optical fiber embedded in tissue would be 25° for $n_0 = 1.4$.

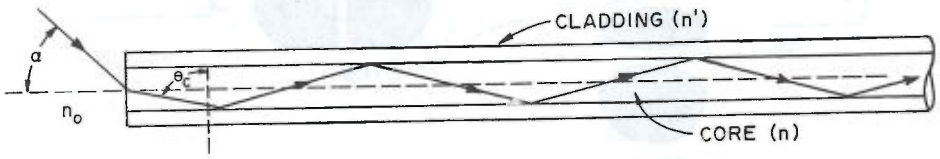


Figure 2.19 Optical paths in a step-index optical fiber. The angle a is the maximum acceptance angle for propagation of the incident ray. The numerical aperture is given by equation 2.21. From Grossweiner, L.I., *The Science of Photobiology*, 2nd ed., Plenum Press, New York, 1989. With permission.

Electromagnetic theory shows that optical fibers are cylindrical optical cavities. This property is relevant mostly for communications fibers. However, similar considerations are involved in laser physics as described in Section 3.6. Light propagation along an optical fiber establishes a set of standing wave patterns or *optical modes* perpendicular to the fiber axis. Each mode corresponds to a specific linear polarization of the transmitted light and a unique spatial pattern of the E-field. A *single-mode fiber* transports only one linearly polarized mode in which the radial E-field intensity has a gaussian (bell-shaped) profile (Figure 2.20a). The properties of commercial optical fibers are specified by the *normalized frequency parameter* or *V- number* defined by

$$V = 2\pi \frac{a}{\lambda_0}(N.A.) \qquad (2.22)$$

where a is the core radius and λ_0 is the wavelength in the external medium. The total number of modes equals $V^2/2$. A step-index fiber becomes single mode when $V < 2.405$.

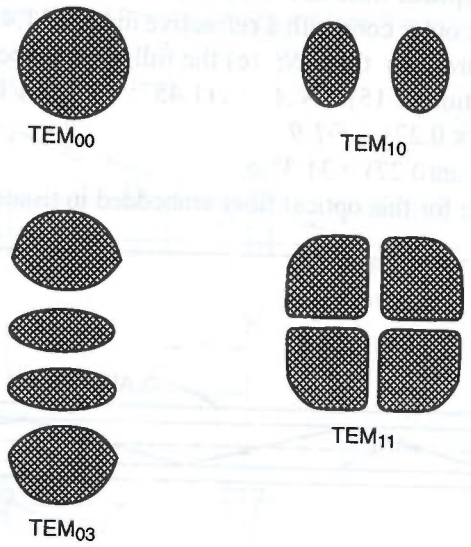


Figure 2.20 Projection of some low-order tansverse modes of a laser cavity. The TEM_{00} mode has a gaussian irradiance profile.

Example 2.13 Modes in a step-index fiber

A 100 μm diameter optical fiber has a *N.A.* equal to 0.27 at 633 nm. Calculate: (a) the *V*-number; (b) the number of linearly polarized modes; (c) the maximum radius for single-mode operation.

(a) $V = 2\pi \times 0.27 \times 50 \times 10^{-6}/633 \times 10^{-9} = 134$

(b) $134^2/2 = 9000$ modes.

(c) The maximum radius for single-mode operation is given by: $2\pi \times 0.27 \times a / 633 \times 10^{-9} < 2.405$, leading to $a = 0.90$ μm.

A multi-mode fiber can transport a much higher optical power than a single-mode fiber, about 3000 times higher than in Example 2.13. The advantage of single-mode operation lies in lower *inter-modal dispersion*. Since each mode in a multi-mode fiber propagates with a different optical path, there is a large variation in the arrival times of the various modes. This time delay between the

arrival of the slowest and fastest modes sets an upper limit on the optical bandwidth and limits the amount of information that can be transmitted. Inter-modal dispersion is absent in a single-mode fiber and the bandwidth is limited by the dependence of refractive index on wavelength. Only laser sources can couple sufficient power into a small-diameter single-mode fiber and accurate alignment is required at inter-fiber junctions. Coupling a laser into an optical fiber is illustrated in Figure 2.21. The general guidelines for efficient coupling are: (1) The focused spot should be comparable to the core size; (2) The *N.A.* of the lens should not exceed the *N.A.* of the fiber.

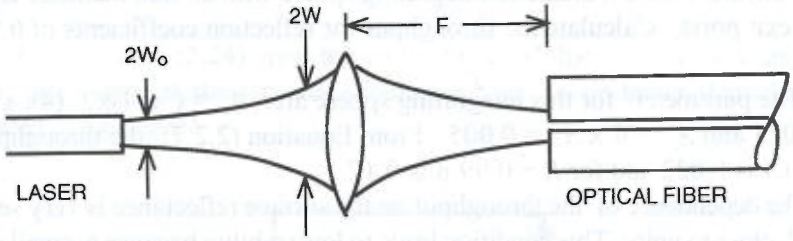


Figure 2.21 Coupling a laser beam into an optical fiber. For optimal coupling the diameter of the focused laser beam equals the diameter of the fiber core, and the numerical aperture of the lens equals the numerical aperture of the optical fiber.

2.6.4 The Integrating Sphere

The integrating sphere is a venerable device with many new applications. The basic construction of an integrating sphere consists of a hollow spherical shell with a highly reflecting inner surface and small ports to admit and extract radiation (Figure 2.22). Light rays entering the sphere at different angles are "homogenized" by multiple reflections. The integrating sphere is an ideal lambertian light source. An entering ray that strikes any point of the interior surface from any angle illuminates the entire surface of the sphere at constant irradiance. Integrating spheres are used for measurements of total radiant power, to collect light from a wide angle source, as uniform light sources, laser beam attenuators, and for light beam mixing. The addition of an integrating sphere to a spectrophotometer makes it possible to measure reflection spectra of opaque materials. Reflection meaurements are especially useful for thin layers such as paper and for highly opaque materials. The throughput (H) of an integrating sphere is the ratio of the exiting flux to the input flux. A ray tracing calculation leads to

$$H = A_e \frac{R}{[1 - R(1 - A_p)]} \tag{2.23}$$

where A_e is the ratio of the exit port area to the surface area of the entire sphere, A_p is the ratio of all port areas to the sphere area, and R is the reflection coefficient of the interior surface.

Example 2.14 Throughput of an integrating sphere

Consider a 20 cm diameter integrating sphere with 20 mm diameter entrance and exit ports. Calculate the throughput for reflection coefficients of 0.90 and 0.99.

The parameters for this integrating sphere are: $A_e = (\pi/4)x2^2/(4\pi \times 10^2) = 0.00025$ and $A_p = 2 \times A_e = 0.005$. From Equation (2.2.7), the throughput for $R = 0.9$ is 0.022 and for $R = 0.99$ it is 0.17.

The dependence of the throughput on the surface reflectance is very sensitive for R close to unity. This condition leads to low stability because a small change in the reflecting coating can have a large effect on the properties. A less efficient integrating sphere with $R \approx 0.90$ is preferable.

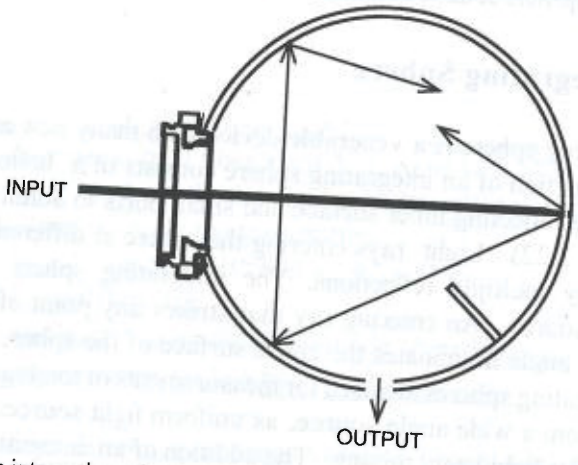


Figure 2.22 An integrating sphere. Light entering the sphere at any angle is diffused by multiple reflections within the sphere. Adapted from Oriel Corporation *catalog Light Sources, Monochromators and Spectrographs, Detectors and Detector Systems, Fiber Optics*, Vol II, 1994, Stratford, CT. With permission.

2.6.5 Coupling of Optical Components

The basic principles of optical coupling are illustrated for a two-lens system in Figure 2.23, where h_1 and h_2 are the heights of a source and receiver, respectively, and θ_1 and θ_2 are the half-angles subtended by the collecting and the focusing lens. The theory of paraxial ray imaging leads to the relationship:

$$\theta_1 h_1 = \theta_2 h_2 \tag{2.24}$$

If $h_1 = h_2$, Equation (2.24) gives the "rule of thumb" that the *F/N* subtended by the optical system at the receiver should be equal to or lower than the *F/N* subtended at the source.

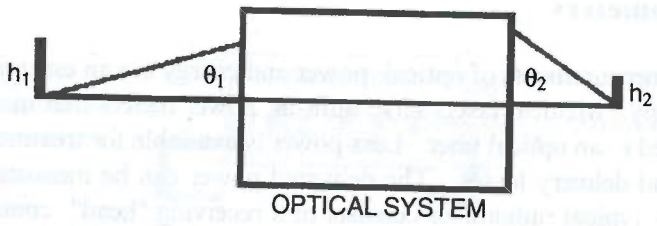


Figure 2.23 Coupling of optical components. An object of height h_1 subtends an acceptance angle θ_1. The image formed by the optical system has a height h_2 and subtends an exit angle θ_2. The most efficient coupling satisfies Equation (2.24).

Example 2.15 Optical coupling of a small lamp to a receiver

An arc lamp is focused on a 2 mm high receiver using an *F/1.5* collecting lens and a *F/4* focusing lens. (a) Calculate the maximum length of source that can be imaged. (b) If both lens have 50 mm apertures, calculate the focal length of each lens and the magnification.

(a) The approximate values of θ for the collecting and focusing lenses are: $\theta_1 = 1/(2 \times 1.5) = 0.33$ rad $= 19°$; $\theta_2 = 1/(2 \times 4) = 0.125$ rad $= 7°$. According to Equation (2.28): $\theta_1 h_1 < 2 \times 7 = 14$ mm-deg. The longest source imaged is: $14/19 = 0.7$ mm.

(b) The focal lengths are: $F_1 = 50 \times 1.5 = 75$ mm; $F_2 = 50 \times 4 = 200$ mm. The magnification of the source at the receiver is given by: $M = F_2/F_1 = 200/75 = 2.7$.

Equation (2.24) is applicable for ordinary light sources. Lasers require a different treatment that is described below in connection with launching light into an optical fiber. Equation (2.24) is a special case of a more general relationship relating the flux exiting a source to the flux at the receiver

$$A_1\Omega_1=A_2\Omega_2 \tag{2.25}$$

where A_1 and A_2 are the cross-sectional areas of the source and receiver and Ω_1 and Ω_2 are the solid angles subtended by the input and exit apertures, respectively.[10] The angle-averaged radiance is defined by: $L_{ave} = \Gamma/A\Omega$, where Γ is the radiant flux . Equating the flux leaving the source to that reaching the receiver shows that L_{ave}(source) = L_{ave} (receiver). This calculation proves that *an image cannot be brighter than the source.*

2.6.6 Radiometers

Accurate measurements of optical power and energy are an essential aspect of phototherapy. Medical lasers have built-in power meters that measure the power delivered to an optical fiber. Less power is available for treatment owing to coupling and delivery losses. The delivered power can be measured with a radiometer. A typical radiometer consists of a receiving "head" containing an optical sensor connected to a calibrated readout device. Radiometers are deceptively simple instruments because the validity of the reading depends on the measurement geometry and spectral match between the light source and the photodetector. Two basic types of photodetectors are available. A *thermal detector* responds to the heat generated by absorption of radiant energy. A *photoelectric detector* converts light to an electrical signal. The operating characteristics of radiometers include the useful wavelength range, the lowest and highest power levels that can be measured, and the ability to measure pulsed and CW light. Commercial photodetectors are characterized by standard parameters: (1) responsivity - detector output per unit optical power input; (2) response time constant - time required for the photodetector output to reach 63% of the maximum value; (3) linearity - range of input irradiance over which the output signal is a linear function of the input; (4) detectivity , a figure of merit indicating the minimum measurable optical power exceeding electrical noise.

The response of a thermal detector is independent of wavelength over a very wide spectral range. The thermopile is an early type of thermal detector. This device consists of an array of closely-spaced junctions of two dissimilar metals

(Figure 2.24a). The "hot" junctions are illuminated by the optical radiation and the "cold" junctions are maintained at a constant temperature. The temperature difference between the hot and cold junctions generates a thermoelectric voltage proportional to the absorbed radiant power. The junctions in a modern thermopile are fabricated from a thin metal film coated with an absorbing layer of black paint or metal oxide. Thermopiles have a slow time response and are most useful for CW measurements. The low responsivity limits the applications to relatively strong light sources. Another type of thermal detector is fabricated from a *pyroelectric crystal* such as lead zirconate and lithium tantalate. In these crystals a permanent electric field exists across opposite faces. Thermal expansion induced by absorption of radiation induces a voltage change (Figure 2.24b). A typical pyroelectric sensor is about 100 times more sensitive than a thermopile and has a faster time response. However, only pulsed and "chopped" CW signals can be measured. The "flat" spectral response of thermal detectors makes them useful for measurements of total power delivered by a very broad wavelength light source such as sunlight.

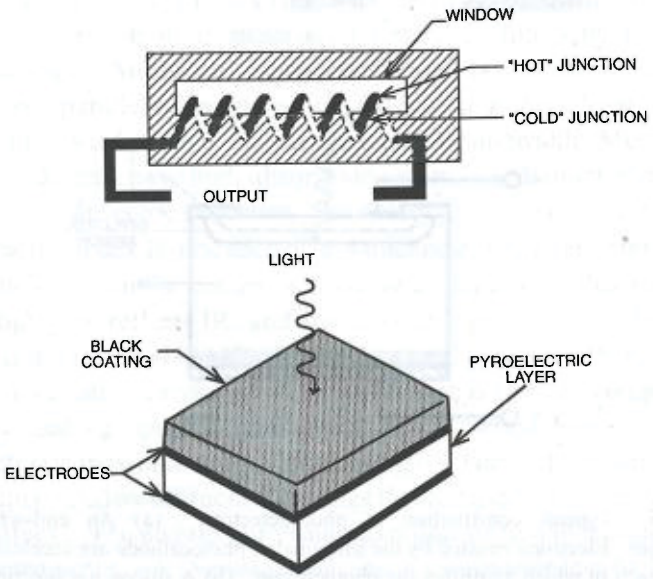


Figure 2.24 Typical construction of thermal detectors. (a) A thermopile is an alternating array of two dissimlar metals. (b) A pyroelectric detector is a ferroelectric crystal in which heating alters the "built-in" electric field across the external faces.

A photodetector converts incident photons to an electric current. The useful wavelength range and spectral sensitivity vary widely with the light-sensitive element. A vacuum phototube is an early type of photoelectric detector

in which electrons are emitted by an illuminated photocathode and a collected by an anode. The primary photocurrent is amplified many times in a *photomultiplier* (Figure 2.25a). Electrons emitted by the photocathode are accelerated by series of "dynodes", each of which emits a higher electrical current than it receives. Photomultipliers are fabricated with many different photocathodes, covering the spectral range from 80 nm in the far-VU to 1100 nm in the near-IR. They are rugged, highly sensitive devices, and respond to both CW and light pulses in the nanosecond range. A faster time response is achieved in a *microchannel plate* that uses an array of approximately one million electron-multiplying elements. Microchannel plates are used in scanning electron multipliers and image

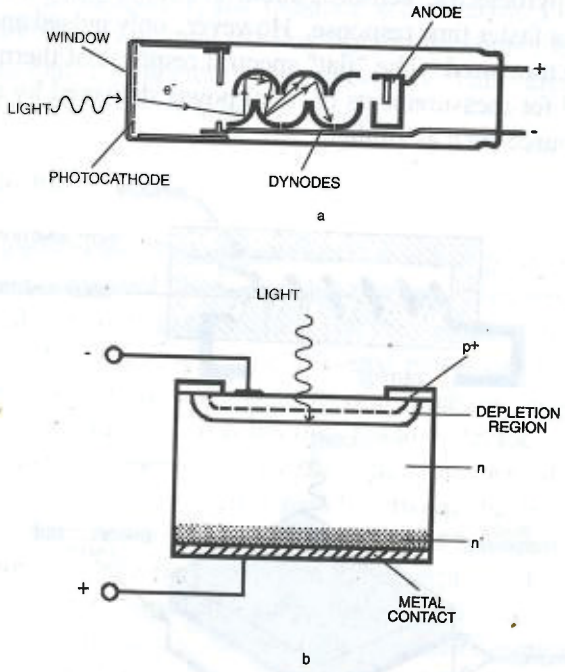


Figure 2.25 Typical construction of photodetectors. (a) An end-window vacuum photomultiplier. Electrons emitted by the illuminated photocathode are accelerated by a series of dynodes, each of which amplifies the photocurrent. (b) A silicon p-n junction photodiolde. Light absorption in the p^+, depletion, and n regions generates electrons and holes that are accelerated by the built-in voltage into the n^- and p^+ regions, respectively. In unbiased operation, the charge separation is measured with a sensitive voltmeter. Application of reverse bias generates a photocurrent. Adapted from Oriel Corporation *catalog Light Sources, Monochromators and Spectrographs, Detectors and Detector Systems, Fiber Optics*, Vol II, 1994, Stratford, CT. With permission.

intensifiers. Except for measurements of ultrashort pulses, photomultipliers have been largely superceded by solid state devices. A review of semiconductor physics in Section 3.6.3 is applicable for both solid state photodetectors and semiconductor lasers.

2.7 WAVELENGTH SELECTION

A wavelength selection device is used to modify the spectral output of a light source. Virtually any optical phenomenon can be adapted for wavelength selection, including absorption, dispersion, interference, diffraction, birefringence, and light scattering. Color glasses incorporating inorganic salts with selective absorptions are the least expensive type of color filter, although the characteristics are limited by the available materials. A "long-pass" filter is strongly absorbing below a cutoff wavelength and transmitting at longer wavelengths and *vice versa* for a "short-pass" filter. A "band-pass" filter is highly transmitting over one or several wavelength bands. More expensive interference filters may be fabricated as required, adding an integral blocking layer. Narrow bandwidth is achieved at the expense of throughput in practical interference filters by using highly reflective coatings. An optical cavity consisting of a dielectric (or air) layer located between parallel reflectors is referred to as a *Fabry-Perot*. The Fabry-Perot units can be stacked for a further reduction in bandwidth. Metallic mirrors used in earlier designs have high absorption losses. Modern interference filters utilize multilayer dielectric reflectors. A stack of alternate high-refractive index and low-refractive index layers, each of $\lambda/4$ thickness, has a very high reflectance at wavelength λ. A similar design is used for a *hot mirror* that transmits UV and visible light and reflects IR, and for a *cold mirror* that reflects UV and visible light and transmits IR. Hot mirrors are used to remove IR heat rays from arc lamps. A variable wavelength selection device is the key component in all instruments used for spectral measurements. A *monochromator* selects a narrow spectral bandwidth out of incident light. The early prism instruments have been superceded by diffraction gratings that are used on virtually all modern monochromators.. The addition of visual detection system to a monochromator gives a *spectrometer*. A *spectrograph* is a spectrometer in which the output spectrum is recorded on a photographic plate or a photodiode array. Fast temporal changes in output spectra can be measured with a *streak camera* in which a photographic film is displaced at a high speed. A *spectroradiometer* is a spectrometer plus a photodetector for measurements of radiant power or energy. A *spectrophotometer* consists of a monochromator with a light source at the input and a photodector at the output. This device is used to record absorption or emission spectra of samples placed in the optical path.

NOTES

1. A Fourier series expresses an arbitrary function as an infinite sum of discrete sine and cosine functions. This representation is very convenient for periodic wave motion. Each term in a Fourier series corresponds to a monochromatic sinusoidal wave. A Fourier integral extends the Fourier series to the expansion of non-periodic functions. In a Fourier integral the discrete terms in the Fourier series are replaced by an integral over pairs of conjugate variables such as frequency and time.

2. The International System of units (SI) are standard in physics. The SI base units are: length (meter, m), mass (kilogram, kg), time (second, s), and amount of substance (mole, mol). Standard prefixes are used when dealing with very large or very small numbers, Some of the most widely used prefixes are kilo (10^3), mega (10^6), giga (10^9), and tera (10^{12}) for large quantities and deci (10^{-1}), centi (10^{-2}), milli (10^{-3}), micro (10^{-6}), nano (10^{-9}) pico (10^{-12}), and femto (10^{-15}) for small quantities. Other SI units are defined for specific properties including energy (joule, J), power (watt, W), electric potential difference (volt, V), and magnetic flux (weber, Wb).

3. Many molecules including water have a built-in separation of electric charge which is equivalent to an electric dipole. An alternating electric field induces a back and forth motion of the dipoles referred to as a dielectric interaction. The friction associated with this motion heats the material. Microwave ovens utilize this effect.

4. The electron volt (eV) is the amount of kinetic energy acquired by an electron when it accelerated by a potential difference of one volt. The utility of this unit is based on the fact that the same energy increment applies for any particle having the same charge as the electron, for example, a proton. However, the 1 eV proton is moving about 40 times slower than a 1 eV electron owing to the heavier mass.

5. This calculation is slightly ambiguous because according to Equation (2.1), changes in V can be related to changes in f, λ, or both. Example 2.2 assumes that f is the same in air and glass while λ differs in each medium. This assumption is justified by classical electromagnetic theory. For the simpler case of elastic waves in adjoining media the particles at an interface between different media must oscillate at the same frequency, for example, vibrations of a thin string joined to a heavier cord.

6. The reader familiar with ultrasound may observe that Equation (2.4) resembles the expression for reflection ultrasound at the interface between two regions of different specific acoustic impedance Z. Equation (2.6) for normal incidence gives the ultrasound reflection coefficient with n_r replaced by Z_2/Z_1.

7. Soliton motion describes a unique type of wave in which spreading does not take place. Soliton waves were first reported in 1834 by Russell in the form of a high-speed "solitary" wave that continued without change of form or speed in the Glasgow-

Edinburgh canal. Very short optical pulses can propagate in optical fibers as solitons.

8. In circularly polarized light, the mutually perpendicular vectors representing the E-field and B-field rotate together at the wave frequency and constant amplitude. It is easily shown that linearly polarized light can be resolved into two equal-amplitude circular polarizations rotating in opposite directions.

9. More complicated optical systems can be analyzed with ray tracing computer programs. A general analytical technique applicable for paraxial rays uses ray matrices. The action of each optical element is represented by a "2 x 2" array or "ABCD" matrix. The effect of a complex optical system on an incident ray is calculated by the rules of matrix algebra. This method is suitable for cascaded groups of lenses, mirrors, and spatial regions of variable refractive index or "ducts".

10. A geometrical calculation for paraxial rays shows that W is related to q by: $W = 2p(1 - \cos q)^2 \approx pq^2$. Substituting for W and taking the square root of both sides leads to Equation (2.24).